

微生物学报 *Acta Microbiologica Sinica*
52 (12) : 1531 – 1539; 4 December 2012
ISSN 0001 – 6209; CN 11 – 1995/Q
<http://journals.im.ac.cn/actamicrocn>

海洋细菌 *Pseudoalteromonas* sp. NJ631 中 NRPS 基因簇及核心模块的发掘

陈威, 朱鹏*, 何山, 金海晓, 严小军

宁波大学, 应用海洋生物技术教育部重点实验室, 宁波 315211

摘要: 【目的】运用基因组信息发掘技术 (Genome Mining) 对海洋细菌 *Pseudoalteromonas* sp. NJ631 基因组中的非核糖体肽合成酶 (Nonribosomal peptides synthetases, NRPSs) 基因簇资源及其核心模块进行发掘分析, 旨在为 *Pseudoalteromonas* 属中非核糖体肽 (Nonribosomal peptides, NRPs) 的发现提供理论依据和数据支持。【方法】依托第二代测序技术获得的 *Pseudoalteromonas* sp. NJ631 基因组序列草图, 在分析其次生代谢产物编码基因的基础上, 利用 NRPS-PKS knowledgebase 在线预测软件鉴定潜在的 NRPSs 基因簇, 并对其基因组中的 NRPSs 核心模块腺苷酰化 (Adenylation, A) 结构域编码基因信息进行发掘。【结果】在 NJ631 基因组序列草图中发现 3 个典型结构组成的 NRPSs 基因簇, 命名为 NGC1、NGC2 和 NGC3, 分别位于 scaffold6, 9 和 11。进一步的结构域预测分析表明, 3 个 NRPSs 基因簇均含有 3 个 ORFs, 其中 NGC1 编码 7 个 NRPSs 模块; NGC2 和 NGC3 均编码 6 个 NRPSs 模块。对 A 结构域的信息发掘显示 NJ631 基因组中含有 38 个 A 结构域编码基因, 特异性选择 18 种氨基酸底物。【结论】通过运用基因组信息发掘技术对海洋细菌 *Pseudoalteromonas* sp. NJ631 全基因组信息进行 NRPSs 基因簇及核心模块 A 结构域的发掘分析, 结果提示, 通常只在放线菌或真菌中发现的 NRPSs 基因资源也在 *Pseudoalteromonas* 属中大量存在。研究结果也为今后 *Pseudoalteromonas* 属中非核糖体肽 (Nonribosomal peptides, NRPs) 的发现提供理论依据。

关键词: 海洋细菌, 非核糖体肽合成酶, 腺苷酰化结构域, 基因组信息发掘

中图分类号: Q933 文献标识码: A 文章编号: 0001-6209 (2012) 12-1531-09

作为海洋环境中特有的菌属, 交替假单胞菌属 (*Pseudoalteromonas*) 由于其能够分泌大量具有潜在药用价值的次生代谢产物, 如胞外酶、毒素、抗生素等^[1-2] 而成为近年来众多科研工作者的关注热点。而非核糖体肽 (Nonribosomal peptides, NRPs) 和聚酮类 (Polyketides, PKs) 则是微生物次生代谢产物中最具有药用开发价值的两大类物质, 其分别依靠微生物体内的非核糖体肽合成酶 (Nonribosomal peptides

synthetases, NRPSs) 和聚酮合成酶 (Polyketide synthetases, PKSs), 这两大酶系催化合成^[3-5]。NRPSs 和 PKSs 均为复合酶, 拥有相似的结构组分和催化合成机制^[6-8]。因此, 对交替假单胞菌属 (*Pseudoalteromonas*) 中这两大酶系基因资源的发掘利用将对该菌属中具有药用开发价值的非核糖体肽和聚酮类化合物的发现有着重要的指导意义。同时, 随着细菌基因组数据爆炸性增长以及次生代谢产物预测软

基金项目: 国家自然科学基金 (40906080); 浙江省海洋生物技术产业科技创新团队 (2012R10029)

* 通信作者。Tel: +86-574-87609572; Fax: +86-574-87600590; E-mail: zhupeng@nbu.edu.cn

作者简介: 陈威 (1987-), 男, 浙江温州人, 硕士研究生, 主要从事海洋药用微生物次生代谢功能基因研究。E-mail: weinerchen@126.com

收稿日期: 2012-06-18; 修回日期: 2012-10-07

件的出现,为 NRPSs 和 PKSs 的基因搜索以及潜在的次生代谢产物合成基因簇的发现提供了新的平台^[9-10]。

在前期的研究中,我们从珍珠膜海绵 (*Hymeniacidon perleve*) 中分离到一株具有较强抗菌活性和细胞毒活性的交替假单胞菌 *Pseudoalteromonas* sp. NJ631。通过 PCR 技术和构建 Fosmid 文库发现一条约 60kb 大小的 NRPS-PKS 杂合基因簇^[11-13]。并尝试分离该功能基因簇引导合成的次生代谢产物,但结果令人失望。分析原因,可能是在常规培养条件下,该次生代谢基因簇未能进行表达;又或者是在 NJ631 的基因组中还存在其他的 NRPS 基因簇,这些基因簇的存在和表达干扰了目标基因簇的产物检测。而对于后者,本研究以第二代测序技术获得的 *Pseudoalteromonas* sp. NJ631 基因组序列草图为基础,利用基因组信息发掘技术对其含有的 NRPS 基因簇进行深入发掘,以期了解其基因组中 NRPS 基因簇及核心结构域的分布组成,同时也为今后 *Pseudoalteromonas* 属中非核糖体肽的发现提供理论依据。

