LBNL-62942
Preprint

ERNEST ORLANDO LAWRENCE
BERKELEY NATIONAL LABORATORY

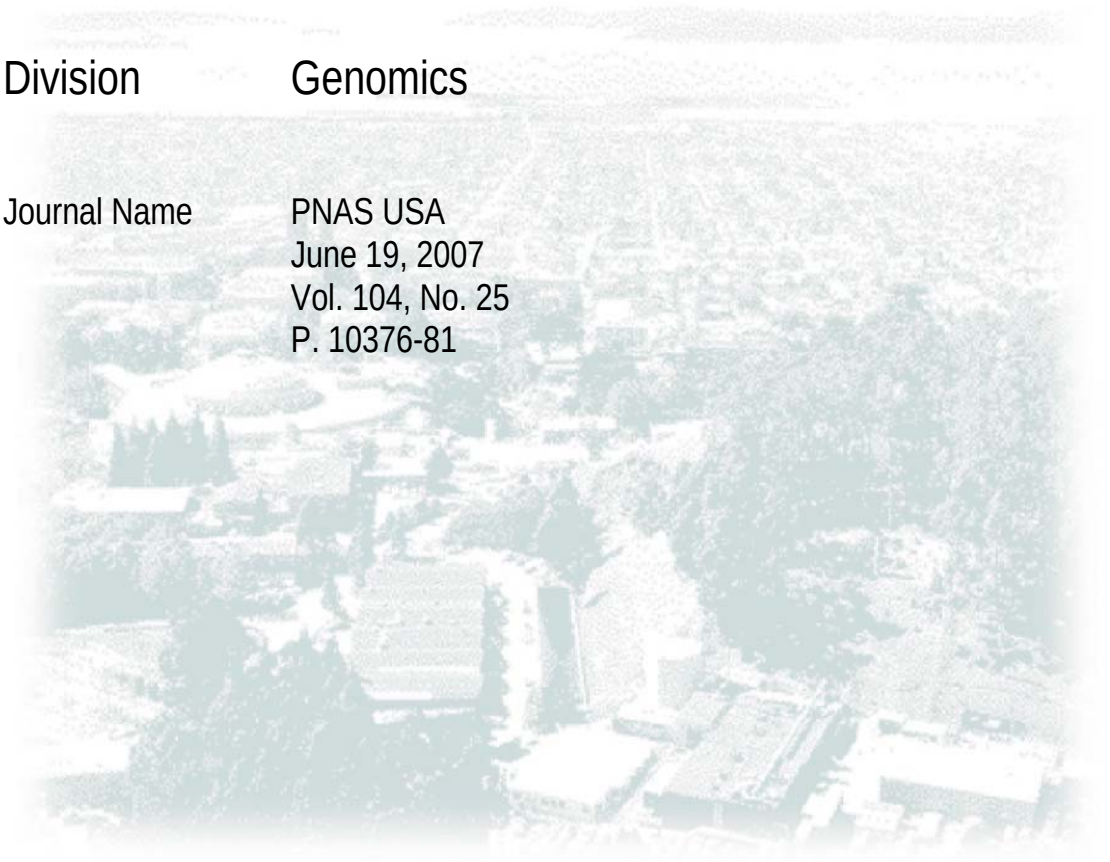
Title

Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*

Author(s)
Daniel W. Udwary, Lisa Zeigler, et al.

Division
Genomics

Journal Name
PNAS USA
June 19, 2007
Vol. 104, No. 25
P. 10376-81



Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*

Daniel W. Udwary,[*] *Lisa Zeigler,*[*] *Ratnakar Asolkar,*[*] *Vasanth Singan,*[†] *Alla Lapidus,*[†] *William Fenical,*[*] *Paul R. Jensen,*[*] *and Bradley S. Moore*[*,‡,§]

[*]Scripps Institution of Oceanography and [‡]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California at San Diego, La Jolla, CA 92093-0204, and [†]Department Of Energy, Joint Genome Institute-LBNL, Walnut Creek, CA 94598

[§] To whom correspondence should be addressed. E-mail: bsmoore@ucsd.edu.

