



## Genome and metabolome mining of marine obligate *Salinispora* strains to discover new natural products

Süleyman ÖZAKIN<sup>1,2</sup> , Ebru İNCE<sup>1,\*</sup> 

<sup>1</sup>Department of Biology, Faculty of Science, Dicle University, Diyarbakır, Turkey

<sup>2</sup>Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California San Diego, La Jolla, California, USA

Received: 17.07.2018 • Accepted/Published Online: 29.11.2018 • Final Version: 07.02.2019

**Abstract:** Marine microorganisms are receiving more attention as a promising potential source of new natural products. In the present study, we performed genomic and metabolomic analyses to explore the metabolic potential of the obligate marine actinomycete genus *Salinispora*. The genomes of thirty *Salinispora* strains were prospected in search of biosynthetic gene clusters including polyketide synthase (PKS), nonribosomal peptide synthetase (NRPS), terpene, indole, lantibiotics, and siderophores. We determined considerable diversity of natural product biosynthetic gene clusters in their genome. There were a total of 1428 putative gene clusters involved in the biosynthesis of various bioactive natural products. Furthermore, 1509 ketosynthase (KS) and condensation (C) domains were detected by using NapDoS belonging to PKS and NRPS genes, respectively. Metabolic profiling was performed by a nontargeted LC-MS/MS approach combined with spectral networking using Global Natural Product Social Molecular Networking (GNPS). Dereplication and tentative identification of natural products were evaluated for common chemical properties and their associated pathways. Significant bioactive natural products such as lomaiviticin C, 7-OH-staurosporine, staurosporine, and cyanosporaside B were determined. More importantly, an unknown glycosylated compound associated with an NRPS/PKS-I hybrid gene cluster in *Salinispora pacifica* CNY703 was established through chemical and genomic analyses.

**Key words:** *Salinispora*, genome mining, metabolomics, mass spectrometry, GNPS, marine natural products

### 1. Introduction

Natural products are biological molecules produced by organisms such as fungi, plants, and microorganisms. The biotechnological potential of natural products from microorganisms is receiving more attention for the discovery of novel bioactive compounds (Qin et al., 2017). Of all microbes, the actinomycetes traditionally represent one of the most important sources. Marine actinomycetes belonging to the genus *Salinispora* have long been an important source of structurally diverse and biologically active natural products, several of which have inspired the development of new classes of therapeutic agents (Feling et al., 2003; Fenical and Jensen, 2006; Jensen and Mafnas, 2006). Polyketide- and peptide-derived metabolites are among the most diverse and include many clinically important compounds (Fischbach and Walsh, 2006; Jang et al., 2013).

In recent years, genomics and metabolomics have been combined to identify new bioactive metabolites. The mining of actinomycete genomes has proven to be useful in the identification of secondary metabolite biosynthetic

gene clusters (Jensen and Mafnas, 2006; Zerikly and Challis, 2009). In untargeted metabolomic studies, liquid chromatography followed by mass spectrometry (LC/MS) has been widely used to detect the highest number of metabolites in small amounts of sample. Identification of these compounds is based on tandem MS (MS/MS) data, produced by fragmenting the compound and determining the masses of the fragments. Global Natural Products Social Molecular Networking (GNPS, <http://gnps.ucsd.edu>) is an open-access knowledge base for public sharing of processed and annotated MS/MS spectrometry data. The molecular networks created by GNPS enable dereplication (rapid identification of known metabolites) and structural identification of metabolites through spectrum library matching. The MS-guided genome mining technique helps to bridge the gaps between genes, pathways, and chemical features of metabolites. It creates an algorithm capable of comparing characteristic fragmentation patterns, thus composing molecular groups with the same structural features and probably the same biosynthetic origin (Wang et al., 2016). In this respect, we used a combined

\* Correspondence: [eince@dicle.edu.tr](mailto:eince@dicle.edu.tr)

genomic and metabolomic mining approach to highlight the natural product biosynthetic capacity of 30 marine obligate *Salinispora* strains.

## 2. Materials and methods

### 2.1. Bacteria and fermentation studies

The names of the 30 *Salinispora* strains used in this study and their genome accession numbers are listed in Table 1. Glycerol stock solutions of all bacteria were prepared by inoculating 10 µL of cell stock into 25 mL of A1 medium

