Secondary metabolite gene evolution in three closely related marine actinomycete species

Kelle C. Freel, Sang-Jip Nam, William Fenical, and Paul R. Jensen

Scripps Institution of Oceanography
University of California San Diego
La Jolla, California USA 92093
pjensen@ucsd.edu
858-534-7322

Abstract

The marine actinomycete genus *Salinispora* is composed of three closely related species. These bacteria are a rich source of secondary metabolites, which are produced in species-specific patterns. This study examines the distribution and phylogenetic relationships of genes involved in the biosynthesis of secondary metabolites in the salinosporamide and staurosporine classes, which have been reported from *S. tropica* and *S. arenicola*, respectively. The focus is on “*S. pacifica*”, the most recently discovered and phylogenetically diverse member of the genus. Of 61 “*S. pacifica*” strains examined, 15 tested positive for a ketosynthase domain linked to the biosynthesis of salinosporamide K, a new compound in the salinosporamide series. Compound production was confirmed in two strains and the domain phylogeny supports vertical inheritance.
from a common ancestor shared with S. tropica, which produces related
compounds in the salinosporamide series. There was no evidence for
interspecies recombination among salA KS sequences providing further support
for the geographic isolation of these two salinosporamide producing lineages. In
addition, staurosporine production is reported for the first time in “S. pacifica”,
with 24 of 61 strains testing positive for staD, a key gene involved in the
biosynthesis of this compound. High levels of recombination were observed
between staD alleles in “S. pacifica” and the co-occurring yet more distantly
related S. arenicola, which produces a similar series of staurosporines. The
distributions and phylogenies of the biosynthetic genes examined provide insight
into the complex processes driving the evolution of secondary metabolism
among closely related bacterial species.

Introduction

Microbial secondary metabolism is the source of many of today’s most
useful pharmaceutical agents. Bacteria in the order Actinomycetales have been
particularly important in this regard, accounting for over 50% of the bioactive
microbial compounds discovered as of 2002 (1). Studies of bacterial secondary
metabolism have largely targeted the discovery of new compounds and the
mechanisms of their biosynthesis. As a result, we know relatively little about the
ecological functions of secondary metabolites or the evolutionary histories of the
biosynthetic genes responsible for their production. In cases where functions
have been addressed, it is clear that secondary metabolites can act as
allelochemicals, signaling molecules, and siderophores (3), while it has been suggested that suites of compounds can work synergistically against competitors (2). The evolutionary histories of the associated biosynthetic genes are proving to be equally complex (17) yet are providing new opportunities to incorporate phylogenetics into the discovery process (14).

Secondary metabolites are generally produced by large gene collectives that can exceed 100 kb and include genes involved in regulation, resistance, and transport (4, 11). The horizontal exchange of genes in these pathways is well documented (18, 21, 23, 26) and provides a rapid mechanism for bacteria to test the selective advantage afforded by the small molecule product(s) of complex biosynthetic pathways (12, 23). Documenting the distributions of specific biosynthetic pathways among bacteria has the potential to add new insight into the extent to which these pathways are exchanged and how gene clusters evolve to create new chemical diversity.

We have been studying the ecology and secondary chemistry of the marine actinomycete genus *Salinispora*. To date *S. tropica* and *S. arenicola* have been formally described (24) while a third species, “*S. pacifica*”, has been proposed (19). The three species share 99% 16S rRNA gene sequence identity (19) and thus are at the limits of resolution attainable with this taxonomic marker. They also display different geographic distributions with reports to date indicating that *S. tropica* is restricted to the Caribbean, “*S. pacifica*” occurs worldwide except for the Caribbean, and *S. arenicola* is broadly distributed and co-occurs with both species (19).
The genus *Salinispora* has proven to be a robust source of secondary metabolites, which represent the largest class of functional traits differentiating the species (30). These differences include the consistent production of compounds in the salinosporamide class by *S. tropica* and the staurosporine class by *S. arenicola* (20). Among these compounds, salinosporamide A is a potent proteasome inhibitor that is currently advancing through clinical trials as an anticancer agent (10). Salinosporamide A was originally discovered from *S. tropica* strain CNB-392 (9), which subsequently yielded additional compounds in this series (9, 36). The staurosporines are a well known class of protein kinase inhibitors (34) that were originally discovered from a *Streptomyces* sp. (13, 34).

The consistent production of specific classes of secondary metabolites by different *Salinispora* species was used to support the ecological importance of secondary metabolism and to link this functional trait to unresolved ecological differences among the species. Species-specific secondary metabolite production has also been documented in fungi and used to suggest that this trait may be used as a taxonomic marker (22).

Mining of the draft genome sequence of “*S. pacifica*” strain CNT-133 led to the recent discovery of salinosporamide K (7), a new compound in the salinosporamide series. Salinosporamide K lacks the chloro-ethyl side chain at the C-2 position of the salinosporamide bicyclic ring system. As expected, the associated salinosporamide biosynthetic pathway (*Sp_sal*) characterized from this strain lacked the *salL* chlorinase (8) and associated genes responsible for the creation of the unique PKS extender unit observed in the *S. tropica St_sal*
pathway (6). The exact replacement of these genes with transposases in “S. pacifica” strain CNT-133, coupled with the conservation of the remainder of the biosynthetic genes in the two pathways (7), suggests gene loss in “S. pacifica” as opposed to gene gain in S. tropica. However, it is not clear if this genetic difference is a consistent feature of “S. pacifica”. In addition, preliminary chemical studies of “S. pacifica” provided the first evidence of staurosporine production in this species. These observations, coupled with the isolation of the salinosporamide-related cinnabaramide series in a Streptomyces strain (32) and the detection of the staurosporine gene cluster on a Streptomyces giant linear plasmid (25), raised new questions about the distributions and evolutionary histories of the associated biosynthetic pathways in Salinispora species.