1 材料和方法

1.1 材料

1.1.1 菌株: 海洋微生物 *Pseudoalteromonas* sp. NJ631 分离于珍珠膜海绵 (*Hymeniacidon perleve*), 由宁波大学浙江省海洋生物工程重点实验室保藏。

1.1.2 主要试剂和仪器: 细菌基因组 DNA 抽提试剂盒购自生工生物工程(上海)股份有限公司; DNA 分子标准量 Marker 购自大连宝生物公司 (TaKaRa); 纯水制备系统购自美国颇尔公司 (Pall); 台式高速冷冻离心机购自德国艾本德公司 (Eppendorf); 凝胶成像仪购自上海天能科技有限公司; NanoDrop ND1000 微量分光光度计购自美国赛默飞世尔科技有限公司 (Thermo); 测序仪器 Hiseq2000 购自美国 illumine 公司。

1.2 基因组 DNA 制备

按照细菌基因组 DNA 抽提试剂盒 (BBI, Canada) 要求提取 NJ631 基因组 DNA, 并达到用于基因组测序的质量要求 (DNA 浓度为 276 ng/ μ L, 总量为 23.184 μ g)。

1.3 基因组测序与拼接

NJ631 全基因组序列测定采用鸟枪法。基因组 DNA 经超声波随机打断后,回收所需大小的 DNA 片段。利用 T4 DNA 聚合酶进行末端修复,并在末

端添加特殊接头,从而构建含约 500 bp 大小的插入片段的基因组文库。基因组测序工作由深圳华大基因研究院利用第二代测序平台 (Next Generation Sequencing, NGS) illumine Hiseq2000 (Illumina, USA) 完成,获得有效序列大小为 550 Mb。序列的组装与拼接采用华大基因研发的 SOAPdenovo 软件包^[14],从而获得 NJ631 基因组序列草图,序列中每一碱基的精确度为 99.99%。

1.4 ORF 预测与注释

用 Glimmer 3.0 基因预测软件对 NJ631 基因组序列草图进行开放阅读框 (open reading frames, ORFs) 的预测;将预测得到的 ORFs 与现有的 NR (non-redundant), COG (Clusters of Orthologous), Swiss-Prot 和 TrEMBL (translational EMBL) 的蛋白数据库进行 Blastp 比对,获得其功能信息以及近缘序列。

1.5 NRPSs 基因簇分析

在与 COG 蛋白数据库比对结果所提示的次生代谢产物基因中,利用基因组发掘的方法获得 NRPS 的相关基因。利用 DNASTAR v8.02 软件分析获得蛋白序列,通过 NRPS-PKS knowledgebase 在线网站 (<http://www.nii.res.in/nrps-pks.html>)^[15] 预测 NRPS 的结构组成与 A 结构域的底物特异性分析,并利用 Blast 软件获得 A 结构域的近缘序列。

2 结果

2.1 基因组序列草图概况

Pseudoalteromonas sp. NJ631 基因组序列草图分析表明,其基因组序列由 55 个基因框架 (scaffold) 组成,总长为 5322614 bp,其中最长片段为 970668 bp,最小片段为 510 bp;基因组的 (G + C) 含量为 43.29%,含有 4658 个 ORFs,平均长度为 1006 bp,编码基因占基因组序列的 88.05%。将 4658 个 ORFs 与提交 COGs 蛋白数据库进行功能比对,结果表明这些 ORFs 可分为 21 个功能组、未知功能组、未分类组和 COGs 无比对结果组,共 24 组 (表 1)。除大部分的与 NJ631 基本生长代谢相关 ORFs 功能组如 COG-code E、K、T 等外,仅有包含 115 个 ORFs 的功能组 COG-code Q 与其次生代谢产物合成、转运和代谢相关,这也提示我们在这些与次生代谢产物合成、转运有关的 ORFs 中或许存在 NRPS 生物合成基因簇及核心模件。

2.2 NRPS 基因簇的分析

利用 NRPS-PKS knowledgebase 在线预测软件对

表 1 基于 COG 数据库的 NJ631 中 ORFs 的功能分类检索结果

Table 1 Classification of ORFs of NJ631 based on Clusters of Orthologous of protein database