**Table 1.** *Salinispora* strains used in this study and genome accession numbers.

Strains	Genome accession numbers
<i>Salinispora pacifica</i> CNS801	2561511036
<i>Salinispora pacifica</i> CNY703	2563366517
<i>Salinispora pacifica</i> CNS860	2518285563
<i>Salinispora pacifica</i> CNR909	2561511038
<i>Salinispora pacifica</i> CNY666	2563366532
<i>Salinispora pacifica</i> CNY239	2524614561
<i>Salinispora pacifica</i> CNT796	2515154182
<i>Salinispora pacifica</i> CNT603	2515154185
<i>Salinispora pacifica</i> CNT124	2517572159
<i>Salinispora pacifica</i> CNQ768	2517572155
<i>Salinispora pacifica</i> CNT851	2517572162
<i>Salinispora arenicola</i> CNR107	2519103194
<i>Salinispora arenicola</i> CNY011	2517572153
<i>Salinispora arenicola</i> CNY230	2561511115
<i>Salinispora arenicola</i> CNR425	2528311033
<i>Salinispora arenicola</i> CNY256	2518285559
<i>Salinispora arenicola</i> CNS820	2565956528
<i>Salinispora arenicola</i> CNS299	2524614529
<i>Salinispora arenicola</i> CNT800	2515154088
<i>Salinispora arenicola</i> CNS673	2519103185
<i>Salinispora arenicola</i> CNH877	2519103192
<i>Salinispora arenicola</i> CNH963	2524023246
<i>Salinispora arenicola</i> CNY679	2561511113
<i>Salinispora arenicola</i> CNS325	2571042009
<i>Salinispora arenicola</i> CNT798	2515154186
<i>Salinispora arenicola</i> CNH643	2561511037
<i>Salinispora arenicola</i> CNT850	2515154135
<i>Salinispora arenicola</i> CNH962	2519103193
<i>Salinispora arenicola</i> CNT799	2526164509
<i>Salinispora tropica</i> CNT250	2540341193

containing 10 g/L starch, 4 g/L yeast extract, 2 g/L peptone, and 22 g/L Instant Ocean sea salt (Instant Ocean®) at pH 6.5 and were incubated at 25 °C for 6–10 days.

For the production of metabolites, all isolates were grown in triplicate in 100 mL of A1M1 medium containing 5 g/L starch, 2 g/L yeast extract, 2 g/L peptone, and 22 g/L Instant Ocean sea salt at pH 6.5 and were incubated at 25 °C with shaking at 160 rpm for 14–20 days.

### 2.2. Genome mining

A total of 30 *Salinispora* genomes were downloaded from the Joint Genome Institute's Integrated Microbial Genomes (IMG) database (<http://img.jgi.doe.gov>). The draft genome sequences of all *Salinispora* strains were analyzed by NapDoS (Ziemert et al., 2012) and antiSMASH 2.0 (Blin et al., 2013). NapDoS was used to detect and extract KS and C domains in PKS and NRPS genes in the genomes, respectively. antiSMASH 2.0 was used to detect secondary metabolite biosynthetic gene clusters with the whole range of known secondary metabolite compound classes, including polyketides, nonribosomal peptides, lantipeptides, oligosaccharide antibiotics, phenazines, thiopeptides, homo-serine lactones, phosphonates, and furans. Homologous clusters for predicted biosynthetic pathways were analyzed using the MultiGeneBLAST program (Medema et al., 2013). DoBISCUIT software was also used to screen a variety of gene clusters for secondary metabolite biosynthesis (Ichikawa et al., 2012).

### 2.3. Extraction and spectroscopic analysis of metabolites

The supernatant and pellets were extracted with ethyl acetate (1:1, v/v). The extracts were dried with Na<sub>2</sub>SO<sub>4</sub> and evaporated to dryness under reduced pressure to yield crude extracts. After being weighed, the extracts were dissolved in methanol to obtain a final concentration of 1 mg/mL. LC and LC-MS/MS analyses were carried out after filtration through 0.2-µm Acrodisc MS syringe filters (25 mm).

The samples were injected as 20 µL into an Agilent 1260 LC system with an Agilent Extend-C18 RP UPLC column (2.1 × 100 mm, 1.8 µm) connected to an Agilent 6530 Accurate-Mass Q-TOF LC/MS. The LC gradient was as follows: 10% (v/v) acetonitrile (ACN) (0.1% water, 0–3 min), 10–100% (v/v) ACN (0.1% water)/0.1% water (3–23 min), 100% ACN (0.1% water, 23–25 min), 10% (v/v) ACN (0.1% water, 25–30 min). The column compartment temperature was 25 °C.

Q-TOF MS settings during the LC gradient were as follows: acquisition mass range m/z 100–1600, MS scan rate 1s<sup>-1</sup>, MS/MS scan rate 2s<sup>-1</sup>, fixed collision energy 20 eV, source-gas temperature 300 °C, gas flow 11 L min<sup>-1</sup>, nebulizer 45 psi, ion polarity positive; scan source parameters—VCap 3000, Fragmentor 100, Skimmer1 65, OctopoleRFPeak 750. The MS was autotuned using Agilent tuning solution in positive mode before each

measurement. LC (DAD) and MS data were analyzed with ChemStation and MassHunter software (Agilent), respectively.