The aims of this study were to examine the distributions and phylogenies of genes involved in the biosynthesis of compounds in the salinosporamide and staurosporine classes among Salinispora species. The resulting data were used to assess the roles of vertical inheritance, recombination, and horizontal gene transfer in the evolutionary histories of these biosynthetic genes. The results provide evidence of the complex processes driving the evolution of secondary metabolism in three closely related bacterial species.

Materials and Methods

Strains and nucleic acid extraction

The 61 “S. pacifica” strains used in this study were cultured from marine sediment samples and identified based on 16S rRNA gene sequencing as
previously described (7, 19). In addition, six *S. arenicola* and four *S. tropica* strains were also included in the analyses. Genomic DNA was extracted according to the DNeasy protocol (QIAGEN INC., Valencia, CA) with previously described changes (15) and used immediately or stored at -20°C.

**Gene amplification and sequencing**

Specific and degenerate primer sets were designed based on the KS domains identified in the salinosporamide polyketide synthase gene *salA* in *S. tropica* (CNB-440) and “*S. pacifica*” (CNT-133) (7, 35). The specific primer set *salAks2F* (5'-GCGGAAATCGACGATACGT-3') and *salAks2R* (5'-TCCACATAGTCTACGAGCCA-3') targeted approximately 700 base pairs while the degenerate primer set *salAks3F* (5'-CATMGCRCCCGGYARCCTCG-3') and *salAks3R* (5'-TYCACRTAGCTRCGASCCA-3') targeted approximately 750 base pairs. Each PCR reaction consisted of a 50 μl mixture containing 10X PCR buffer (Applied Biosciences), 2.5 mM MgCl₂ (Applied Biosciences), 0.7% dimethyl sulfoxide (DMSO), 10 mM dNTP’s, 1.5 U of AmpliTaq Gold DNA polymerase (Applied Biosciences, Foster City, CA) and 18.75 μmol of each primer. The program for the PCR reaction included a primary denaturation step at 95°C for 15 minutes followed by 30 cycles of 95°C for 1 minute, 60°C for 1 minute (55°C for the degenerate primers), and 72°C for 1 minute, followed by a final extension at 72°C for 7 minutes. “*S. pacifica*” strain CNT-133 was used as a positive control while *S. arenicola* strain CNS-205, which does not possess the *sal* pathway (30), was used as a negative control for all PCR reactions. PCR
products were purified using the Zymo DNA clean and concentrator purification kit (Zymo Research Incorporated, Irvine CA) and sequenced using the salAks3F primer at SeqXcel, Sorrento Valley CA. Previously published primers targeting the salL chlorinase (8), which is involved in the biosynthesis of salinosporamide A, were tested on all strains that yielded a sequence-verified salA KS product. The PCR conditions were as described above (annealing temperature 55°C).

The specific primer set staD2F 5'-TGTGGGGSCACTACAACGA-3' and staD1R 5'-SGGRTCGCACATCTGCCAGAT-3' was designed based on an alignment that included staD sequences from S. arenicola strain CNS-205 (NC_009953), two Streptomyces spp. (AB071406, AB088119), two S. arenicola draft genomes, and homologs of rebD (AB090952) and vioB (GQ266676, AF172851). The PCR reagents were the same as those listed above with the addition of 0.5 μL of bovine serum albumin (New England BioLabs, Inc., Beverly, MA). A touchdown PCR that consisted of an initial soak at 94°C for 12 minutes followed by 3 cycles of 94°C for 1 minute, 67°C for 1 minute, and 72°C for 1 minute followed by 27 additional cycles of 94°C for 1 minute, 62°C for 1 minute, and 72°C for 1 minute, with a final extension at 72°C for 10 minutes was used. Appropriately sized PCR products were purified and sequenced as described above using the staD2F forward primer.

Phylogenetic analyses

Nucleotide sequences were edited using the Sequencher software package (version 4.5; Gene Codes Co., Ann Arbor, MI), aligned using Muscle
(5), and visually edited using MacClade (version 4.07; Dave and Wayne Maddison, Sinauer Associates, Inc., Sunderland, MA). Maximum likelihood, maximum parsimony, and neighbor joining phylogenetic trees were created using PAUP (33) and Phyml (maximum likelihood methods) (16). Bootstraps were calculated using 1000 replicates and resampling at all sites.

Fermentation/extraction and LC-MS analyses

Initial LC-MS analyses of five “S. pacifica” strains that tested positive for the salA KS did not reveal the presence of salinosporamide K. Of these, the strain with the best growth (CNS-863) was selected for more detailed studies along with strain CNT-133 (positive control), the original source of this compound. Both strains were cultured in 2.8 L Fernbach flasks containing 1 L of medium A1BFe+C (10 g starch, 4 g yeast extract, 2 g peptone, 1 g CaCO₃, 40 mg Fe₂(SO₄)₃•4H₂O, 100 mg KBr, 1 L seawater) and shaken at 230 rpm and 27°C. Autoclaved XAD-7 resin (20 g) was added to the culture after 24 hours. After six days, the resin was collected by filtration through cheesecloth, washed with deionized water, and eluted with acetone. The acetone was removed under reduced pressure and the resulting aqueous layer extracted with ethyl acetate (2 × 300 mL). The ethyl acetate soluble fraction was dried under vacuum to obtain a crude extract, which was fractionated by silica gel flash chromatography eluting with increasing amounts of acetonitrile (CH₃CN) in dichloromethane (CH₂Cl₂) (100% CH₂Cl₂, 100:1, 50:1, 20:1, 5:1, 1:1, and 100% CH₃CN). All fractions were subjected to LC-MS analysis using a Hewlett-Packard series 1100 LC-MS
system with a reversed-phase C18 column (Phenomenex Luna, 4.6 mm × 100 mm, pore size 5 μm) using a solvent gradient from 5% to 100% CH3CN over 23 min, a flow rate of 0.7 ml/min, and UV detection. Salinosporamide K production was determined by retention time and comparison of UV and mass spectral data with an authentic standard. Low-resolution mass data were obtained in the positive mode (ESI voltage 6.0 kV, capillary temperature 200°C, auxiliary and sheath gas pressure 5 units and 70 lb/in², respectively).