Classification: Biological Sciences/Biochemistry

Manuscript information: 15 Text pages, 3 Figures, and 2 Tables.

Word and character count: Abstract: 174 words; Paper: - total characters.

Abstract

Recent fermentation studies have identified actinomycetes of the marine-dwelling genus *Salinispora* as prolific natural product producers. To further evaluate their biosynthetic potential, we analyzed all identifiable secondary natural product gene clusters from the recently sequenced 5,184,724 bp *S. tropica* CNB-440 circular genome. Our analysis shows that biosynthetic potential meets or exceeds that shown by previous *Streptomyces* genome sequences as well as other natural product-producing actinomycetes. The *S. tropica* genome features nine polyketide synthase systems of every known formally classified family, non-ribosomal peptide synthetases and several hybrid clusters. While a few clusters appear to encode molecules previously identified in *Streptomyces* species, the majority of the 15 biosynthetic loci are novel. Specific chemical information about putative and observed natural product molecules is presented and discussed. In addition, our bioinformatic analysis was critical for the structure elucidation of the novel polyene macrolactam salinilactam A. This study demonstrates the potential for genomic analysis to complement and strengthen traditional natural product isolation studies and firmly establishes the genus *Salinispora* as a rich source of novel drug-like molecules.

Introduction

Actinomycetes are an exceptionally prolific source of secondary metabolites accounting for more than half of all microbial antibiotics discovered to date (1). Remarkably, the vast majority of these compounds are derived from the single actinomycete genus *Streptomyces*, raising the intriguing possibility that additional, chemically prolific taxa await discovery. Further incentive to explore actinomycetes as a source of novel secondary metabolites comes from the genome sequences of *S. coelicolor* (2) and *S. avermitilis* (3), both of which revealed many unanticipated biosynthetic gene clusters thus demonstrating that even well-studied taxa have the potential to yield new metabolites. Such genomic-based information has been used not only to predict the chemical structures of previously unobserved metabolites but also to develop fermentation methods that enhance their production (4-7). Bioinformatics-based approaches to natural product discovery have also been used successfully at the industrial level, where genome scanning has led to the discovery of significant new chemical entities (8, 9). These methods have great potential to eliminate the redundant isolation of previously described compounds while allowing detailed fermentation studies or molecular cloning experiments to be focused on strains that possess a high probability of producing new chemical structures.

Genomics has already been particularly useful to microbial natural products studies because actinomycete secondary metabolites such as polyketides, nonribosomal peptides, and hybrids thereof are often biosynthesized by large, multifunctional synthases that in an assembly line process sequentially assemble small carboxylic acid and amino acid building blocks into their products (10). The biosynthetic genes responsible for the production of these metabolites are almost invariably tightly packaged into operon-like clusters that include regulatory elements and resistance mechanisms (2). In the case of modular polyketide synthase (PKS) and nonribosomal peptide synthetase (NRPS) systems, the repetitive domain structures associated with these megasynthases generally follow a co-linearity rule (11) which, when combined with bioinformatics and biosynthetic

precedence, can be used to predict the chemical structures of new polyketide and peptide-based metabolites.

Marine-derived actinomycetes have become a focus in the search for novel secondary metabolites (12). Among the strains cultured from marine samples is the genus *Salinispora*, which was recently described as the first seawater-requiring marine actinomycete (13). This genus is widely distributed in tropical and subtropical ocean sediments and is currently comprised of the two formally described species, *S. tropica* and *S. arenicola*, and a third species for which the name *S. pacifica* has been proposed (14). These actinomycetes are proving to be an exceptionally rich source of structurally diverse secondary metabolites, which are produced in species-specific patterns (15). In the case of *S. tropica*, the compounds observed to date from this bacterium include the potent proteasome inhibitor salinosporamide A (16), which is currently in phase I human clinical trials for the treatment of cancer, the unprecedented halogenated macrolides sporolides A and B (17), lymphostin, which was observed by scientists at Nereus Pharmaceuticals during salinosporamide A fermentation development (R. Lam, personal communication) and was first reported from a *Streptomyces* species (18) and, salinilactam, the structure of which was solved as part of the present study.