#### 2.4. GNPS molecular networking and dereplication analyses

The raw MS/MS data of 30 *Salinispora* strains were converted from MassHunter data files (.d.) to mzXML file format using the Trans Proteomic Pipeline.

MS/MS data of molecular network and dereplication analyses were determined with GNPS molecular networking. Before the analyses, spectral networks were imported into Cytoscape 3.1.0 for visualization as a network. Cytoscape software was employed to visualize the molecular networks and biological pathways. It integrates these networks with annotations, gene expression profiles, and other state data ([www.cytoscape.org](http://www.cytoscape.org)). Some parent ions obtained from LC/MS data were analyzed with METLIN (Smith et al., 2005) and MarinLit (<http://pubs.rsc.org/marinlit/>) databases. Glycogenomic analyses were carried out as described previously (Kersten et al., 2013).

### 3. Results and discussion

#### 3.1. Genomic analysis of secondary metabolite genes

Based on the genome analysis, *Salinispora* strains appeared to be rich sources for production of various chemical entities and new secondary metabolites. They have a large proportion of natural product biosynthetic gene clusters (about 9.9%) in their genomes in comparison to terrestrial counterparts *Streptomyces coelicolor* (4.5%) and *Streptomyces avermitilis* (6%) (Bentley et al., 2002; Ikeda et al., 2003). According to antiSMASH results, 1428 putative natural product gene clusters were found to be involved in the biosynthesis of various pathways such as NRPS, PKS, terpenes, butyrolactones, siderophores, bacteriocins, and lantibiotics. In total, 1509 ketosynthase (KS) and condensation (C) domains were detected in all strains belonging to their PKS and NRPS genes by using NapDoS, respectively. NRPS and PKS gene clusters were remarkably higher than other biosynthetic genes in almost all strains (Table 2). The *Salinispora arenicola* strains, particularly CNR425, were identified as hot spots for PKS and NRPS biosynthetic capacity (Figure 1). Among the 30 *Salinispora* strains, *S. arenicola* CNS673 and *S. arenicola* CNT799 both contain eleven PKS-I pathways. In addition, the highest number of PKS-II biosynthetic gene clusters was identified in both *S. pacifica* CNT851 and CNT796 strains. It has been previously reported that certain *Salinispora* genomes were enriched in PKS and NRPS biosynthetic pathways (Ziemert et al., 2014). Furthermore, the average number of natural product gene clusters identified per genome was significantly greater in *S. arenicola* than in *S. pacifica* and *S. tropica* (Letzel et al., 2017). In our study, all *S. arenicola* strains produced the most popular polyketide

compound, rifamycin, many of whose analogs have long been used to treat mycobacterial infections. Apart from *S. pacifica* CNS860, the other 29 strains host lymphostin biosynthetic gene clusters in their genomes. Highly conserved lymphostin gene clusters were also reported to exist in *Salinispora* bacteria by Miyanaga et al. (2011). *S. pacifica* CNS860 and *S. tropica* CNT250 strains have been able to produce salinosporamide A, which is currently being employed in clinical trials for the treatment of cancer (Feling et al., 2003). Sporolide biosynthetic gene clusters were determined in the genomes of two strains: *S. tropica* CNT250 and *S. arenicola* CN963. Cyanosporaside, salinilactam, and cyclomarin pathways were rarely seen in the analyzed strains. A total of 23 *Salinispora* strains could produce siderophores. These biologically active iron chelators have been playing important roles in adaption to unstable environmental conditions.

In the present study, it is observed that *S. arenicola* strains (CNR107, CNS673, CNS325, CNY230, CNR425, CNS820, CNH962, CNH963, and CNY299) encode biosynthetic genes related to glycosylated pathways in NPRS and PKS gene clusters. Three different putative glycosylated pathways were detected in *S. pacifica* CNY703 by the antiSMASH program. These pathways lead to the synthesizing of the following compounds: (a) staurosporine, (b) an unknown hybrid NRPS-PKS, and (c) a NRPS. The gene clusters of the first two pathways (staurosporine and an unknown hybrid NRPS-PKS) have common specific sugar biosynthetic genes, including NDP sugar epimerases, aminotransferase, 4,6-dehydratase, Glu1P N transferase, CH<sub>3</sub> transferase, glycosyltransferase, and NDP hexose 2,3-dehydratase. These genes are important markers for the discovery of glycosylated natural products by glycogenomic approaches (Kersten et al., 2013). However, O-CH<sub>3</sub> transferase, which is the specific glycosylation gene, was discovered only in the gene cluster of the unknown NRPS-PKS compound (b), and not in the staurosporine (a). It has been determined that the staurosporine and unknown NRPS/PKS-I gene clusters are homologs, but their products were synthesized by different pathways when analyzed by MultiGeneBlast.