For staurosporine production, “S. pacifica” strain CNS-863 and S. arenicola strain CNS-205 (positive control) were grown in 25 ml of medium A1 (10 g yeast, 4 g yeast extract, 2 g peptone, 1 L seawater) in 125-ml Erlenmeyer flasks for 3 to 7 days prior to transfer to 1 L of medium A1BFe+C in 2.8 L Fernbach flasks. All cultures were grown with shaking at 230 rpm at 25-27°C for six days then extracted once with 1 L of ethyl acetate. The organic layers were separated, concentrated to dryness under vacuum, and fractionated by silica gel flash chromatography eluting with increasing amounts of methanol (CH3OH) in dichloromethane (CH2Cl2) (100% CH2Cl2, 100:1, 50:1, 20:1, 5:1, 1:1, and 100% CH3OH). The fractions containing staurosporines were combined and analyzed by LC-MS as above (4.6 × 250 mm column, UV = 292 nm) using an isocratic solvent system of 35% CH3CN in water with 0.1% trifluoroacetic acid. Compounds were identified as staurosporines by comparison with data from Antibase (Wiley-VCH Verlag GmbH & Co) and standards stored in an internal database.
Results

salA KS distributions and phylogeny

Sixty-one “S. pacifica” isolates derived from 34 independent marine sediment samples, collected from six geographically distinct sites, were screened by PCR for sequences related to the KS domains in the salA polyketide synthase genes in S. tropica (35) and “S. pacifica” (7). This gene is associated with the biosynthesis of salinosporamide A in S. tropica (6) and salinosporamide K in “S. pacifica” (7). The homolog cinA was recently identified in Streptomyces cinnabarinus as part of the pathway responsible for the biosynthesis of the salinosporamide-related cinnabaramide series in (31). In total, the salA KS was detected in 15 of 61 “S. pacifica” strains including five of seven 16S rRNA gene sequence types originating from three of the six locations sampled (Guam, Palau, and Fiji; table 1). The same 15 strains tested positive using both specific and degenerate PCR primers. SalA KS sequences were also obtained from three additional S. tropica strains derived from the Bahamas to complement that originally reported from strain CNB-440 (35). The “S. pacifica” KS sequences share 84-85% nucleotide identity with the S. tropica sequences and 99-100% nucleotide identity with each other. The key amino acids VDTACSSSLVAVHLACQS involved in forming the binding pocket of the KS domain (28) are conserved in all of the sequences.

The 16S rRNA tree (figure 1) depicts the previously reported relationship of S. tropica and “S. pacifica” as sister taxa that share a common ancestor with S. arenicola (Freel et al., submitted as supplementary material for review). The
seven “S. pacifica” 16S sequence types included in this study (figure 1) represent more than half of the 13 reported to date for this species (Freel et al., submitted as supplementary material for review). Phylogenetic analyses of the Salinispora salA KS sequences reveal two well-supported lineages that are congruent with the 16S tree (figure 2). There is no evidence for interspecies recombination at this locus. However, the detection of clonal KS sequences among different “S. pacifica” 16S sequences types (eg., strains CNR-551, CNS-251, CNS-799, and CNS-863) reveals considerable levels of intraspecific recombination. In addition, the KS sequence from the recently characterized cinnabaramide biosynthetic pathway (31) displays a close phylogenetic relationship with the Salinispora spp. KS sequences (75-78% nucleotide identity) despite originating from Streptomyces cinnabarinus, which belongs to a separate family in the Actinomycetales. KS sequences that share >68% nucleotide identity to those observed in “S. pacifica” have also been reported from taxonomically diverse actinomycetes and myxobacteria (figure 2) and depict a phylogeny that is highly incongruent with the taxonomic relationships of these organisms. These results provide strong evidence that the KS sequences associated with the biosynthesis of compounds in the salinosporamide class have been subjected to horizontal gene transfer (HGT). SalA KS sequences were not detected in one complete (30) and two draft (unpublished data) S. arenicola genome sequences. In addition, compounds in the salinosporamide class were not detected among 30 S. arenicola strains examined in a prior study (20). Taken together, these results provide strong evidence that the sal pathway was acquired prior to the S. tropica
and "S. pacifica" speciation event and subsequently evolved independently in these two lineages.

**Staurosporine staD distributions and phylogeny**

Twenty-four of the 61 "S. pacifica" strains examined, including five of seven 16S sequence types, generated sequence-verified PCR products using primers specific for the dichloro chromopyrrolic acid synthetase gene staD (table 1), which is involved in the biosynthesis of the staurosporine aglycone (29). The six *S. arenicola* strains tested also yielded sequence verified staD products as expected given that this species is known to consistently produce compounds in this class (20). The *S. arenicola* sequences complement those originally observed in the CNS-205 genome (accession number NC_009953) and subsequently in a draft genome sequence of *S. arenicola* strain CNH-643. None of the four *S. tropica* strains tested yielded a product with the staD primer set employed nor was this gene observed in the genome sequence of *S. tropica* strain CNB-440 (accession number NC_009380).

The phylogeny of the *S. arenicola* and "S. pacifica" staD sequences reveals two well-supported clades, as would be expected if they had been inherited from a common ancestor and subsequently evolved independently in the two lineages. Unlike the salA KS sequences however, the staD sequences observed in "S. pacifica" display considerable diversity. More importantly, seven examples of interspecies recombination are observed out of 32 sequences examined (22%). In each case (eg, strains CNT-603 and CNT-131), an "S.
"pacific" 16S sequence type clades within the S. arenicola staD lineage and, in six of seven cases, maintains an allele that is also observed in S. arenicola. There were no examples where a strain identified as S. arenicola based on 16S sequence maintained an "S. pascifica" staD allele or claded within the "S. pascifica" lineage. The previously documented co-occurrence of S. arenicola and "S. pascifica" (19) provides spatial opportunities for these interspecies recombination events to occur. As in the salA KS tree, there are multiple examples where different "S. pascifica" 16S sequence types maintain the same staD allele (eg, strains CNT-124 and CNT-150) providing evidence that intraspecific recombination has also occurred at this locus. The remarkably high level of staD and 16S rRNA gene sequence conservation within S. arenicola (>99% at both loci) relative to "S. pascifica" provides evidence of extensive recombination or a recent period selection event in this species.