COG-code	Gene function	Gene number
A	RNA processing and modification	1
B	Chromatin structure and dynamics	3
C	Energy production and conversion	179
D	Cell cycle control, cell division, chromosome partitioning	31
E	Amino acid transport and metabolism	316
F	Nucleotide transport and metabolism	68
G	Carbohydrate transport and metabolism	153
H	Coenzyme transport and metabolism	121
I	Lipid transport and metabolism	110
J	Translation, ribosomal structure and biogenesis	200
K	Transcription	303
L	Replication, recombination and repair	175
M	Cell wall/membrane/envelope biogenesis	212
N	Cell motility	130
O	Posttranslational modification, protein turnover, chaperones	152
P	Inorganic ion transport and metabolism	214
Q	Secondary metabolites biosynthesis, transport and catabolism	115
R	General function prediction only	441
T	Signal transduction mechanisms	357
U	Intracellular trafficking, secretion, and vesicular transport	97
V	Defense mechanisms	103
S	Function unknown	298
Cog no hits	Cog no hits	390
No cogs	No cogs	489

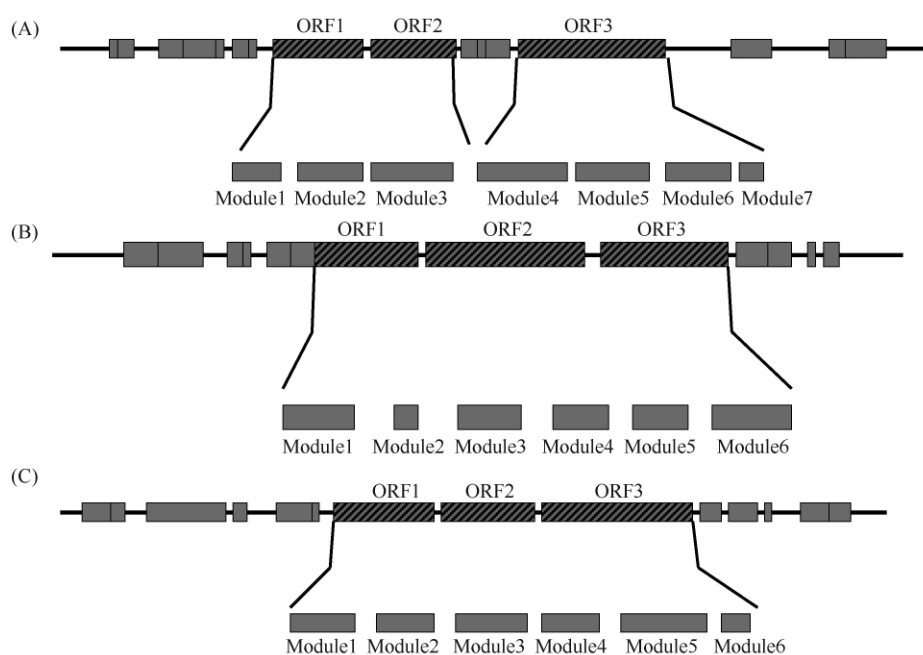


图 1 NJ631 基因组序列草图中的 NRPSs 基因簇所包含的模块组成示意图

Fig. 1 Modular organization of NRPSs gene clusters in the genome of *Pseudoalteromonas* sp. NJ631. Character A, B, C represent three NRPSs gene clusters from scaffold6, scaffold9 and scaffold11 respectively. Thin bars up the map represents partial ORFs of three scaffolds, cross-hatching indicates ORFs containing NRPSs or PKSs domains.

COG-code Q 中的 115 个次生代谢产物合成的相关 ORFs 进行 NRPS 基因簇信息发掘,结果显示,在组成 NJ631 基因组序列草图的 55 个 scaffold 中有 3 个 scaffold (scaffold6、scaffold9 和 scaffold11) 中各含有 1 个结构组成相对完整的 NRPSs 基因簇,分别命名为 NGC1、NGC2 和 NGC3 (图 1, 表 2)。NGC1 位于 scaffold6 中,为 NRPS-PKS 杂合基因簇,含有 3 个 ORFs,分别编码 2 个 NRPSs 模块、1 个 PKSs 模块和 3 个 NRPSs 模块,其中 ORF3 中含有异构化结构域 (E)。序列比对发现 NGC1 与前期通过构建 NJ631 基因组 Fosmid 文库获得的 NRPS-PKS 基因簇序列一致^[13]。NGC2 和 NGC3 分别位于 scaffold9 和 scaffold11 中,也各含有 3 个 ORFs,编码 6 个 NRPSs

模块。对于 NRPS 模块,除连续的“C A T”常规的核心三元件外,NGC1 和 NGC2 均含有用于 NRPSs 起始反应的腺苷酰化(A)和巯基化(T)结构域。而 NGC1 还含有对产物进行异构化作用的异构化结构域(E)。Pseudoalteromonas sp. NJ631 基因组中发现的 3 个 NRPSs 基因簇共含有 9 个 ORFs,编码 52 个功能结构域的 18 个 NRPSs 模块和 1 个 PKSs 模块。然而,除 NGC1 之外,NGC2 和 NGC3 包含的 NRPS 核心模块(A 结构域)与之前通过 PCR 和构建基因组文库在 NJ631 中所筛选获得的 A 结构域在蛋白序列和底物特异性上均不同,这意味着在 Pseudoalteromonas sp. NJ631 中可能存在更多 A 结构域,也促使我们对其基因组中 A 结构域分布组成进行深入发掘与分析。