#### 3.2. Dereplication and molecular networking analysis by GNPS

Dereplication is an important approach to rapidly identify known natural products in complex extracts. To create a molecular network for dereplication and structural identification of metabolites, the GNPS database was used. Molecular networking is a platform that provides a summary of mass spectrometry-based metabolomics by comparing molecular properties with fragmentation patterns to highlight chemical relationships. Each node represents a consensus spectrum and edges represent related fragmentation patterns. Thus, clusters in the

**Table 2.** The pathways detected by AntiSMASH and NapDOS for *Salinispora* strains. The presence and absence of pathways are shown by filling the boxes with gray and black colors, respectively.

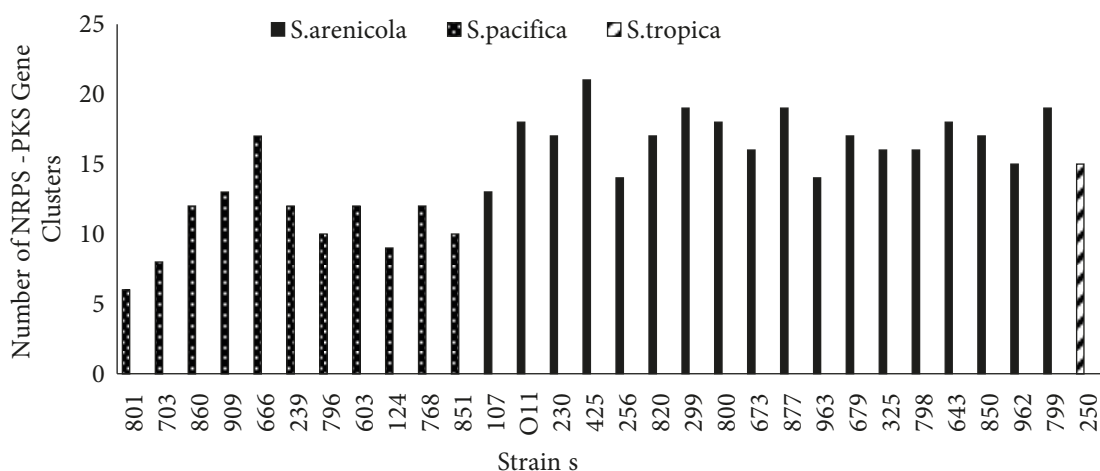
Salinispora strains	Pathways												
	NRPS	PKSI	PKSII	NRPS-PKS	Lan	Sid	Rif	Sal A	Lym	Cya	Slm	Cym	Spo
CNS801	Gray	Black	Gray	Gray	Gray	Gray	Black	Black	Gray	Black	Black	Black	Black
CNY703	Gray	Gray	Gray	Gray	Gray	Gray	Black	Black	Gray	Gray	Black	Black	Black
CNS860	Gray	Gray	Gray	Gray	Gray	Gray	Black	Gray	Black	Black	Black	Black	Black
CNR909	Gray	Gray	Gray	Black	Gray	Black	Black	Black	Gray	Gray	Black	Black	Black
CNY666	Gray	Gray	Gray	Black	Gray	Black	Gray	Black	Gray	Gray	Black	Gray	Black
CNY239	Gray	Gray	Gray	Gray	Gray	Black	Black	Black	Gray	Gray	Black	Black	Black
CNT796	Gray	Gray	Gray	Black	Gray	Black	Black	Black	Gray	Gray	Black	Black	Black
CNT603	Gray	Gray	Gray	Black	Gray	Black	Black	Black	Gray	Gray	Black	Black	Black
CNT124	Gray	Gray	Gray	Gray	Gray	Gray	Black	Black	Gray	Black	Black	Gray	Black
CNQ768	Gray	Gray	Gray	Black	Gray	Black	Black	Black	Gray	Gray	Black	Gray	Black
CNT851	Gray	Gray	Gray	Black	Gray	Black	Black	Black	Gray	Gray	Black	Black	Black
CNR107	Gray	Gray	Gray	Gray	Gray	Gray	Black	Black	Gray	Black	Black	Black	Black
CNY011	Gray	Gray	Gray	Gray	Gray	Gray	Black	Black	Gray	Black	Black	Gray	Black
CNY230	Gray	Gray	Gray	Gray	Gray	Gray	Black	Black	Gray	Black	Black	Black	Black
CNR425	Gray	Gray	Gray	Gray	Gray	Gray	Black	Black	Gray	Black	Black	Gray	Black
CNY256	Gray	Gray	Gray	Gray	Gray	Gray	Black	Black	Gray	Black	Black	Black	Black
CNS820	Gray	Gray	Gray	Gray	Gray	Gray	Black	Black	Gray	Black	Black	Black	Black
CNS299	Gray	Gray	Gray	Gray	Gray	Gray	Black	Black	Gray	Black	Black	Black	Black
CNT800	Gray	Gray	Gray	Gray	Gray	Gray	Black	Black	Gray	Black	Black	Black	Black
CNS673	Gray	Gray	Gray	Gray	Gray	Gray	Black	Black	Gray	Black	Black	Black	Black
CNH877	Gray	Gray	Gray	Gray	Gray	Gray	Black	Black	Gray	Black	Black	Gray	Black
CNH963	Gray	Gray	Gray	Gray	Gray	Gray	Black	Black	Gray	Black	Black	Black	Gray
CNY679	Gray	Gray	Gray	Gray	Gray	Gray	Black	Black	Gray	Black	Black	Black	Black
CNS325	Gray	Gray	Gray	Gray	Gray	Gray	Black	Black	Gray	Black	Black	Black	Black
CNT798	Gray	Gray	Gray	Gray	Gray	Gray	Black	Black	Gray	Black	Black	Black	Black
CNH643	Gray	Gray	Gray	Gray	Gray	Gray	Black	Black	Gray	Black	Black	Black	Black
CNT850	Gray	Gray	Gray	Gray	Gray	Gray	Black	Black	Gray	Black	Black	Black	Black
CNH962	Gray	Gray	Gray	Black	Gray	Gray	Black	Black	Gray	Black	Black	Black	Black
CNT799	Gray	Gray	Gray	Gray	Gray	Gray	Black	Black	Gray	Black	Black	Black	Black
CNT250	Gray	Gray	Gray	Gray	Black	Gray	Black	Gray	Gray	Black	Gray	Black	Gray