LC-MS screening

Among the 15 "S. pascifica" strains that tested positive for the salA KS sequence, salinosporamide K production was confirmed in strain CNS-863 (figure 4) in addition to strain CNT-133, the original source of this compound (7). Due to low yields, fractionation of the crude extract was required before compound production could be confirmed based on retention time, UV spectrum, and mass, all of which matched an authentic standard. The production of salinosporamide A was not observed in either strain. Probing all 15 KS positive "S. pascifica" strains failed to detect salL, the chlorinase associated with the
biosynthesis of the chlorinated extender unit in salinosporamide A (6, 8). This result provides additional support for the occurrence of the salinosporamide K, as opposed to the salinosporamide A, pathway in the 15 “S. pacifica” strains.

Of the 24 “S. pacifica” isolates that tested positive for staD, an organic extract from strain CNS-863 was analyzed by LC-MS to confirm the presence of compounds in the staurosporine class. This extract contained three major peaks (figure 5) with masses that correspond to various staurosporine analogs (supplemental figure 1). Although peak 1 is not clearly visible in the CNS-863 trace, the UV and mass data confirm the presence of the same compounds that where observed in strain CNS-205. Thus, both strains appeared to produce the same staurosporines yet in different relative amounts. The UV characteristics of all of these compounds were identical and matched with 99% identity to a staurosporine standard (figure 5). Because each peak is associated with as many as five staurosporine analogues that possess identical mass and UV properties, it was not possible to assign precise structures to the compounds produced by strain CNS-863 using the methods employed.

Discussion

Secondary metabolites are non-essential metabolic products believed to impart significant positive effects on the fitness and ecology of the bacterial populations that produce them. These compounds are the products of large gene collectives that are subject to horizontal gene transfer (12, 18) and whose distributions among closely related bacterial populations remain largely unknown.
Here we report the distributions and phylogenetic relationships of genes involved in the biosynthesis of two classes of biologically active secondary metabolites in the marine actinomycete genus *Salinispora*. The results provide evidence for both HGT and vertical inheritance and a glimpse into the complexity of secondary metabolite evolution in closely related bacterial species.

The genus *Salinispora* has proven to be an interesting model with which to address questions about secondary metabolite production and its relationship to species-level taxonomic assignments. The recent discovery of salinosporamide K from "*S. pacifica*" strain CNT-133 (7) was unexpected considering that compounds in this class had previously been reported exclusively from *S. tropica* (20). Given that *S. tropica* and "*S. pacifica*" are sister taxa, this observation provided the opportunity to test the hypothesis that the associated biosynthetic pathway was inherited from a common ancestor and subsequently evolved independently in the two species. The occurrence of the salA KS in a majority of the "*S. pacifica*" 16S rRNA sequence types, coupled with the congruence of the 16S (figure 1) and KS (figure 2) trees, provides support for this hypothesis. The detection of salinosporamide K in "*S. pacifica*" strain CNS-863 provides a link between the salA KS sequence and the presence of the complete *Sp_sal* biosynthetic pathway. It also provides additional support for the lineage specificity of salinosporamides A and K in *S. tropica* and "*S. pacifica*", respectively. Further support for lineage specificity comes from a prior study in which salinosporamide A production was observed in six *S. tropica* strains but not in 41 "*S. pacifica*" and *S. arenicola* strains (20). Finally, the salL chlorinase,
which is essential for the production of salinosporamide A in S. tropica (6), was not detected in any of the 15 KS positive “S. pacifica” strains. These results support the hypothesis that the “S. pacifica” lineage maintains a salinosporamide pathway that is devoid of the genes associated with the biosynthesis of the ethyl chloride moiety observed in salinosporamide A. Although it can not be determined if these genes were lost in “S. pacifica” or gained in S. tropica, the presence of transposases in the “S. pacifica” Sp_sal pathway at the precise locations where the genes responsible for the biosynthesis of the PKS extender unit occur in the S. tropica St_sal pathway (7) provides support for gene loss in “S. pacifica”.

Despite evidence for the divergence of the salinosporamide A and K pathways in S. tropica and “S. pacifica” respectively, salinosporamide K and the associated KS sequence proved difficult to detect in many of the “S. pacifica” strains examined. Conversely, salinosporamide A and the associated KS sequence has been consistently observed in S. tropica. A number of possible explanations could account for these differences including failure to access KS templates (14) or differences in the regulatory mechanisms of the two pathways. Another interesting possibility is that the sporadic distribution of the salinosporamide K pathway is linked to the decreased cytotoxic potency of this compound (7) and any associated loss in fitness advantage it confers relative to salinosporamide A. At present, however, the ecological functions of these compounds remain undefined and therefore any links between cytotoxic potency and selective advantage remain highly speculative (7).
The lack of recombination between the *S. tropica* and “*S. pacifica*” salA KS sequences provides an unexpected line of support for both culture-dependent (Freel et al., submitted as supplementary material for review) and culture-independent (27) reports indicating that these two species do not co-occur. The basal position of the recently characterized and closely related KS sequence from the cinnabaramide biosynthetic pathway (figure 2) provides intriguing evidence that the sal pathway was acquired prior to the divergence of *S. tropica* and “*S. pacifica*” and that these pathways have subsequently maintained independent evolutionary trajectories in the two *Salinispora* lineages. Regardless of the evolutionary history of this pathway, the results provide support for a link between geographic isolation and the diversification of secondary metabolite biosynthesis in two closely related species. Based on these results, it can be proposed that the successful cultivation of new *Salinispora* lineages may lead to the discovery of new chemical diversity in the salinosporamide series.

