The Journal of Antibiotics (2015) 68, 767–770
© 2015 Japan Antibiotics Research Association All rights reserved 0021-8820/15
www.nature.com/ja

NOTE

Genome-based analysis of type-I polyketide synthase and nonribosomal peptide synthetase gene clusters in a novel strain taxonomically close to the genus *Salinispora*

Hisayuki Komaki[1], Tomohiko Tamura[1], Natsuko Ichikawa[2], Akio Oguchi[2], Moriyuki Hamada[1], Ken-ichiro Suzuki[1] and Nobuyuki Fujita[2]

The Journal of Antibiotics (2015) **68,** 767–770; doi:10.1038/ja.2015.69; published online 1 July 2015

Members of the genus *Salinispora* are obligate marine actinomycetes of the family *Micromonosporaceae*, and require sodium-enriched medium for growth.[1] This genus has attracted attention from researchers seeking novel secondary metabolites, because it represents promising sources of such molecules.[2,3] However, it is difficult to isolate novel strains within or closely related to the genus *Salinispora*; only three species, *Salinispora arenicola, Salinispora pacifica* and *Salinispora tropica,* have been discovered during the past decade.

In our survey of actinomycetal inhabitants of mangrove forests, we isolated a novel strain, designated NBRC 107566, from the rhizosphere of a mangrove growing on Iriomote Island, Okinawa, Japan. The 16S rRNA gene of the strain showed high sequence similarities to those of *S. arenicola* CNH-643^T (98.28%), *Micromonospora pattaloongensis* TJ2-2^T (98.08%), *S. pacifica* CNR-114^T (98.07%) and *S. tropica* CNB-440^T (98.07%). Thus, to assess the potential of NBRC 107566 as a secondary metabolite producer, we performed whole-genome shotgun sequencing and examined type-I polyketide synthase (PKS) and nonribosomal peptide synthetase (NRPS) gene clusters, which are involved in the major secondary metabolite-synthetic pathways in actinomycetes. The taxonomic study of this strain will be reported elsewhere.

The whole-genome sequence was determined by a shotgun sequencing strategy with paired-end sequencing using MiSeq (Illumina, San Diego, CA, USA; 652 Mb, 97-fold coverage). These reads were assembled using Newbler v2.6 (Roche, Basel, Switzerland) and subsequently finished using GenoFinisher,[4] which enabled a final assembly of 182 scaffold sequences of >500 bp each. The sequences have been deposited at DDBJ under accession numbers BBQH01000001–BBQH01000182. The total size of the assembly was 6 704 564 bp, with a G+C content of 72.4%. Coding sequences were predicted by Prodigal v2.6 (http://prodigal.ornl.gov/downloads.php),[5] and domains related to PKS and NRPS were searched using the SMART and PFAM domain databases. PKS and NRPS gene clusters, and their domain organizations, were identified according to a previously reported procedure.[6]

Complete gene cluster sequences were obtained for two type-I PKS (*pks2* and *pks3*) and five NRPS (*nrps1* to *nrps5*) gene clusters. We believe that one PKS gene cluster (*pks1*) was split into 16 scaffolds/contigs in the draft genome sequence. Genes encoding PKSs and NRPSs in these clusters are listed in Table 1. Although the *pks1* gene cluster could not be completely sequenced, this cluster likely contains at least 12 modules because of the presence of 12 ketosynthase domains and 12 acyl carrier protein domains. Hence, the products will be large compounds containing a polyketide chain comprising at least C_{24}. We hypothesized that the *pks2* gene cluster will synthesize an enediyne-type polyketide, because its PKS (open reading frame (Orf) 5–114) showed high sequence similarities to iterative type-I PKSs for enediyne core synthesis, termed enediyne PKS (PksE),[7] and the domain organization was the same as those of *Salinispora*, such as Sare_0551 and Strop_2697, for enediyne compounds.[2] The *pks3* gene cluster was likely novel because the PKS genes showed low sequence identities to the closest homologs. The domain organization was likely ketosynthase–acyltransferase–ketoreductase–dehydratase, which is specific to PksE, although the module split into three Orfs. Hence, the products may be enediyne compounds too. According to each module number and adenylation domain substrate predicted by antiSMASH,[8] we guessed that gene clusters *nrps1* to *nrps5* synthesize hexapeptidic compounds containing three valine molecules; small compounds derived from single amino-acid molecules; tripeptidic compounds comprising glycine, lysine and serine molecules; compounds derived from cysteine; and tetrapeptidic compounds containing a dihydroxybenzoate and two cysteine molecules, respectively.