Lan: Lantibiotic, Sid: siderophore, Rif: rifamycin, Sal A: salinisporamide A, Lym: lymphostin, Cya: cyanosporaside, Slm: salinilactam, Cym: cyclomarin, Spo: sporolide.

network are represented by the molecular ions that emphasize the structural relationships and similarities of the molecules.

Thirty strains were grown in triplicate and 90 bacterial crude extracts were obtained to analyze the metabolites with LC-MS/MS and subsequently with GNPS molecular networking. According to dereplication

analysis, 21 strains were determined to produce some well-known compounds in the GNPS database (Table 3). A marine sponge-derived natural product, mycalamide A, which is known as a protein synthesis inhibitor with potent antitumor activity, was detected in two *S. pacifica* strains (CNY239 and CNT124). The isolation and in vitro antiviral activity of mycalamide A and mycalamide



**Figure 1.** The number of NRPS-PKS pathways of *Salinispora* strains detected by antiSMASH.

**Table 3.** The results of dereplication analysis of *Salinispora* strains.

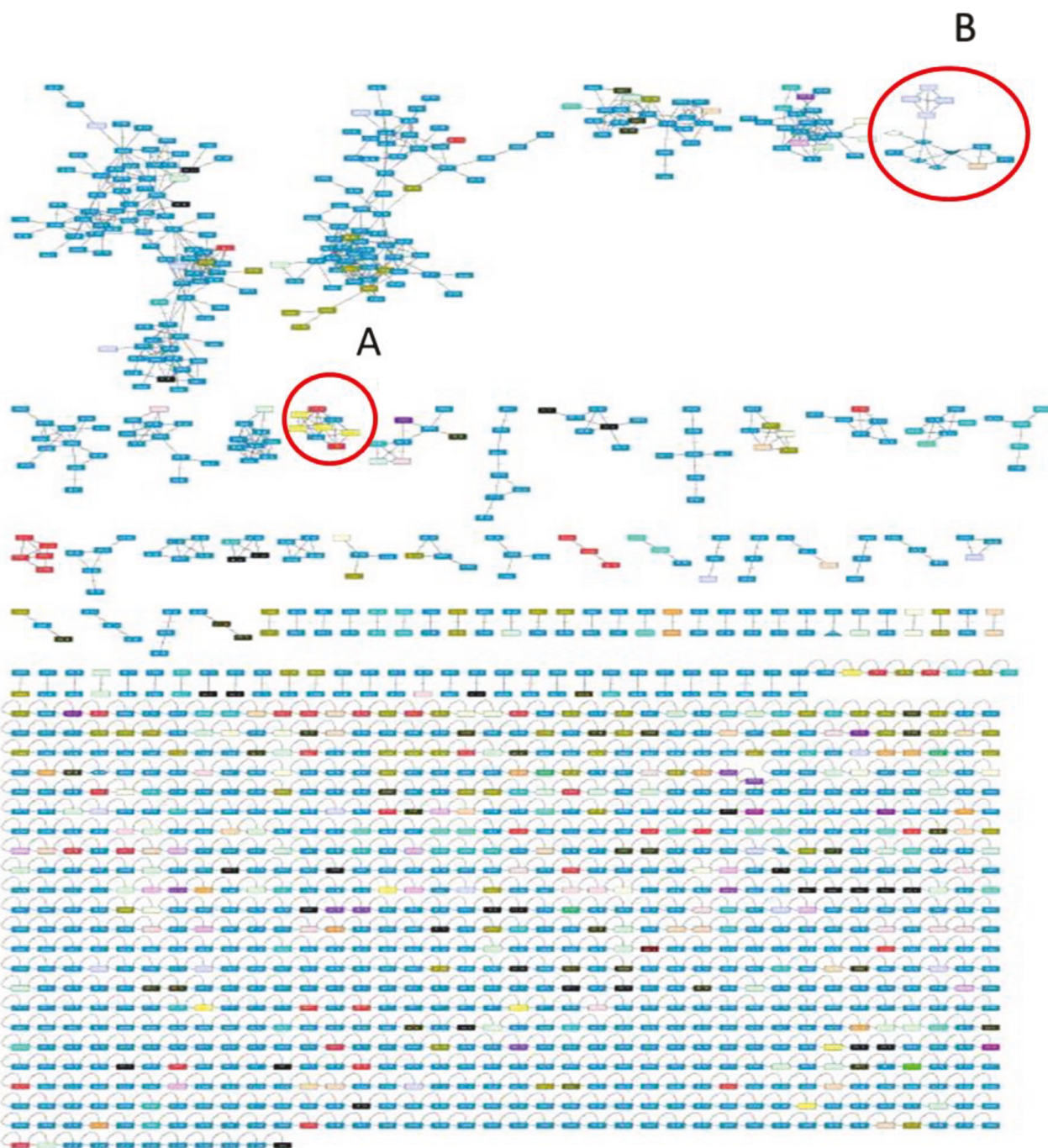
<i>Salinispora</i> strains	GNPS library hits				
<i>S. pacifica</i> CNS801	-				
<i>S. pacifica</i> CNY703	-				
<i>S. pacifica</i> CNS860	-				
<i>S. pacifica</i> CNR909	Cyanosporaside B				
<i>S. pacifica</i> CNY666	-				
<i>S. pacifica</i> CNY239	Cyanosporaside B	Mycalamide A			
<i>S. pacifica</i> CNT796	Enterocin	Lomaiviticin C			
<i>S. pacifica</i> CNT603	Cyanosporaside B	Lomaiviticin C			
<i>S. pacifica</i> CNT124	Mycalamide A				
<i>S. pacifica</i> CNQ768	-				
<i>S. pacifica</i> CNT851	-				
<i>S. arenicola</i> CNR107	7-OH Staurosporine	Staurosporine	Rifamycin S		
<i>S. arenicola</i> CNY011	-				
<i>S. arenicola</i> CNY230	7-OH Staurosporine	Staurosporine			