表 2 菌株 NJ631 中 NRPSs 基因簇中 ORF 的功能结构域的分析

Table 2 Deduced function of ORFs domains from NRPS gene clusters of NJ631

NRPSsGene Cluster Code	Scaffold location	NRPS module	Amino acid residues	Identified domains
NGC1	6	ORF1	1733	
		Module1	~ 638	A T
		Module2	~ 1026	C A T
		ORF2	1465	
		Module3	~ 1440	KS AT KR ACP
		ORF3	4371	
		Module4	~ 1493	C A T E
		Module5	~ 1463	C A T E
		Module6	~ 1039	C A T
		Module7	~ 173	TE
NGC2	9	ORF1	1370	
		Module1	~ 1016	C A T
		Module2	~ 97	TE
		ORF2	2103	
		Module3	~ 989	C A T
		Module4	~ 990	C A T
NGC3	11	ORF3	1660	
		Module5	~ 578	A T
		Module6	~ 1031	C A T
		ORF1	1923	
		Module1	~ 1022	C A T
		Module2	~ 836	C A T
NGC3	11	ORF2	1125	
		Module3	~ 1022	C A T
		ORF3	2432	
		Module4	~ 935	C A T
Module5	~ 1037	C A T		
Module6	~ 94	TE		

C: condensation domain, A: adenylation domain, T: peptidyl carrier domain, TE: thioesterase domain, KS: ketosynthase domain, AT: acyltransferase domain, KR: ketoreductase domain, ACP: acyl carrier protein, E: epimerization domain.

2.3 A 结构域的分析

利用基因组发掘方法对 *Pseudoalteromonas* sp. NJ631 基因组中的 A 结构域编码基因进行发掘分析。结果显示, 在组成 NJ631 基因组序列草图的 55 个 scaffold 中有 10 个 scaffolds 含有 A 结构域, 总数达 38 个, 平均由 506 个氨基酸残基组成, 这是常规的 PCR 扩增筛选与基因组文库构建所无法获得的

信息量。BLAST 比对结果显示, A14 ~ A18 编码基因序列与之前从 NRPS-PKS 杂合基因簇中获得的 A 结构域基因完全相同, 而其他 A 结构域与 GenBank 中已有同源性序列的相似度均低于 60% (表 3)。但其氨基酸序列 BLAST 比对提示为 NRPS 的 A 结构域片段, 这表明在 *Pseudoalteromonas* sp. NJ631 中存在大量的 A 结构域且具有多样性和新颖性。

表 3 NJ631 中 A 结构域底物特异性预测及其相似序列

Table 3 The closest relatives and substrate specificity of A domains *Pseudoalteromonas* sp. NJ631

Gene id	Amino acid residues	Accession No.	Closest similarity	Specificity-conferring pocket	Substrate specificity prediction
A1	519	JQ782618	putative gramicidin S synthase, <i>Brevibacillus laterosporus</i> LMG 15441 (ZP_08642624), 47%	DFWNIGMVHK	Thr
A2	510	JQ782619	arthrofactin synthetase/syringopeptin synthetase-related non-ribosomal peptide synthetases, <i>Bradyrhizobium</i> sp. BTAi1 (YP_001242608), 44%	DAWFLVNVVK	leu
A3	520	JQ782620	amino acid adenylation protein, <i>Cyanothece</i> sp. PCC 7424 (YP_002381045), 42%	DVESTGAISK	Val-D
A4	496	JQ782621	amino acid adenylation domain protein, <i>Microcoleus vaginatus</i> FGP-2 (ZP_08494734), 35%	DLTKVGHVVK	Asp
A5	537	JQ782622	non-ribosomal peptide synthetases, <i>Xanthomonas albilineans</i> GPE PC73 (YP_003375559), 38%	DATKVGEVVK	Asn
A6	485	JQ782623	NosC, <i>Nostoc</i> sp. GSV224 (AAF17280), 38%	DATKVGEVVK	Asn
A7	534	JQ782624	non-ribosomal peptide synthetase, terminal component, <i>Pseudoalteromonas tunicata</i> D2 (ZP_01131645), 47%	DATKVGEVCK	Asn
A8	530	JQ782625	Amino acid adenylation, <i>Pseudoalteromonas tunicata</i> D2 (ZP_01131647), 48%	DATKVGEVCK	Asn
A9	533	JQ782626	amino acid adenylation domain-containing protein, <i>Nostoc punctiforme</i> PCC 73102 (YP_001866790), 45%	DLPRVGHIGK	Asp
A10	513	JQ782627	Peptide synthetases, <i>Cyanothece</i> sp. CCY0110 (ZP_01728834), 49%	DILQ-GMV-K	Gly
A11	515	JQ782628	amino acid adenylation, <i>Anabaena variabilis</i> ATCC 29413 (YP_320287), 46%	DGTTIAEVWK	His
A12	479	JQ782629	non-ribosomal peptide synthase, <i>Nostoc</i> sp. <i>Peltigera membranacea</i> cyanobiont (ADL59762), 58%	DVWHFSLID-	Ser
A13	536	JQ782630	amino acid adenylation domain protein, <i>Microcoleus vaginatus</i> FGP-2 (ZP_08494734), 43%	DLIKIANISK	Asp
A14	546	ABY71211	amino acid adenylation domain-containing protein, <i>Variovorax paradoxus</i> EPS (YP_004156607), 51%	none	none
A15	483	ABY71211	amino acid adenylation, <i>Nitrospira multiformis</i> ATCC 25196 (YP_412519), 46%	DAEDHGTVEK	Gln
A16	512	ABY71213	amino acid adenylation domain-containing protein, <i>Variovorax paradoxus</i> S110 (YP_002945626), 53%	TCGHLSLVDK	Ser
A17	510	ABY71213	amino acid adenylation domain-containing protein, <i>Variovorax paradoxus</i> S110 (YP_002945626), 57%	DVWHLSLVDK	Ser-D