[1]Biological Resource Center, National Institute of Technology and Evaluation, Chiba, Japan and [2]Biological Resource Center, National Institute of Technology and Evaluation, Tokyo, Japan
Correspondence: Dr H Komaki, Biological Resource Center, National Institute of Technology and Evaluation, 2-5-8 Kazusakamatari, Kisarazu, Chiba 292-0818, Japan.
E-mail: komaki-hisayuki@nite.go.jp
Received 27 February 2015; revised 3 April 2015; accepted 17 May 2015; published online 1 July 2015

Table 1 PKSs and NRPSs of type-I PKS and NRPS gene clusters in *Micromonosporaceae* strain NBRC 107566

Gene cluster	Orf	Size (aa)	Domain organization	BLAST search: Closest homolog (accession no.), origin	BLAST search: %[a]	Predicted product
pks1	16-70[b]	>1708	ACP-KS/AT/DH/KR	Hypothetical protein (WP_026242989), *Streptomyces*	64/74	Large polyketide
	319-1[b]	>541	ACP-KS	Erythronolide synthase (EWM17454), *Kutzneria* sp. 744	67/77	
	101-1[b]	>958	DH/KR/ACP	Hypothetical protein, partial (WP_030874026), *Streptomyces* sp. NRRL F-2747	49/62	
	101-2[b]	>322	KS	Hypothetical protein, partial (WP_030870332), *Streptomyces* sp. NRRL F-2747	77/86	
	105-1[b]	>1190	DH/KR/ACP-KS	Type I PKS (BAO66542), *Streptomyces* sp. MJ635-86F5	59/70	
	111-1[b]	>980	DH/KR/ACP-KS	Hypothetical protein (WP_018680914), *Actinokineospora enzanensis*	55/64	
	318-1[b]	>544	ACP-KS	Hypothetical protein (WP_026242989), *Streptomyces*	77/86	
	366-1[b]	>118	ACP-KS	PKS, partial (WP_030844246), *Streptomyces hygroscopicus*	56/68	
	119-1[b]	>822	KS/AT/DH	Erythronolide synthase (EWM17449), *Kutzneria* sp. 744	58/70	
	116-1[b]	>856	KR/ACP-KS	PKS, partial (WP_030843675), *Streptomyces hygroscopicus*	57/67	
	85-3[b]	>908	DH/KR/ACP	Modular PKS, partial (AGI61661), *Micromonospora* sp. CMS I2-32	43/52	
	85-2	1568	KS/AT/KR/ACP	**Hypothetical protein (WP_018521997), *Streptomyces***	**61/72**	
	85-1[b]	>348	KS	Hypothetical protein, partial (WP_030874000), *Streptomyces* sp. NRRL F-2747	78/88	
	82-1[b]	>2722	DH/KR/ACP-KS/AT/DH/KR/ACP	Hypothetical protein (WP_026242989), *Streptomyces*	63/72	
	82-2[b]	>515	KS	Hypothetical protein, partial (WP_031087535), *Streptomyces* sp. NRRL S-1831	68/79	
	355-1[b]	>288	KS	Hypothetical protein, partial (WP_030248090), *Streptomyces* sp. NRRL S-350	69/77	
pks2	5-114	1929	KS/AT/KR/DH	**Erythronolide synthase (WP_028678638), *Salinispora arenicola***	**69/78**	Enediyne
pks3	18-76	307	AT	**Hypothetical protein, partial (WP_031052262), *Streptomyces ochraceiscleroticus***	**35/50**	Enediyne
	18-78	664	KS	Hypothetical protein IL38_10790 (KGI81430), *Actinopolyspora erythraea*	38/53	
	18-79	799	KR/DH	Hypothetical protein IL38_23040 (KGI79490), *Actinopolyspora erythraea*	34/47	
nrps1	16-6[c]	618	A(val)/T	**Hypothetical protein (WP_032713311), *Salinispora arenicola***	**62/74**	Val-containing hexapeptide
	16-15	2182	CoL/T-C/A/T-C	**Hypothetical protein (WP_028191616), *Salinispora arenicola***	**57/66**	
	16-16	1010	A/T-C	**Hypothetical protein (WP_028191617), *Salinispora arenicola***	**59/68**	
	16-17	92	T	**Hypothetical protein (WP_028191618), *Salinispora arenicola***	**60/74**	
	16-18	2123	C/A/T-C/T-C	**Hypothetical protein (WP_028191619), *Salinispora arenicola***	**54/66**	
	16-22	523	A(val)	**Hypothetical protein, partial (WP_030883900), *Streptomyces* sp. NRRL F-5053**	**61/70**	
	16-25[c]	667	A(val)/T	**Hypothetical protein, partial (WP_030921967), *Streptosporangium amethystogenes***	**69/78**	
	16-27	544	T-C	Hypothetical protein (WP_028191619), *Salinispora arenicola*	64/75	
nrps2	8-63	1369	C/A/T-TE	**Amino acid adenylation domain protein (ACU75141), *Catenulispora acidiphila* DSM 44928**	**70/78**	Unknown
nrps3	8-101	4664	C/A(gly)/T-C/A(lys)/T-C/T-C/A(ser)/T-TE	Amino acid adenylation enzyme/thioester reductase family protein (EXG82720), *Cryptosporangium arvum* DSM 44712	68/77	Gly-Lys-Ser
nrps4	42-35	112	T	**Hypothetical protein (WP_030303910), *Streptomyces* sp. NRRL F-6131**	**56/70**	Cys-containing molecule
	42-34	452	C	Hypothetical protein Z951_28640 (EXU64848), *Streptomyces* sp. PRh5	40/55	
	42-32[c]	825	C/A(cys)	Hypothetical protein, partial (WP_030606521), *Streptomyces sclerotialus*	49/62	
	42-30	543	T-C	Hypothetical protein (WP_028191619), *Salinispora arenicola*	62/73	
nrps5	8-2	534	A(dhb)	**AMP-dependent synthetase and ligase (ACY95967), *Thermomonospora curvata* DSM 43183**	**56/68**	DHB and Cys-containing tetrapeptide
	8-5	1558	T-C/A(cys)/T	Peptide synthetase (WP_031074647), *Streptomyces* sp. NRRL WC-3742	50/59	
	8-9	1827	C/A(cys)/MT/T	Pyochelin synthetase (WP_031074637), *Streptomyces* sp. NRRL WC-3742	61/71	
	8-10	1167	A/T-Red	**Oxidoreductase (WP_033090657), *Nocardia seriolae***	**50/64**	