<i>S. arenicola</i> CNR425	Salinisporazine A				
<i>S. arenicola</i> CNY256	7-OH Staurosporine	Staurosporine	Rifamycin S	Saliniketal A	Aphidicolin
<i>S. arenicola</i> CNS820	7-OH Staurosporine	Staurosporine	Rifamycin S	Saliniketal A	
<i>S. arenicola</i> CNS299	7-OH Staurosporine	Staurosporine	Rifamycin S	Rifamycin W	
<i>S. arenicola</i> CNT800	Staurosporine				
<i>S. arenicola</i> CNS673	7-OH Staurosporine	Staurosporine	Rifamycin S	Saliniketal A	Aphidicolin
<i>S. arenicola</i> CNH877	7-OH Staurosporine	Staurosporine	Rifamycin S	Saliniketal B	
<i>S. arenicola</i> CNH963	-				
<i>S. arenicola</i> CNY679	Staurosporine	Rifamycin S	Saliniketal A	Saliniketal B	
<i>S. arenicola</i> CNS325	Rifamycin S	Saliniketal B			
<i>S. arenicola</i> CNT798	Staurosporine	Saliniketal A	Saliniketal B		
<i>S. arenicola</i> CNH643	Staurosporine	Rifamycin S			
<i>S. arenicola</i> CNT850	Staurosporine	Rifamycin S			
<i>S. arenicola</i> CNH962	-				
<i>S. arenicola</i> CNT799	Staurosporine	Rifamycin S	Dinactin		
<i>S. tropica</i> CNT250	Lomaiviticin C				



B were reported from a New Zealand sponge, *Mycale* (Perry et al., 1990). Lomaiviticin C is another bioactive natural product established in dereplication analysis. This antitumor compound was determined to be produced by two *S. pacifica* (CNT796, CNT603) and one *S. tropica* strain (CNT250). Salinisporazine A and enterocin were found in *S. arenicola* CNR425 and *S. pacifica* CNT796

extracts, respectively. The well-known antibiotic rifamycin was dereplicated from most of the *S. arenicola* strains. The indolocarbazole compound staurosporine, which has protein kinase inhibitory activity, is also a common compound in the bacterial crude extracts of 18 strains.