The detection of compounds in the staurosporine class among strains of “*S. pacifica*” was also surprising given that these compounds had previously been observed exclusively in *S. arenicola* (20). An analysis of the staD genes associated with staurosporine biosynthesis in the two *Salinispora* species reveals two distinct lineages, as would be expected of a pathway inherited from a common ancestor. However, incongruence with the 16S tree includes seven recombination events, which suggests considerable allelic exchange among these two co-occurring species (19). These recombination events were recorded
between strains that originated from distant locations, e.g., *S. arenicola* CNS-673 from Fiji and *S. pacifica* CNS-055 from Palau, providing insight into the geographic scales at which these events can be observed. However, given that only one of 14 genes in the staurosporine biosynthetic pathway was examined (29), it is not possible to determine the extent to which these recombination events affected the entire operon.

The close relationship of the *staD* gene sequence in a *Streptomyces* sp. and those observed in both *Salinispora* spp. provides strong evidence that this pathway has been exchanged horizontally, however it is not clear at what point in the *Salinispora* phylogeny it may have been acquired. The *staD* gene was not detected in the *S. tropica* CNB-440 genome (35), in three draft genomes of the closely related genus *Micromonospora*, or in the recently released and closely related *Verrucosispora maris* genome (accession number NC_015434). Staurosporine was also not observed in previous studies of *S. tropica* secondary metabolism (20). These observations suggest that if *staD* was acquired prior to the divergence of *S. arenicola*, it was subsequently lost in *S. tropica*. Alternatively, it may have been acquired by *S. arenicola* and then horizontally transferred to “*S. pacifica*”. Regardless of the evolutionary history of this pathway, both lineages produce the same four staurosporine analogs (by LC-MS), indicating that the differences observed in the *staD* sequences (figure 3) do not appear to be linked to the production of different compounds.

The persistence of pathways acquired by HGT is due to the selective advantage their small molecule products confer to the host (12, 23). The
incongruence of HGT and species-specific secondary metabolite production suggests a complex interplay between gene acquisition and natural selection that creates mechanisms for the generation of new structural diversity and the fixation of adaptive products via periodic selection. Inferring the evolutionary histories of the biosynthetic pathways associated with secondary metabolism remains complex but provides opportunities to understand how nature creates new structural diversity and the extent to which this diversity is linked to specific taxonomic groups. The three *Salinispora* species provide a well-defined model system within which to assess secondary metabolite gene evolution. The results from the present study reveal considerable levels of genetic exchange and clues to the mechanisms of secondary metabolism evolution in closely related taxa.

**Figure legend**

**Figure 1.** 16S rRNA gene phylogeny. Neighbor joining phylogenetic tree based on 512 nucleotide positions from strains that yielded sequence verified salA KS or staD PCR products. Species names are followed by strain identifier, source location, 16S sequence type ("S. pacifica" only, grey scale boxes, A-F with original sequence type not labeled), and accession number (in parentheses). Bootstrap values >60% for neighbor-joining and maximum parsimony trees are shown for 1000 replicates at the respective nodes. *Micromonospora globosa* was used to root the tree.

**Figure 2.** SalA KS phylogeny. Neighbor joining phylogenetic tree based on 625 nucleotide positions. Species names are followed by strain identifier, 16S
sequence type (“S. pacifica” only, grey scale boxes, A-F with original sequence type not labeled), and accession number (in parentheses). Structures of compounds produced by *S. tropica*, “*S. pacifica*”, and *Streptomyces cinnabarinus* are shown next to their respective lineages with differences in the C-2 substitution patterns indicated in grey. Confirmed producing strains are in bold. Bootstrap values >60% for neighbor-joining and maximum parsimony trees are shown for 1000 replicates at the respective nodes. A KS sequence from *C. croactus* was used to root the tree.

**Figure 3.** StaD phylogeny. Neighbor joining phylogenetic tree based on 741 nucleotide positions. Species names are followed by strain identifier, 16S sequence type (“*S. pacifica*” only, grey scale boxes, A-F with original sequence type not labeled), and accession number (in parentheses). Bootstrap values >60% for neighbor joining, maximum likelihood, and maximum parsimony trees are shown for 1000 replicates at the respective nodes. A homolog associated with rebeccamycin biosynthesis (*rebD*) was used to root the tree.

**Figure 4.** LC-MS data supporting salinosporamide K production. A) LC trace for authentic salinosporamide K, a chromatography fraction obtained from a culture extract of strain CNS-863 co-injected with salinosporamide K, and the CNS-863 fraction alone. B) UV trace of salinosporamide K and the peak observed at 9.14 min in strain CNS-863. C) Low resolution mass data of authentic salinosporamide K and the 9.14 minute peak.

**Figure 5.** LC-MS data supporting staurosporine production. A) LC trace for “*S. pacifica*” strain CNS-863 and *S. arenicola* CNS-205 (positive control) with
detection at the UV maximum for staurosporines (292 nm). The similar retention
times of the four peaks in both strains indicate the production of the same
compounds. B) UV data for peaks 1-4 in strain CNS-863. The UV spectrum for
peak 4 is overlaid with that of a staurosporine standard with a mass of 480. MS
data for peaks 1-4 in strain CNS-863. All of these masses correspond to
previously reported staurosporines. Similar UV and mass data were recorded for
CNS-205.

Acknowledgements

This research is the result of financial support from the National Institutes of
Health (NIH), Fogerty Center International Cooperative Biodiversity Groups
program (grant U01-TW007401-01 to PRJ) and NIH grant 1R01GM086261-O1
to PRJ). We thank A. Lechner for the salL chlorinase primers and the people of
Fiji for their hospitality and permission to collect samples in their territorial waters.
We also thank W. Aalbersberg (University of the South Pacific) for providing
laboratory space and facilitating field collections.

References


driving forces for the evolution of multiple secondary metabolite production


proteasome inhibitor from a novel microbial source, a marine bacterium of

10. Fenical, W., P. R. Jensen, M. A. Palladino, K. S. Lam, G. K. Lloyd, and
    B. C. Potts. 2009. Discovery and development of the anticancer agent

    polyketide and nonribosomal peptide antibiotics: logic, machinery, and


    Omura. 1978. X-Ray crystal structure of staurosporine: a new alkaloid
    from a Streptomyces strain. Journal of the Chemical Society, Chemical
    Communications:800-801.