Abbreviations: A, adenylation; ACP, acyl carrier protein; AT, acyltransferase; C, condensation; CoL, CoA ligase; DH, dehydratase; dhb/DHB, dihydroxybenzoate; KR, ketoreductase; KS, ketosynthase; MT, methyltransferase; NRPS, nonribosomal peptide synthetase; Orf, open reading frame; PKS, polyketide synthase; Red, reductase; T, thiolation; TE, thioesterase.
Orf numbers are shown in combination with scaffold/contig numbers.
The closest homolog whose domain organization is identical to that of NBRC 107566 Orf is boldfaced. Predicted substrates of A domains are shown in brackets.
[a]Identity/similarity.
[b]Not completely sequenced.
[c]Encoded on the complementary strand.

Among the eight gene clusters, five (*pks1*, *pks3*, *nrps3*, *nrps4* and *nrps5*) were new and specific to strain NBRC 107566, because each PKS and NRPS showed low sequence similarity and distinct domain organization from its closest homolog. In contrast, homologs of the remaining three gene clusters (*pks2*, *nrps1* and *nrps2*) were present in other strains (Table 1). PksE genes such as *pks2* are present commonly in the genus *Salinispora*,[3] and also in the closely related genus *Micromonospora*.[9] In the *nrps1* gene cluster, six genes showed high sequence similarities (57–64% identities) to those of *S. arenicola*, which are specific in *S. arenicola* strain CNX891 and are not present in the other *Salinispora* strains whose genome sequences are available. However, the products of strain NBRC 107566 and *S. arenicola*

Table 2 Type-I PKS and NRPS gene clusters conserved or specific in NBRC 107566 and *Salinispora* species