After generating the molecular networks, node connectivity was visualized (Figure 2). One node (colored



**Figure 2.** Molecular networking analysis of 30 strains of *Salinispora*. Clusters A and B represent the lomaiviticin and staurosporine clusters, respectively.



366.214, and 380.23 in this cluster. These nodes (pale purple) come only from *S. pacifica* CNY703. Node m/z 366.214 in the quartet is connected to the staurosporine node (m/z 467.182). Although *S. pacifica* CNY703 has an indolocarbazole biosynthetic gene cluster in its genome, the production of staurosporine has not been detected in this strain. When examining the MS/MS fragmentation behavior of *S. pacifica* CNY703 for cluster B, it seems that the m/z 338.183, 366.214, and 380.23 signals are not parent but fragmented ions. The LC/MS chromatogram was investigated in the range of m/z 600–700. Two parent ions, m/z 661 [M+H] and 683 [M+Na], were identified in the chromatogram connected to fragmented ion m/z 366.214. Furthermore, the UV library search of the two related ions has a top hit matched with an unknown NRPS/PKS-I compound, CNS205. The molecular weight of the related compound was in the range of 650–700, which may correspond to glycine, tryptophan, valine/leucine, ketide chains, and sugar. These chemical features have been found in the core structure of the unknown NRPS/PKS-I hybrid compound according to monomer prediction by antiSMASH.

In a previous study, high levels of pathway diversity associated with polyketide and nonribosomal peptide biosynthesis were found in the 75 analyzed *Salinispora* genome sequences (Ziemert et al., 2014). Although 15 out of 30 strains were identical in both studies (Ziemert et al., 2014 and present study), we carried out this study not only in terms of the genomic but also the chemical perspective to identify genes and corresponding metabolites. Herein, we report a putative novel glycosylated NRPS/PKS-I hybrid metabolite and its gene cluster from *S. pacifica* CNY703 by using the MS-guided genome mining strategy. The combination of genomic and metabolomic data has been used to discover some novel compounds including anticancer retimycin A (Duncan et al., 2015), antifungal

thanamycin (Kersten et al., 2011), and antibacterial vitropocins produced by marine *Vibrio* sp. (Liaw et al., 2015). Kersten et al. (2013) established the connection between predictable glycosylation fragments from MS/MS experiments and underlined the glycosylation genes in microbial genomes. Cinerubin B is an exemplified compound that has been characterized as a glycosylated anthracycline antibiotic from a *Streptomyces* strain. Moreover, glycogenomics has facilitated the discovery of arenimycins, which are N-glycosylated aromatic polyketides from *S. arenicola* exhibiting significant anti-MRSA activity (Kersten et al., 2013).

### 3.3. Conclusions

In this study, we adopted genomics and metabolomics tools in order to highlight the metabolic potential of 30 marine obligate *Salinispora* strains. Fifteen strains of 30 were investigated for the first time with the goal of identifying genes and corresponding metabolites. Genome analyses of these strains by using two different genome mining software programs have provided comparisons of the natural product biosynthetic potential of *Salinispora* strains. The most important finding of this chemoinformatic study is to reveal the previously unknown NRPS/PKS-I gene cluster and unique signals belonging to its product in *S. pacifica* CNY703. Further studies are required to purify the compound and elucidate its structure.

### Acknowledgments

We are grateful to Dr Paul Jensen from the Scripps Institution of Oceanography, Center for Marine Biotechnology and Biomedicine (San Diego, CA, USA) for his guidance and insight throughout this research. This work was supported by a fellowship awarded to Süleyman Özakin by the Scientific and Technological Research Council of Turkey (TÜBİTAK 2214/A Fellowship Program).

### References

- Bentley SD, Chater KF, Cerdeño-Tárraga AM, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D et al. (2002). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* 417: 141-147.
- Blin K, Medema MH, Kazempour D, Fischbach MA, Breitling R, Takano E, Weber T (2013). antiSMASH 2.0—A versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res* 41: W204-212.
- Duncan KR, Crusemann M, Lechner A, Sarkar A, Li J, Ziemert N, Wang MX, Bandeira N, Moore BS, Dorrestein PC et al. (2015). Molecular networking and pattern-based genome mining improves discovery of biosynthetic gene clusters and their products from *Salinispora* species. *Chem Biol* 22: 460-471.
- Feling RH, Buchanan GO, Mincer TJ, Kauffman CA, Jensen PR, Fenical W (2003). Salinosporamide A: a highly cytotoxic proteasome inhibitor from a novel microbial source, a marine bacterium of the new genus *Salinispora*. *Angew Chem Int Ed Engl* 42: 355-357.
- Fenical W, Jensen PR (2006). Developing a new resource for drug discovery: marine actinomycete bacteria. *Nat Chem Biol* 2: 666-673.
- Fischbach MA, Walsh CT (2006). Assembly-line enzymology for polyketides and nonribosomal peptide antibiotics: logic, machinery and mechanisms. *Chem Rev* 106: 3468-3496.