续表 3

Gene id	Amino acid residues	Accession No.	Closest similarity	Specificity-conferring pocket	Substrate specificity prediction
A18	500	ABY71213	peptide synthase, <i>Pseudomonas aeruginosa</i> 152504 (EGM17495), 56%	DGHVYFGVTK	Bmt
A19	540	JQ782631	Gramicidin S synthetase 2, <i>Photorhabdus asymbiotica</i> subsp. <i>asymbiotica</i> ATCC 43949 (YP_003042186), 54%	DFWNIGMVHK	Thr
A20	525	JQ782632	non-ribosomal peptide synthetase module-containing protein, <i>Coxiella burnetii</i> Dugway 5J108-111 (YP_001425246), 43%	DAFWLGATFK	Val
A21	512	JQ782633	non-ribosomal peptide synthetase, <i>Salmonella enterica</i> subsp. <i>enterica</i> serovar <i>Weltevreden</i> str. 2007-60-3289-1 (CBY98966), 39%	DAFFLGCTFK	Val
A22	479	JQ782634	syringopeptin synthetase c, <i>Photorhabdus asymbiotica</i> subsp. <i>asymbiotica</i> ATCC 43949 (YP_003041057), 38%	DAESIGEVTK	Lys
A23	499	JQ782635	phenylalanine racemase, <i>Xenorhabdus bovienii</i> SS-2004 (YP_003466709), 36%	DACFLAEVTK	Leu
A24	482	JQ782636	Gramicidin S synthetase 2, <i>Photorhabdus asymbiotica</i> subsp. <i>asymbiotica</i> ATCC 43949 (YP_003042186), 41%	DAFAIAGVCK	Phe-D
A25	498	JQ782637	Gramicidin S synthetase 2, <i>Photorhabdus asymbiotica</i> subsp. <i>asymbiotica</i> ATCC 43949 (YP_003042186), 47%	DFWNIGMVHK	Thr
A26	494	JQ782638	peptide synthetases, <i>Crocospaera watsonii</i> WH 0003 (EHJ13347), 39%	DILQVGGVYK	Gly
A27	480	JQ782639	Bacitracin synthetase 1, <i>Bacillus thuringiensis</i> serovar <i>israelensis</i> ATCC 35646 (ZP_00740970), 42%	DA-DLGVVVK	Gln
A28	497	JQ782640	OciC, <i>Planktothrix agardhii</i> NIES-205 (ABW84366) 45	DAFFMGITFK	Ile
A29	484	JQ782641	non-ribosomal peptide synthetases, <i>Pectobacterium carotovorum</i> subsp. <i>brasiliensis</i> PBR1692 (ZP_03828843), 41%	DA-DIGVVVK	Glu-D
A30	527	JQ782642	non-ribosomal peptide synthetase	DAFFNGAVYK	Ile
A31	510	JQ782643	<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar <i>Weltevreden</i> str. 2007-60-3289-1 (CBY98967), 40%	DTEDICTVVK	Lys
A32	529	JQ782644	non-ribosomal peptide synthetase modules-like protein Hahella chejuensis KCTC 2396 (YP_437629), 46%	DIFHFGLILK	Hpg
A33	519	JQ782645	peptide synthetases, <i>Chromobacterium violaceum</i> ATCC 12472 (NP_902472) 47	DLYDFSLVWK	Cys
A34	508	JQ782646	amino acid adenylation domain protein, <i>Paenibacillus curdlanolyticus</i> YK9 (ZP_07387014), 53%	DILQVSLIWK	Val-D
A35	473	JQ782647	amino acid adenylation protein, <i>Herpetosiphon aurantiacus</i> DSM 785 (YP_001544632), 45%	DVGEIGSIDK	Orn-D
A36	541	JQ782648	hybrid polyketide synthase and nonribosomal peptide synthetases, <i>Lysobacter enzymogenes</i> (ABL86391), 39%	DLNGRAAICK	Ala
A37	554	JQ782649	saframycin Mx1 synthetase B, <i>Blastopirellula marina</i> DSM 3645 (ZP_01092227), 43%	IFMSDSCLFK	Val-D
A38	308	JQ782650	AMP-dependent synthetase and ligase, <i>Cyanotheca</i> sp. ATCC 51142 (YP_001805172), 48%	—DIL	Gly
			McnC, <i>Microcystis</i> sp. NIVA-CYA 172/5 (AAZ03552), 33%		

D: aspartic acid, F: phenylalanine, W: tryptophan, N: asparagine, I: isoleucine, G: glycine, M: methionine, V: valine, H: histidine, K: lysine, A: alanine, E: glutamic acid, S: serine, T: threonine, R: arginine, Q: glutamine, C: cystine, L: leucine, Y: Tyrosine, Hpg: 4-hydroxy-phenyl-glycine, Orn: ornithine, Bmt: Butenyl-4-methyl-L-threonine.