Here we report all identified secondary metabobolic biosynthetic gene clusters deriving from the complete genome sequence of the marine actinomycete *S. tropica* strain CNB-440. This strain is now found to hold the most diverse assemblage of polyketide biosynthetic mechanisms observed in a single organism, as well as the largest percentage of a genome devoted to natural product biosynthesis to date. Bioinformatic analysis was used to facilitate the structure elucidation of the novel polyene macrolactam salinilactam A. Sequence analyses further revealed that many of these clusters are likely to have been introduced into this genome as a result of recent horizontal gene transfer, which has important implications to origin of this organism's secondary metabolome.

Results and Discussion

General features of the genome and associated secondary metabolome. The complete genome sequence of *S. tropica* strain CNB-440 revealed a single circular chromosome comprising 5,184,724 bp, with no plasmids, and an average G+C content of 69.5% (Fig. 1). *S. tropica* has approximately 4,500 predicted protein-coding sequences and is similar in size when compared to other actinomycetes that harbor circular chromosomes [*Mycobacterium tuberculosis* (19), *Frankia* sp. CcI3 (GenBank accession CP000249), and *Nocardia farcinica* (20)], yet is substantially smaller than those with linear chromosomes [*S. coelicolor* (2) and *Rhodococcus* sp. RHA1 (21)] (Table 1).

Biosynthetic gene clusters were initially identified from a draft genome sequence provided by the Joint Genome Institute (publicly released September 1, 2006), and checked against a single contiguous sequence provided in advance of public release. Each putative open reading frame (ORF) was compared against a representative library of all known PKS and NRPS domains, as well as known biosynthetic accessory-type genes (including, but not limited to, cytochrome P450s, terpene cyclases, prenyl transferases, methyl transferases, NAD(P)H-dependent oxidoreductases, and CoA/AMP ligases).

This analysis revealed 15 secondary metabolic biosynthesis gene clusters that we predict to be involved in siderophore, melanin, polyketide, non-ribosomal peptide, and aminocyclitol production (Table 1). The combined length of these gene clusters is estimated at 492 kb, establishing that, amongst sequenced bacteria to date, *S. tropica* devotes the largest percentage of its genome (~9.5%) to natural product assembly (Table 1). These analyses also confirm that the biosynthetic potential of this strain is considerably greater than that observed by fermentation, as was previously reported in analyses of the *S. coelicolor* and *S. avermitilis* genome sequences (2, 3).

Two thirds of the clusters are concentrated in a single quadrant of the chromosome nearly antipodal to the replication-controlling *dnaA* gene (Fig. 1). This section of the genome also harbors

a significant proportion of the approximate 128 chromosomal mobile genetic elements such as transposases, integrases, resolvases and phage-related ORFs. Many reside in close proximity to secondary metabolic clusters (*pks1*, *sal*, *spo, amc*, *sid3*, *slm*) while several clusters have strong similarity in ORF and domain organization to existing gene clusters (*sid2*, *sid3*, *pks3*), strongly suggesting that many of these clusters may have originated from donor genomes via horizontal gene transfer.

Biosynthetic capabilities of *S. tropica*. The majority of identified *S. tropica* secondary metabolite gene clusters utilize carrier protein-based biosynthetic logic in the assembly of their products. Hence, biosynthetic precursors and intermediates are expected to be largely protein-bound during the assembly process and are not likely to crosstalk with other primary or secondary metabolic pathways. *Salinispora tropica* harbors at least five (Stro0922, Stro2003, Stro3049, Stro3460, Stro3545) 4'-phosphopantetheinyl transferase (PPTase) (22) encoding genes whose products transfer the 4'-phosphopantetheine moiety from coenzyme A to the ~35 conserved serine residues in fatty acid synthase-, PKS-, and NRPS-associated carrier proteins. Three are found in identified carrier protein-containing biosynthetic clusters (Stro0922 with *sid3*, Stro2003 with *pks3*, and Stro3545 with *nrps2*) and thus may be specific to their pathways.