- Ichikawa N, Sasagawa M, Yamamoto M, Komaki H, Yoshida Y, Yamazaki S, Fujita N (2012). DoBISCUIT: A database of secondary metabolite biosynthetic gene clusters. *Nucleic Acids Res* 41: 8-14.
- Ikeda H, Ishikawa J, Hanamoto A, Shinose M, Kikuchi H, Shiba T, Sakaki Y, Hattori M, Omura S (2003). Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat Biotechnol* 21: 526-531.
- Jang KH, Nam SJ, Locke JB, Kauffman CA, Beatty DS, Paul LA, Fenical W (2013). Anthracimycin, a potent anthrax antibiotic from a marine-derived actinomycete. *Angew Chem Int Ed Engl* 52: 7822-7824.
- Jensen PR, Mafnas C (2006). Biogeography of the marine actinomycete *Salinispora*. *Environ Microbiol* 8: 1881-1888.
- Kersten RD, Yang YL, Xu Y, Cimermancic P, Nam SJ, Fenical W, Fischbach MA, Moore BS, Dorrestein PC (2011). A mass spectrometry-guided genome mining approach for natural product peptidogenomics. *Nat Chem Biol* 7: 794-802.
- Kersten RD, Ziemert N, Gonzalez DJ, Duggan BM, Nizet V, Dorrestein PC, Moore BS (2013). Glycogenomics as a mass spectrometry-guided genome-mining method for microbial glycosylated molecules. *P Natl Acad Sci USA* 110: 4407-4416.
- Letzel AC, Li J, Amos GCA, Millán-Aguinaga N, Ginigini J, Abdelmohsen UR, Gaudencio SP, Ziemert N, Moore BS, Jensen PR (2017). Genomic insights into specialized metabolism in the marine actinomycete *Salinispora*. *Environ Microbiol* 19: 3660-3673.
- Liaw CC, Chen PC, Shih CJ, Tseng SP, Lai YM, Hsu CH, Dorrestein PC, Yang YL (2015). Vitroprocines, new antibiotics against *Acinetobacter baumannii*, discovered from marine *Vibrio* sp. QWI-06 using mass-spectrometry-based metabolomics approach. *Sci Rep* 5: 1-11.
- Medema MH, Takano E, Breitling R (2013). Detecting sequence homology at the gene cluster level with MultiGeneBlast. *Mol Biol Evol* 30: 1218-1223.
- Miyanağa A, Janso JE, McDonald L, He M, Liu H Barbieri L, Eustáquio AS, Fielding EN, Carter GT, Jensen PR et al. (2011). Discovery and assembly-line biosynthesis of the lymphostin pyrroloquinoline alkaloid family of mTOR inhibitors in *Salinispora* bacteria. *J Am Chem Soc* 133: 13311-13313.
- Perry NB, Blunt JW, Munro MHG, Thompson AM (1990). Antiviral and antitumor agents from a New Zealand sponge, *Mycale* sp. 2. Structures and solution conformations of mycalamides A and B. *J Org Chem* 55: 223-227.
- Qin Z, Munnoch JT, Devine R, Holmes NA, Seipke RF, Wilkinson KA, Wilkinson B, Hutchings MI (2017). Formicamycins, antibacterial polyketides produced by *Streptomyces formicae* isolated from African Tetraponera plant-ants. *Chem Sci* 8: 3218-3227.
- Smith CA, O'Maille G, Want EJ, Qin C, Trauger SA, Brandon TR, Custodio DE, Abagyan R, Siuzdak G (2005). METLIN: A metabolite mass spectral database. *Ther Drug Monit* 27: 747-751.
- Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, Kaponi CA, Luzzatto-Knaan T et al. (2016). Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotech* 34: 828-837.
- Zerikly M, Challis GL (2009). Strategies for the discovery of new natural products by genome mining. *ChemBioChem* 10: 625-633.
- Ziemert N, Lechner A, Wietz M, Millan-Aguinaga N, Chavarria KL, Jensen PR (2014). Diversity and evolution of secondary metabolism in the marine actinomycete genus *Salinispora*. *P Natl Acad Sci USA* 111: E1130-1139.
- Ziemert N, Podell S, Penn K, Badger JH, Allen E (2012). The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLoS ONE* 7: e34064.