    Sequence-based analysis of secondary-metabolite biosynthesis in marine

    diversity of gram-positive bacteria cultured from marine sediments. Appl.
    Environ. Microbiol. 73:3272-3282.


Figure 2

Chondromyces croactus (AM179409)

Micromospora echinospora (FJ915123)

Streptomyces avermitilis (AB070940)

Streptomyces venezuelae (AF079138)

Sorangium cellulosum (GQ981380)

Streptomyces cinnabarinus (FR687018)

S. tropica CNB-440 (HQ642869)
S. tropica CNH-898 (HQ642870)
S. tropica CNS-193 (HQ642871)
S. tropica CNS-699 (HQ642872)

"S. pacifica" CNR-942 (HQ642855)
"S. pacifica" CNS-735 (HQ642859)
"S. pacifica" CNT-045 (HQ642861)
"S. pacifica" CNS-103 (HQ642856)
"S. pacifica" CNT-494 (HQ642854)
"S. pacifica" CNS-251 (HQ642857)
"S. pacifica" CNS-799 (HQ642860)
"S. pacifica" CNS-863 (HQ642858)
"S. pacifica" CNT-569 (HQ642866)
"S. pacifica" CNT-609 (HQ642868)
"S. pacifica" CNT-603 (HQ642867)
"S. pacifica" CNT-133 (HQ642863)
"S. pacifica" CNT-138 (HQ642864)
"S. pacifica" CNT-148 (HQ642856)
"S. pacifica" CNT-084 (HQ642862)

0.05 substitutions/site
Figure 3

Lechevalieria aerocolonigenes (AB0711405) rebD
Streptomyces sp. TP-A0274T (AB071406)
Streptomyces sp. (AB088119)
S. arenicola CNH-643T (HQ642932)
S. arenicola CNT-850 (HQ642938)
S. arenicola CNR-040 (HQ642934)
"S. pacifica" CNT-603 (HQ642959)
"S. pacifica" CNT-131 (HQ642954)
S. arenicola CNP-173 (HQ642933)
"S. pacifica" CNT-609 D (HQ642960)
S. arenicola CNS-205 (HQ642963)
S. arenicola CNS-673 (HQ642936)
"S. pacifica" CNS-055 A (HQ642939)
"S. pacifica" CNS-844 (HQ642943)
"S. pacifica" CNT-084 D (HQ642951)
"S. pacifica" CNT-148 A (HQ642956)
S. arenicola CNT-005 (HQ642937)
S. arenicola CNR-581 (HQ642935)
"S. pacifica" CNS-237 B (HQ642940)
"S. pacifica" CNS-251 (HQ642941)
"S. pacifica" CNT-124 C (HQ642953)
"S. pacifica" CNT-150 B (HQ642957)
"S. pacifica" CNS-799 (HQ642942)
"S. pacifica" CNS-890 (HQ642946)
"S. pacifica" CNT-094 C (HQ642952)
"S. pacifica" CNT-584 C (HQ642958)
"S. pacifica" CNT-854 C (HQ642961)
"S. pacifica" CNT-853 D (HQ642962)
"S. pacifica" CNS-860 (HQ642944)
"S. pacifica" CNS-863 C (HQ642945)
"S. pacifica" CNS-996 (HQ642947)
"S. pacifica" CNT-045 C (HQ642950)
"S. pacifica" CNT-044 C (HQ642949)
"S. pacifica" CNT-029 F (HQ642948)
"S. pacifica" CNT-138 E (HQ642955)

0.01 substitutions/site
Figure 4

A) MAU

B) UV

C) MS

m/z 274.0 [M+Na]+

m/z 274.1 [M+Na]+

Max: 5537

Max: 1650
Figure 5

A

CNS-863

CNS-205

CNS-863+CNS-205

B

UV peak 1 (12.1 min)

UV peak 2 (13.0 min)

UV peak 3 (13.9 min)

UV peak 4 (14.4 min) overlaid with staurosporine standard

C

m/z 483.1 [M+H]^+

MS peak 1

m/z 497.1 [M+H]^+

MS peak 2

m/z 467.1 [M+H]^+

MS peak 3

m/z 481.2 [M+H]^+

MS peak 4
Table 1. Strains used in this study. Only those that tested positive for either the salA KS or staD gene are listed of the 61 total that were screened.