Cluster	*Type*	*NBRC 107566*	*S. arenicola CNH-205*	*S. tropica CNB-440*	*S. pacifica DSM 45543*	*Actual or predicted product*[a]
1	**PKS**	***pks2***	**Sare_0551**	**Strop_2697**	**SALPAC_RS0114200**	**Enediyne**
2	**PKS/NRPS**	—	**Sare_2071-2093**	**Strop_2647-2655**	**SALPAC_RS0121725-0121735**	**Yersiniabactin-like siderophore**
3	**NRPS**	—	**Sare_4890-4895**	**Strop_4416-4420**	**SALPAC_RS0124550-0124570**	**Tetrapeptide**
4	**PKS**	—	**Sare_3282**	**Strop_3056**	—[b]	**Lymphostin**
5	**PKS/NRPS**	—	—[c]	**Strop_1022-1024**	**SALPAC_RS0116580-0116590**	**Salinisporamide**
6	**PKS**	—	**Sare_2029**	**Strop_0598**	—	**Enediyne**
7	**NRPS**	—	**Sare_0353-0363**	—	**SALPAC_RS0101640-0101690**	**Pentapeptide**
8	PKS	*pks1*	—	—	—	Large polyketide
9	PKS	*pks3*	—	—	—	Enediyne
10	NRPS	*nrps1*	—	—	—	Val-containing hexapeptide
11	NRPS	*nrps2*	—	—	—	Unknown
12	NRPS	*nrps3*	—	—	—	Gly-Lys-Ser
13	NRPS	*nrps4*	—	—	—	Cys-containing molecule
14	NRPS	*nrps5*	—	—	—	DHB and Cys-containing tetrapeptide
15	PKS	—	Sare_1246-1250	—	—	Rifamycin
16	PKS	—	Sare_2179	—	—	Enediyne
17	PKS	—	Sare_3151-3156	—	—	Macrolide
18	PKS	—	Sare_4951	—	—	Enediyne
19	NRPS	—	Sare_2948-2962	—	—	Tetrapeptide
20	NRPS	—	Sare_4562	—	—	Cyclomarin
21	PKS/NRPS	—	Sare_2407	—	—	PK-NRP hybrid
22	PKS	—	—	Strop_2768-2781	—	Salinilactam
23	NRPS	—	—	Strop_0673-0688	—	Dipeptide
24	NRPS	—	—	Strop_2806	—	Dihydroaeruginoic acid
25	NRPS	—	—	Strop_2820-2821	—	Coelibactin
26	PKS	—	—	—	SALPAC_RS0115200	Unknown
27	NRPS	—	—	—	SALPAC_RS0110150-0110175	Pentapeptide
28	NRPS	—	—	—	SALPAC_RS0111835	Unknown
29	NRPS	—	—	—	SALPAC_RS0122730	Unknown
30	PKS/NRPS	—	—	—	SALPAC_RS0111095-25875	PK-NRP hybrid
Total		8	13	10	10	

Abbreviations: DHB, dihydroxybenzoate; NRPS, nonribosomal peptide synthetase; PK-NRP, polyketide-nonribosomal peptide; PKS, polyketide synthase.
Conserved gene clusters are boldfaced.
[a]Products are from Table 1 or Nett *et al.*[10]
[b]This cluster is not present in completely genome-sequenced strain DSM 45543, but present in ~30 *S. pacifica* strains.[3]
[c]This cluster is not present in most *S. arenicola* strains including CNH-205, but present in five strains such as CNH941, CNH964, CNP105, CNP193 and CNY280.

CNX891 probably differ because the *nrps1* gene cluster encoded two extra NRPSs in addition to those found in *S. arenicola* CNX891, and some domain organizations between these two strains were different (Supplementary Figure S1). The *nrps2* cluster showed high sequence similarities to a *Catenulispora acidiphila* gene, whose products have not yet been identified, and whose domain organizations were the same, suggesting that the products of the two strains will be similar. A putative acetyl-lysine deacetylase and a lysine biosynthesis enzyme LysX were also encoded upstream of *nrps2* and the *C. acidiphila* gene cluster (data not shown). However, we could not predict the products only by genome information. To identify the products, further studies are needed.