A striking feature of the *S. tropica* genome is its remarkable diversity of polyketide biosynthetic pathways, which is greater in variety than in other sequenced bacterial genomes (Table 1). Pathways include those for modular type I PKSs (*slm*), iterative enediyne type I PKSs (*spo* and *pks1*), hybrid type I PKS–NRPSs (*sal*, *sid2*, *sid3*, and *lym*), heterodimeric type II PKSs (*pks2* and *pks3*), and a homodimeric type III PKS (*pks4*) (Fig. 1). Interestingly, none of the type I PKS-associated modules, of which there are 15, contain a full set of reductive domains required to completely reduce the transient β-carbonyl group to a saturated methylene carbon during the polyketide elongation process. Hence, *S. tropica* PKS-derived products are expected to be highly oxidized, as is the case for the characterized metabolites salinosporamide A (*sal*), sporolide A (*spo*), lymphostin (*lym*), and salinilactam A (*slm*) (Fig. 2).

A primary motivation for the sequencing of the CNB-440 strain of *S. tropica* was that it is a producer of the potent anticancer agent salinosporamide A (16). Its biosynthetic gene cluster (*sal*) encodes a 29-ORF, 41-kb unprecedented hybrid PKS–NRPS pathway (Fig. 2E) involving new mechanisms in chlorination and β-lactone synthesis as well as 20S proteasome resistance (23). In addition to the *sal* cluster, *S. tropica* harbors two additional hybrid PKS–NRPS pathways. The *lym* cluster is proposed to synthesize the lymphocyte kinase inhibitor lymphostin, which was first isolated from *Streptomyces* sp. KY11783 (18). We propose that the uniquely organized two-module synthase LymA catalyzes a single polyketide extension followed by reductive offloading and *N*-acetylation of a non-proteinogenic tricyclic planar amino acid derived from tryptophan, with subsequent *O*-methylation catalyzed by an adjacent SAM-dependent methyl transferase LymB (Fig. 2F). A set of four genes with homology to a domain of unknown function (DUF611) may play a role in the synthesis of the tryptophan-derived primer unit. The *sid2* cluster bears strong domain organization (though lacks direct sequence similarity) to the yersiniabactin cluster (*ybt*) from *Yersinia pestis* (24) and other organisms (Fig. 2B). Although no siderophore molecules have yet been isolated from *Salinispora* species, the notable similarity in domain structure very strongly suggests that the resulting molecule will have the yersiniabactin polyketide/non-ribosomal peptide core with possible alterations in sites of methylation and/or oxidation.

Two independent, novel enediyne PKS biosynthetic gene clusters (*spo* and *pks1*) are found in the *S. tropica* genome. Enediyne natural products, such as the potent DNA cleaving agents calicheamicin, dynemicin, C-1027, and neocarzinostatin are polyketide secondary metabolites produced by members of the order *Actinomycetales*, many of which (like the genus *Salinispora*) belong to the family *Micromonosporaceae* (25). The enediyne polyketide structural motif falls into two classes that possess either 9- or 10-membered rings, and their associated PKSs are phylogenetically distinct. Phylogenetic analysis of the *S. tropica* enediyne PKSs associated with the *spo* and *pks1* clusters revealed that they are putatively involved in the biosynthesis of distinct 9- and 10-membered enediyne polyketides, respectively (Fig. 3). While no enediyne natural products have yet been directly characterized from *S. tropica*, sporolides A and B (17) are putatively derived from

a 9-membered enediyne PKS product that is uniquely halogenated with chloride during the aromatization of the enediyne unit to yield the tricyclic hydrocarbon nucleus (26). The ~71-kb *spo* cluster harbors as many as 53 ORFs, which include other conserved enediyne PKS-associated genes, as well as those encoding the postulated biosynthesis and coupling of the sporolide cyclohexenone unit deriving from the amino acid tyrosine. The 10-membered enediyne PKS in gene cluster *pks1* also bears the canonical enediyne PKS domain architecture, and further analysis of its 30-kb gene cluster did not reveal accessory genes involved in deoxyhexose biosynthesis and attachment, nor additional polyketide pathways typically associated with 10-membered enediyne polyketide biosynthesis (*i.e.*, calicheamicin and dynemicin), suggesting that this orphan pathway codes for the biosynthesis of a novel natural product.