<table>
<thead>
<tr>
<th>Strain number</th>
<th>Species</th>
<th>Location</th>
<th>Year</th>
<th>16S rRNA type (1)</th>
<th>16S rRNA accession #</th>
<th>SalA KS</th>
<th>KS accession #</th>
<th>StaD</th>
<th>StaD accession #</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNB-440</td>
<td>S. tropica</td>
<td>BA</td>
<td>1989</td>
<td>~</td>
<td>AY040617</td>
<td>Yes</td>
<td>HQ642869</td>
<td>No</td>
<td>NP</td>
</tr>
<tr>
<td>CNH-898</td>
<td>S. tropica</td>
<td>BA</td>
<td>2000</td>
<td>~</td>
<td>HQ642876</td>
<td>Yes</td>
<td>HQ642870</td>
<td>NT</td>
<td>~</td>
</tr>
<tr>
<td>CNR-699</td>
<td>S. tropica</td>
<td>BA</td>
<td>2003</td>
<td>~</td>
<td>JN161822</td>
<td>Yes</td>
<td>HQ642872</td>
<td>NT</td>
<td>~</td>
</tr>
<tr>
<td>CNS-193</td>
<td>S. tropica</td>
<td>BA</td>
<td>2004</td>
<td>~</td>
<td>JN161823</td>
<td>Yes</td>
<td>HQ642871</td>
<td>No</td>
<td>~</td>
</tr>
<tr>
<td>CNH-643</td>
<td>S. arenicola</td>
<td>BA</td>
<td>1999</td>
<td>~</td>
<td>AY371897</td>
<td>No</td>
<td>~</td>
<td>Yes</td>
<td>HQ642932</td>
</tr>
<tr>
<td>CNP-173</td>
<td>S. arenicola</td>
<td>USVI</td>
<td>2001</td>
<td>~</td>
<td>HQ642847</td>
<td>NT</td>
<td>~</td>
<td>Yes</td>
<td>HQ642933</td>
</tr>
<tr>
<td>CNR-040</td>
<td>S. arenicola</td>
<td>Guam</td>
<td>2002</td>
<td>~</td>
<td>JN161824</td>
<td>NT</td>
<td>~</td>
<td>Yes</td>
<td>HQ642934</td>
</tr>
<tr>
<td>CNR-581</td>
<td>S. arenicola</td>
<td>Guam</td>
<td>2002</td>
<td>~</td>
<td>JN161825</td>
<td>NT</td>
<td>~</td>
<td>Yes</td>
<td>HQ642935</td>
</tr>
<tr>
<td>CNS-205</td>
<td>S. arenicola</td>
<td>Palau</td>
<td>2004</td>
<td>~</td>
<td>CP000850</td>
<td>No</td>
<td>~</td>
<td>Yes</td>
<td>HQ642963</td>
</tr>
<tr>
<td>CNS-673</td>
<td>S. arenicola</td>
<td>Fiji</td>
<td>2006</td>
<td>~</td>
<td>JN161826</td>
<td>NT</td>
<td>~</td>
<td>Yes</td>
<td>HQ642936</td>
</tr>
<tr>
<td>CNT-005</td>
<td>S. arenicola</td>
<td>Fiji</td>
<td>2006</td>
<td>~</td>
<td>JN161827</td>
<td>NT</td>
<td>~</td>
<td>Yes</td>
<td>HQ642937</td>
</tr>
<tr>
<td>CNT-850</td>
<td>S. arenicola</td>
<td>Hawaii</td>
<td>2008</td>
<td>~</td>
<td>HQ642848</td>
<td>NT</td>
<td>~</td>
<td>Yes</td>
<td>HQ642938</td>
</tr>
<tr>
<td>CNR-551</td>
<td>&quot;S. pacifica&quot;</td>
<td>Guam</td>
<td>2002</td>
<td>A</td>
<td>HQ642881</td>
<td>Yes</td>
<td>HQ642854</td>
<td>No</td>
<td>~</td>
</tr>
<tr>
<td>CNR-942</td>
<td>&quot;S. pacifica&quot;</td>
<td>Palau</td>
<td>2004</td>
<td>E</td>
<td>HQ642877</td>
<td>Yes</td>
<td>HQ642855</td>
<td>No</td>
<td>~</td>
</tr>
<tr>
<td>CNS-055</td>
<td>&quot;S. pacifica&quot;</td>
<td>Palau</td>
<td>2004</td>
<td>A</td>
<td>DQ224159</td>
<td>No</td>
<td>~</td>
<td>Yes</td>
<td>HQ642939</td>
</tr>
<tr>
<td>CNS-103</td>
<td>&quot;S. pacifica&quot;</td>
<td>Palau</td>
<td>2004</td>
<td>~</td>
<td>DQ224160</td>
<td>Yes</td>
<td>HQ642856</td>
<td>No</td>
<td>~</td>
</tr>
<tr>
<td>CNS-237</td>
<td>&quot;S. pacifica&quot;</td>
<td>Palau</td>
<td>2004</td>
<td>B</td>
<td>HQ642850</td>
<td>No</td>
<td>~</td>
<td>Yes</td>
<td>HQ642940</td>
</tr>
<tr>
<td>CNS-251</td>
<td>&quot;S. pacifica&quot;</td>
<td>Palau</td>
<td>2004</td>
<td>~</td>
<td>HQ642879</td>
<td>Yes</td>
<td>HQ642857</td>
<td>Yes</td>
<td>HQ642941</td>
</tr>
<tr>
<td>CNS-735</td>
<td>&quot;S. pacifica&quot;</td>
<td>Fiji</td>
<td>2006</td>
<td>~</td>
<td>HQ642880</td>
<td>Yes</td>
<td>HQ642859</td>
<td>No</td>
<td>~</td>
</tr>
<tr>
<td>CNS-799</td>
<td>&quot;S. pacifica&quot;</td>
<td>Fiji</td>
<td>2006</td>
<td>C</td>
<td>HQ642895</td>
<td>Yes</td>
<td>HQ642860</td>
<td>Yes</td>
<td>HQ642942</td>
</tr>
<tr>
<td>CNS-844</td>
<td>&quot;S. pacifica&quot;</td>
<td>Fiji</td>
<td>2006</td>
<td>~</td>
<td>HQ642897</td>
<td>No</td>
<td>~</td>
<td>Yes</td>
<td>HQ642943</td>
</tr>
<tr>
<td>CNS-860</td>
<td>&quot;S. pacifica&quot;</td>
<td>Fiji</td>
<td>2006</td>
<td>C</td>
<td>HQ642886</td>
<td>No</td>
<td>~</td>
<td>Yes</td>
<td>HQ642944</td>
</tr>
<tr>
<td>CNS-863</td>
<td>&quot;S. pacifica&quot;</td>
<td>Fiji</td>
<td>2006</td>