We compared similarities and differences of type-I PKS and NRPS gene clusters among strain NBRC 107566 and three *Salinispora* species (Table 2). Orthologous gene clusters are aligned in the same row of the table. Strain NBRC 107566, *S. arenicola* CNH-205, *S. tropica* CNB-440 and *S. pacifica* DSM 45543 possess 8, 13, 10 and 10 PKS and NRPS gene clusters, respectively, suggesting NBRC 107566 has as many and various gene clusters as *Salinispora* strains. In the genus *Salinispora*, three PKS and NRPS gene clusters (iterative enediyne PKS, yersiniabactin-like siderophore NRPS and tetrapeptide NRPS, shown as clusters 1–3 in Table 2) are well conserved in all the species, and four gene clusters (clusters 4–7) are conserved at least between two species.[3,10] In contrast, strain NBRC 107566 did not possess a yersiniabactin-like siderophore NRPS gene cluster, a tetrapeptide NRPS gene cluster or the four gene clusters, but does have an iterative enediyne PKS gene cluster (*pks2*). According to the report on the ancestry of secondary metabolite gene clusters in the genus *Salinispora*, the well-conserved gene clusters are derived from the common ancestors, but only the iterative enediyne PKS gene cluster among them is shared with the closely related genus *Micromonospora*; the remaining clusters are considered to have been acquired by horizontal gene transfer at the beginning of the evolution of this genus or each species.[3] Strain NBRC 107566 is phylogenetically related to the *Salinispora* species, but its taxonomic position is outside of the clade of the genus *Salinispora* in the phylogenetic tree based on 16S rRNA gene sequences (Tamura *et al.*, unpublished). This strain may have evolved without acquiring these gene clusters, except for *pksE* (*pks2*), conserved in the genus *Salinispora*. Interestingly, no homologs of the five NBRC 107566-specific gene clusters (*pks1*, *pks3*, *nrps3*, *nrps4*, *nrps5*) were observed, even in genome-sequenced strains belonging to the related genus *Micromonospora* (Table 1). To date, only 10 *Micromonospora* strains have been genome sequenced; therefore, it is not clear whether these five clusters were transmitted vertically or

acquired by horizontal gene transfer during evolution of the family *Micromonosporaceae*. This study showed that isolation of phylogenetically novel strains, such as strain NBRC 107566, in this family could aid the search for attractive, novel and diverse secondary metabolites.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We are grateful to Ms Tomoko Hanamaki and Ms Yuko Kitahashi for assistance with the whole-genome sequencing and annotation of PKS and NRPS genes, respectively.

1 Maldonado, L. A. *et al. Salinispora arenicola* gen. nov., sp. nov., and *Salinispora tropica* sp. nov., obligate marine actinomycetes belonging to the family *Micromonosporaceae*. *Int. J. Syst. Evol. Microbiol.* **55**(Pt 5), 1759–1766 (2005).
2 Udwary, D. W. *et al.* Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*. *Proc. Natl Acad. Sci. USA* **104**, 10376–10381 (2007).
3 Ziemert, N., *et al.* Diversity and evolution of secondary metabolism in the marine actinomycete genus *Salinispora*. *Proc. Natl. Acad. Sci. USA* **111**, E1130–E1139 (2014).
4 Ohtsubo, Y., Maruyama, F., Mitsui, H., Nagata, Y. & Tsuda, M. Complete genome sequence of *Acidovorax* sp. strain KKS102, a polychlorinated-biphenyl degrader. *J. Bacteriol.* **194**, 6970–6971 (2012).
5 Hyatt, D., *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
6 Komaki, H., Ichikawa, N., Oguchi, A., Hanamaki, T. & Fujita, N. Genome-wide survey of polyketide synthase and nonribosomal peptide synthetase gene clusters in *Streptomyces turgidiscabies* NBRC 16081. *J. Gen. Appl. Microbiol.* **58**, 363–372 (2012).
7 Horsman, G. P., Van Lanen, S. G. & Shen, B. Iterative type I polyketide synthases for enediyne core biosynthesis. *Methods Enzymol.* **459**, 97–112 (2009).
8 Blin, K. *et al.* antiSMASH 2.0—a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res.* **41**, W204–W212 (2013).
9 Trujillo, M. E. *et al.* genome features of the endophytic actinobacterium *Micromonospora lupini* strain Lupac 08: on the process of adaptation to an endophytic life style? *PLoS One* **9**, e108522 (2014).
10 Nett, M., Ikeda, H. & Moore, B. S. Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Nat. Prod. Rep.* **26**, 1362–1384 (2009).

Supplementary Information accompanies the paper on The *Journal of Antibiotics* website (http://www.nature.com/ja)