In addition to the abundance of diverse PKS pathways, *S. tropica* harbors four NRPS operons (*nrps1*, *nrps2*, *sid3*, and *sid4*) (Fig. 1). While the products of the dimodular *nrps1* and novel tetramodular *nrps2* gene clusters are unknown and could not be accurately predicted on the basis of bioinformatics alone (27, 28), *sid3* and *sid4* are homologous to gene sets encoding known siderophore pathway biosynthetic enzymes. When combined with the highly homologous *S. coelicolor* NRPS-independent siderophore desferrioxamine (29) cluster *sid1* (Fig. 2A) and the *ybt* homologous PKS–NRPS hybrid cluster *sid2* (Fig. 2B), a total of four identifiable siderophore-like clusters reside in *S. tropica*. The three assembly line siderophore synthetases encoded by gene sets *sid2–4* each have related priming mechanisms involving homologous ATP-dependent aryl-CoA ligases, which typically incorporate aromatic residues such as salicylate or 2,3-dihydroxybenzoate into their NRP siderophore products such as yersiniabactin, enterobactin, and pyochelin (30). The *sid4* gene cluster has strongest homology with a biosynthetic cluster for an as-yet unisolated siderophore ('coelibactin') from *S. coelicolor* (Fig. 2D) (2). The *S. tropica* sequence, however, differs from that in *S. coelicolor* in that it is flanked by an additional transport system typical of siderophore export and iron-bound uptake. The smaller *sid3* cluster is adjacent to the *sid4* cluster and encodes a highly unusual type II PKS and a single NRPS module that we propose to extend and cyclize an aromatic acid with cysteine, tentatively producing a siderophore-like intermediate

(Fig. 2C). It is unclear if *sid3* is fully functional or is instead biosynthetically linked with *sid4*. While the latter three gene clusters are traditionally found in γ-proteobacteria, these *S. tropica* gene clusters bear %G+C values similar to the *S. tropica* total genome percentage, and therefore do not suggest a very recent assimilation, despite similarities to existing clusters.

Genome-guided natural product discovery of salinilactam A. The *slm* cluster is the largest *S. tropica* biosynthetic cluster and consists of six genes encoding a ten-module PKS with 49 domains. Due to the highly repetitive nature of the DNA sequence associated with the *slm* loci and an extremely high level of sequence identity (>99% in many regions), assembly of the cluster and, as a consequence, closure of the genome was problematic. Initial inspection of the partial *slm* gene cluster from an early draft genome sequence suggested that it coded for a novel lysine-primed polyene macrolactam polyketide devoid of sugar or other appendages often associated with polyketide natural products. A fermentation broth of *S. tropica* CNB-440 was inspected for new compounds with characteristic UV chromophores associated with polyene units, which led to the isolation of a series of new polyene macrolactams exemplified by salinilactam A (Fig. 2G). NMR and MS characterization of salinilactam A quickly revealed a novel polyene macrolactam framework, which was only consistent with the *slm* cluster. Structural features included two isolated polyene fragments, a 1,2,3,5-tetrahydroxy alkane moiety, three methyl groups, and an amide functionality. However, completion of the structure was initially hindered due to the instability of the compound and the presence of eight conjugated olefins with similar NMR properties.

Inspection of the structure fragments, together with the molecular formula $C_{28}H_{39}NO_5$ revealed by high-resolution mass spectrometry, suggested that salinilactam A was derived from a PKS with at least ten extension modules. This information was useful to help resolve and properly assemble the repetitive DNA sequences associated with *slm* into two operons (*slmABCD* and *slmNO*) separated by nine accessory genes presumed to be involved in starter unit biosynthesis and macrolactam hydroxylation (Fig. 2G). On the basis of the co-linearity rule of modular polyketide biosynthesis, the relationship of *slm* to the terminal region of the vicenistatin PKS (31), and partial

NMR-based structural fragments, we were able to accurately predict the gross chemical structure of salinilactam A. The natural product would thus be produced by chain extension of the novel lysine-derived PKS starter unit 5-aminohex-2-enoate with eight malonate and two methylmalonate units followed by macrolactamization and cytochome P450 hydroxylation at C-8.