经 NRPS-PKS knowledgebase 在线预测软件对 *Pseudoalteromonas* sp. NJ631 中 A 结构域底物特异性的预测显示,除 A14 外,37 个 A 结构域均表现出对特定氨基酸的选择性。37 个 A 结构域共识别 18 种氨基酸,其中识别疏水性脂肪族氨基酸 6 种,亲水性氨基酸 10 种,芳香族氨基酸 2 种。此外,A 结构域识别的氨基酸中含有 3 种稀有氨基酸,分别是 A18 识别丁烯-4-甲基-L-苏氨酸 (Butenyl-4methyl-L-threonine, Bmt), A32 识别 4-羟基苯甘氨酸 (4-Hydroxy-Phenyl-Glycine, Hpg) 以及 A35 识别鸟氨酸 (Ornithine, Orn)。 *Pseudoalteromonas* sp. NJ631 中存在的大量 A 结构域,将极大地丰富现有的 A 结构域数据库,为后续的 A 结构域的深入研究提供重要的数据支撑。

本次研究所获得的 NRPS 基因簇和 A 结构域的核酸蛋白序列已提交 GenBank,并获得序列登录号,分别为 JQ782618-JQ782650, JX173654-JX173656, JX173679-JX173681, ABY71211-ABY71213。

3 讨论

本研究利用基因组信息发掘技术成功发现了 3 个结构组成相对完整的 NRPSs 基因簇 NGC1、NGC2 和 NGC3。其中,NGC1 序列与前期通过构建 NJ631 基因组文库所获得的 NRPS-PKS 杂合基因簇完全相同。此外,我们还发现了 NGC2 和 NGC3,暗示着 *Pseudoalteromonas* sp. NJ631 中含有多个 NRPSs 生物合成基因簇。同时也提示我们那些通常只在放线菌或真菌中发现的 NRPSs 基因资源也在 *Pseudoalteromonas* 属中存在。相对于通过构建基因组文库获得 NRPSs 基因的传统方法,基因组信息发掘技术基于先进的测序平台,优秀的在线分析软件,可以快速全面地对目的基因组中 NRPSs 生物合成基因簇进行发掘分析,极大的方便了对 NRPSs 基因资源的发现。我们的研究结果也证实了基因组信息发掘技术在发现 NRPSs 基因资源中的优越性。

目前,基于对细菌基因组中 NRPSs 基因簇的深入分析,可使我们预测其表达产物,进而指导部分该活性化合物的分离^[9,16-17]。如 Challis^[18] 利用基因组发掘方法成功发现了一种由 NRPS-like 基因簇编码的具有活性的代谢产物;Bergmann^[19] 等利用诱导方法获得沉默的 NRPS-PKS 杂合基因簇编码的代谢

产物。我们从 *Pseudoalteromonas* sp. NJ631 中发现 3 个结构组成相对完整的 NRPSs 基因簇 NGC1、NGC2 和 NGC3。根据 A 结构域的底物特异性,3 个基因簇的合成产物均为含有 5 个氨基酸的环肽化合物,其中 NGC1 编码含 PKS 的杂合环肽。环肽,是微生物产生的具有抗菌和细胞毒活性的一类重要次生代谢产物^[20]。而实验室前期获得的菌株 NJ631 表现出了较强的抗菌活性与抗肿瘤活性,两者之间可能存在着一定的联系。本实验获得的 3 个 NRPSs 基因簇,将为后续的活性产物的分离与抗菌活性肽的发现提供了可供参考的基因数据。此外,我们从 *Pseudoalteromonas* sp. NJ631 基因组中共发掘到 38 个 A 结构域,数量远超 3 个 NRPS 基因簇中包含的 A 结构域数量(14 个)。这表明在 *Pseudoalteromonas* sp. NJ631 的基因组中还含有结构不完整的 NRPSs 基因簇,如仅含 NRPSs 模块的部分结构域编码基因和由多个不完整 NRPSs 模件组成的基因簇,这些基因簇仍需要利用生物信息学方法进一步分析与鉴定。