<td>C</td>
<td>HQ642851</td>
<td>Yes</td>
<td>HQ642858</td>
<td>Yes</td>
<td>HQ642945</td>
</tr>
<tr>
<td>ID</td>
<td>Species</td>
<td>Location</td>
<td>Year</td>
<td>Type</td>
<td>Accession</td>
<td>NT</td>
<td>~</td>
<td>Yes</td>
<td>Accession</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>----------</td>
<td>------</td>
<td>------</td>
<td>-----------</td>
<td>----</td>
<td>----</td>
<td>-----</td>
<td>-----------</td>
</tr>
<tr>
<td>CNS-890</td>
<td>&quot;S. pacifica&quot;</td>
<td>Fiji</td>
<td>2006</td>
<td>C</td>
<td>HQ642889</td>
<td>No</td>
<td>~</td>
<td>Yes</td>
<td>HQ642946</td>
</tr>
<tr>
<td>CNS-996</td>
<td>&quot;S. pacifica&quot;</td>
<td>Fiji</td>
<td>2006</td>
<td>C</td>
<td>HQ642888</td>
<td>No</td>
<td>~</td>
<td>Yes</td>
<td>HQ642947</td>
</tr>
<tr>
<td>CNT-029</td>
<td>&quot;S. pacifica&quot;</td>
<td>Fiji</td>
<td>2006</td>
<td>F</td>
<td>HQ642852</td>
<td>No</td>
<td>~</td>
<td>Yes</td>
<td>HQ642948</td>
</tr>
<tr>
<td>CNT-044</td>
<td>&quot;S. pacifica&quot;</td>
<td>Fiji</td>
<td>2006</td>
<td>C</td>
<td>HQ642893</td>
<td>No</td>
<td>~</td>
<td>Yes</td>
<td>HQ642949</td>
</tr>
<tr>
<td>CNT-045</td>
<td>&quot;S. pacifica&quot;</td>
<td>Fiji</td>
<td>2006</td>
<td>C</td>
<td>HQ642887</td>
<td>Yes</td>
<td>~</td>
<td>Yes</td>
<td>HQ642950</td>
</tr>
<tr>
<td>CNT-084</td>
<td>&quot;S. pacifica&quot;</td>
<td>Fiji</td>
<td>2006</td>
<td>D</td>
<td>HQ642882</td>
<td>Yes</td>
<td>~</td>
<td>Yes</td>
<td>HQ642951</td>
</tr>
<tr>
<td>CNT-094</td>
<td>&quot;S. pacifica&quot;</td>
<td>Fiji</td>
<td>2006</td>
<td>C</td>
<td>HQ642894</td>
<td>No</td>
<td>~</td>
<td>Yes</td>
<td>HQ642952</td>
</tr>
<tr>
<td>CNT-124</td>
<td>&quot;S. pacifica&quot;</td>
<td>Fiji</td>
<td>2006</td>
<td>C</td>
<td>HQ642891</td>
<td>No</td>
<td>~</td>
<td>Yes</td>
<td>HQ642953</td>
</tr>
<tr>
<td>CNT-131</td>
<td>&quot;S. pacifica&quot;</td>
<td>Fiji</td>
<td>2006</td>
<td>~</td>
<td>HQ642896</td>
<td>No</td>
<td>~</td>
<td>Yes</td>
<td>HQ642954</td>
</tr>
<tr>
<td>CNT-133</td>
<td>&quot;S. pacifica&quot;</td>
<td>Fiji</td>
<td>2006</td>
<td>D</td>
<td>HQ218996</td>
<td>Yes</td>
<td>~</td>
<td>Yes</td>
<td>HQ642863</td>
</tr>
<tr>
<td>CNT-138</td>
<td>&quot;S. pacifica&quot;</td>
<td>Fiji</td>
<td>2006</td>
<td>E</td>
<td>HQ642853</td>
<td>Yes</td>
<td>~</td>
<td>Yes</td>
<td>HQ642955</td>
</tr>
<tr>
<td>CNT-148</td>
<td>&quot;S. pacifica&quot;</td>
<td>Fiji</td>
<td>2006</td>
<td>A</td>
<td>HQ642899</td>
<td>Yes</td>
<td>~</td>
<td>Yes</td>
<td>HQ642956</td>
</tr>
<tr>
<td>CNT-150</td>
<td>&quot;S. pacifica&quot;</td>
<td>Fiji</td>
<td>2006</td>
<td>B</td>
<td>HQ642900</td>
<td>No</td>
<td>~</td>
<td>Yes</td>
<td>HQ642957</td>
</tr>
<tr>
<td>CNT-569</td>
<td>&quot;S. pacifica&quot;</td>
<td>Fiji</td>
<td>2008</td>
<td>D</td>
<td>HQ642885</td>
<td>Yes</td>
<td>~</td>
<td>Yes</td>
<td>HQ642866</td>
</tr>
<tr>
<td>CNT-584</td>
<td>&quot;S. pacifica&quot;</td>
<td>Fiji</td>
<td>2008</td>
<td>C</td>
<td>HQ642890</td>
<td>No</td>
<td>~</td>
<td>Yes</td>
<td>HQ642958</td>
</tr>
<tr>
<td>CNT-603</td>
<td>&quot;S. pacifica&quot;</td>
<td>Fiji</td>
<td>2008</td>
<td>~</td>
<td>HQ642878</td>
<td>Yes</td>
<td>~</td>
<td>Yes</td>
<td>HQ642959</td>
</tr>
<tr>
<td>CNT-609</td>
<td>&quot;S. pacifica&quot;</td>
<td>Fiji</td>
<td>2008</td>
<td>D</td>
<td>HQ642883</td>
<td>Yes</td>
<td>~</td>
<td>Yes</td>
<td>HQ642960</td>
</tr>
<tr>
<td>CNT-853</td>
<td>&quot;S. pacifica&quot;</td>
<td>Hawaii</td>
<td>2008</td>
<td>D</td>
<td>HQ642884</td>
<td>No</td>
<td>~</td>
<td>Yes</td>
<td>HQ642962</td>
</tr>
<tr>
<td>CNT-854</td>
<td>&quot;S. pacifica&quot;</td>
<td>Hawaii</td>
<td>2008</td>
<td>C</td>
<td>HQ642892</td>
<td>No</td>
<td>~</td>
<td>Yes</td>
<td>HQ642961</td>
</tr>
</tbody>
</table>

NT = not tested.

~ = not applicable.

BA = Bahamas, USVI = US Virgin Islands.

(1) sequences types as reported in Freel et al, submitted as supplementary material for review only. The original sequence type identified for each species was not assigned a letter.