作为 NRPSs 的核心保守模件,A 结构域对催化 NRPs 的合成具有重要的启动作用与决定作用,有着重要的研究价值。基于 *Pseudoalteromonas* sp. NJ631 基因组序列草图,通过基因组发掘的方法我们共筛选到 38 个 A 结构域,并表现出对 18 种氨基酸的底物特异选择性。大量 A 结构域基因资源的获取是常规 PCR 或基因组文库构建方法所得到的信息无法企及的,为基因组发掘方法的应用潜力提供了事实依据。目前,A 结构域基因克隆、测序和生化特点已得到广泛的研究^[21-22]。其活性位点处的核心氨基酸的改变将会导致 A 结构域活性功能的丢失或发生较大的改变。此外,Misiek^[23] 等研究发现部分 A 结构域能特异性选择一系列的底物氨基;Fewer^[24] 等指出 A 结构域的替换将导致 NRPS 催化合成另类的非核糖体肽,意味着 A 结构域的突变与重组将催生更多新型、未知的药用价值代谢产物。我们所获得的 38 个 A 结构域不仅能较好的补充 A 结构域数据库,而且为 NRPs 活性提升研究和化学组合研究提供丰富的研究素材。

4 结论

本研究首次报道了运用基因组信息发掘方法

(Genome Mining) 对海洋细菌 *Pseudoalteromonas* sp. NJ631 基因组中的 NRPSs 基因簇及核心模块 A 结构域进行全面的发掘与分析, 结果显示其基因组中含有丰富的 NRPSs 基因资源, 共发现 3 个 NRPSs 基因簇和 38 个 A 结构域编码基因。这些研究结果提示我们那些通常只在放线菌或真菌中发现的 NRPSs 基因资源也在 *Pseudoalteromonas* 属中大量存在。同时, 我们的研究结果也为今后 *Pseudoalteromonas* 属中非核糖体肽 (Nonribosomal peptides, NRPs) 的发现提供了重要的理论依据。

参考文献

- [1] Bowman JP. Bioactive compound synthetic capacity and ecological significance of marine bacterial genus *Pseudoalteromonas*. *Marine Drugs*, 2007, 5(4): 220-241.
- [2] Arthaud ID, Rodrigues FA, Jimenez PC, Montenegro RC, Angelim AL, Maciel VM, Silveira ER, Freitas HP, Sousa TS, Pessoa OD, Lotufo TM, Costa-Lotufo LV. Studies on the Secondary Metabolites of a *Pseudoalteromonas* sp. isolated from Sediments Collected at the Northeastern Coast of Brazil. *Chemistry & Biodiversity*, 2012, 9(2): 418-427.
- [3] Stachelhaus T, Marahiel MA. Modular structure of peptide synthetases revealed by dissection of the multifunctional enzyme grsA. *Journal of Biological Chemistry*, 1995, 270(11): 6163-6169.
- [4] Luo L, Burkart MD, Stachelhaus T, Walsh CT. Substrate recognition and selection by the initiation module PheATE of gramicidin S synthetase. *Journal of The American Chemical Society*, 2001, 123(45): 11208-11218.
- [5] Von-Döhren H. A survey of nonribosomal peptide synthetase (NRPS) genes in *Aspergillus nidulans*. *Fungal Genetics and Biology*, 2009, 46(Suppl 1): S45-S52.
- [6] Donadio S, Staver MJ, McAlpine JB, Swanson SJ, Katz L. Modular organization of genes required for complex polyketide biosynthesis. *Science*, 1991, 252(5006): 675-679.
- [7] Wenzel SC, Muller R. Formation of novel secondary metabolites by bacterial multimodular assembly lines: deviations from textbook biosynthetic logic. *Current Opinion in Chemical Biology*, 2005, 9(5): 447-458.
- [8] Wang SY. Advances in the study of the mechanism and application of nonribosomal peptide synthetases. *Acta Microbiologica Sinica*, 2007, 47(4): 734-737. (in Chinese)
- 王世媛. 非核糖体肽合成酶 (NRPSs) 作用机理与应用的研究进展. *微生物学报*, 2007, 47(4): 734-737.
- [9] Rounge TB, Rohrlack T, Nederbragt AJ, Kristensen T, Jakobsen KS. A genome-wide analysis of nonribosomal peptide synthetase gene clusters and their peptides in a *Planktothrix rubescens* strain. *BMC Genomics*, 2009, 10: 396.
- [10] Fehér T, Burland V, Pósfai G. In the fast lane: large-scale bacterial genome engineering. *Journal of Biotechnology*. 2012, doi: 10.1016/j.jbiotec.2012.02.012
- [11] Zheng L, Han X, Chen H, Lin W, Yan X. Marine bacteria associated with marine macroorganisms: the potential antimicrobial resources. *Annals of Microbiology*, 2005, 55(2): 119-124.
- [12] Zhu P, Zheng L, Li J, Shao JZ, Yan XJ. Screening and characterization of marine bacteria with antibacterial and cytotoxic activities, and existence of PKS I and NRPS genes in bioactive strains. *Acta Microbiologica Sinica*, 2007, 47(2): 228-234. (in Chinese)
- 朱鹏, 郑立, 林晶, 邵健忠, 严小军. 抗菌和细胞毒活性海洋细菌的筛选及其次生代谢基因证据. *微生物学报*, 2007, 47(2): 228-234.
- [13] Zhu P, Zheng Y, You Y, Yan X, Shao J. Sequencing and modular analysis of the hybrid non-ribosomal peptide synthase - polyketide synthase gene cluster from the marine sponge *Hymeniacidon perlewe*-associated bacterium *Pseudoalteromonas* sp. strain NJ631. *Canadian Journal of Microbiology*, 2009, 55(3): 219-227.
- [14] Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, Wang J. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics*, 2009, 25(15): 1966-1967.
- [15] Ansari MZ, Yadav G., Gokhale RS, Mohanty D. NRPS-PKS: a knowledge-based resource for analysis of NRPS/PKS megasynthases. *Nucleic Acids Research*, 2004, 32 (Web Server issue): w405-w413.
- [16] Schauwecker F, Pfennig F, Schröder W, Keller U. Molecular cloning of the actinomycin synthetase gene cluster from *Streptomyces chrysomallus* and functional heterologous expression of the gene encoding actinomycin synthetase II. *Journal of Bacteriology*, 1998, 180(9): 2468-2474.
- [17] Li TL, Huang F, Haydock SF, Mironenko T, Leadlay PF, Spencer JB. Biosynthetic Gene Cluster of the glycopeptide antibiotic Teicoplanin: characterization of Two Glycosyltransferases and the key Acyltransferase.

- Chemistry & Biology*, 2004, 11 (1) :107-119.
- [18] Challis GL. Mining microbial genomes for new natural products and biosynthetic pathways. *Microbiology*, 2008, 154 (Pt6) :1555-1569.
- [19] Bergmann S, Schumann, Scherlach K, Lange C, Brakhage AA, Hertweck C. Genomics-driven discovery of PKS-NRPS hybrid metabolites from *Aspergillus nidulans*. *Nature Chemical Biology*, 2007, 3 (4) :213-217.
- [20] Grunewald J, Marahiel MA. chemoenzymatic and template-directed synthesis of bioactive macrocyclic peptides. *Microbiology and Molecular Biology Reviews*, 2006, 70 (1) :121-146.
- [21] Mootz HD, Marahiel MA. The Tyrocidine biosynthesis operon of *Bacillus brevis*: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. *Journal of Bacteriology*, 1997, 179 (21) : 6843-6850.
- [22] Becker JE, Moore RE, Moore BS. Cloning, sequencing, and biochemical characterization of the nostocyclopeptide biosynthetic gene cluster: molecular basis for imine macrocyclization. *Gene*, 2004, 325: 35-42.
- [23] Misiak M, Braesel J, Hoffmeister D. Characterisation of the ArmaA adenylation domain implies a more diverse secondary metabolism in the genus *Armillaria*. *Fungal Biology*, 2011, 115 (8) :775-781.
- [24] Fewer DP, Rouhiainen L, Jokela J, Wahlsten M, Laakso K, Wang H, Sivonen K. Recurrent adenylation domain replacement in the microcystin synthetase gene cluster. *BMC Evolutionary Biology*, 2007, 7:183.

Nonribosomal peptides synthetases gene clusters and core domain in *Pseudoalteromonas* sp. NJ631

Wei Chen, Peng Zhu^{*}, Shan He, Haixiao Jin, Xiaojun Yan

Key Laboratory of Applied Marine Biotechnology, (Ningbo University), Ministry of Education, Ningbo, Zhejiang 315211, China

Abstract: [Objective] We studied nonribosomal peptides synthetases (NRPSs) gene clusters and the core module of NRPSs in *Pseudoalteromonas* sp. NJ631 using genome mining approach. [Methods] The genome of *Pseudoalteromonas* sp. NJ631 was constructed by the next genome sequencing (NGS) technology. We adopted an online available software called NRPS-PKS knowledgebase to identify potential NRPSs gene clusters within genes involved in the biosynthesis of secondary metabolite of *Pseudoalteromonas* sp. NJ631. The genes encoding adenylation (A) domains, the core module of NRPSs, were collected and analyzed using genome mining method. [Results] We identified three typical NRPS gene clusters comprising three ORFs which encode six continuous modular NRPSs. The result of genome mining indicates that genome of *Pseudoalteromonas* sp. NJ631 contains 38 A domain genes which show 60% similarity below to their closest relatives. The substrate of these A domains was predicted to specifically bind 18 types of amino acids using the specificity-conferring selection rule. [Conclusion] This is the first reported on the systematic screening and analysis of NRPSs gene clusters and A domains in genus *Pseudoalteromonas*, suggesting that the genus *Pseudoalteromonas* possesses a vast array of secondary metabolite biosynthesis genes that were previously found mostly in actinomycetes and fungi. The information on secondary metabolite genes from *Pseudoalteromonas* sp. NJ631 will facilitate us to isolate novel nonribosomal peptides.

Keywords: marine bacterium, nonribosomal peptides synthetases, adenylation domain, genome mining

(本文责编:王晋芳)

Supported by the National Natural Science Foundation of China (40906080) and by the Zhejiang Marine Biotechnology Innovation Team (ZMBIT) of China (2012R10029-4)

^{*} Corresponding author. Tel: +86-574-87609572; Fax: +86-574-87600590; E-mail: zhupeng@nbu.edu.cn

Received: 18 June 2012/Revised: 7 October 2012