The salinilactam A structure was verified by comprehensive NMR analyses on the purified natural product to confirm the bioinformatics-based structure assignment. Although the relative and absolute stereochemistry of salinilactam A was not determined spectroscopically, the stereochemistry of the C-7, C-9, and C-11 hydroxyl groups was predicted to be *R*, *S*, and *R*, respectively, based upon strong homology of all *slm* ketoreductase (KR) domains to other "A-type" KRs (32). The intimate interplay between microbial genomics, biosynthetic logic and natural product chemistry was critical not only in the structure elucidation of this new chemical entity, but also in the final closure of the genome sequence.

Conclusions

There is an ongoing resurgence in natural product drug discovery research (33). This renewed interest is driven by the inherent structural diversity and biological activity associated with secondary metabolites, the low productivity of alternative drug discovery strategies, and the application of improved analytical methods that make it possible to solve structures using small quantities of material. A relatively new concept that can be added to this list is the molecular genetics of natural product biosynthesis (34). Advances in this field, in combination with increased access to DNA sequencing, are providing a wealth of information about how natural products are assembled, mechanisms by which natural product gene clusters can be manipulated to yield new product diversity, and the genetic potential of individual organisms. Concerning the later, complete genome sequencing provides unparalleled access to the genes involved in secondary metabolism,

how they are assembled, and in some cases, what products they may yield. Armed with these data, it becomes possible to compare the compounds observed using traditional fermentation procedures with those that are predicted from gene sequences, design fermentation methods that may activate or enhance the production of predicted products, and assess the evolutionary history of individual gene clusters. The genus *Salinispora* has become a model organism to address questions about species-specific patterns of natural product production (15) and, as described in the present study, the application of whole genome sequencing to the process of natural product discovery and structure elucidation.

Genome sequencing of *S. tropica* has revealed an abundance of novel biosynthetic gene clusters, the majority of which were unexpected based on previous fermentation analyses of this and closely related species and strains. With this information now available, there is a clear need for further genome-guided fermentation studies as our analysis clearly confirms the value of the *Salinispora* genus as a source of novel drug-like molecules. In the absence of other *Micromonsporacea* genome sequences, it is unclear if this natural product diversity is general to this family or if the diversity in the genus *Salinispora* derives from environmental novelty. Sequencing of a related species, *Salinispora arenicola* CNS-205, already underway, should shed further light on the metabolic capabilities of this newly described marine actinomycete genus.

Outside of the exploratory aspects of this project, our bioinformatic analysis was of broad practical utility in conjunction with natural product isolation. The isolation of a large polyene molecule led to identification and resolution of a large, repetitive PKS sequence, the analysis of which spurred further refinement and identification of salinilactam A as a novel polyene. With the genome information in hand we expect that it will facilitate similar studies in the future, and this result demonstrates the benefits that can result from greater interplay between information technology and natural product structure elucidation. We recently developed a genetics system in *S. tropica* for rapid gene knockouts via PCR targeting that will further facilitate the genome mining of this metabolically rich bacterium (23).

Currently, the implications for secondary metabolism due to the circular nature of the *S. tropica* chromosome are unclear. A mechanism for the biogenesis of novel secondary metabolic clusters in linear chromosomes was recently reported (35), and makes clear that the instability found at of the terminal ends of *Streptomyces* linear chromosomes has been utilized to provide rapid evolutionary adaptability to the organism. By contrast, no particular region of the circular *S. tropica* chromosome seems unstable to the same degree observed in the chromosomal ends of *Streptomyces* as pseudogenes are largely absent and secondary metabolic clusters and transposable elements are more dispersed throughout the chromosome. This may indicate that *Salinispora* species will have more frequently acquired their secondary metabolic systems horizontally from other species rather than evolving most themselves. It will be appealing to put this hypothesis to the test when the *S. arenicola* CNS-205 genome is completed and directly compared to *S. tropica* CNB-440.

Materials and Methods

Genome sequencing, annotation and analysis. Draft sequencing and automated annotation was provided by the Department of Energy, Joint Genome Institute (JGI) under the Community Sequencing Program and has been deposited with GenBank (AATJ00000000). Automated gene prediction and functional annotation was performed by the Genome Analysis Pipeline (36). Automated prediction of protein-coding genes was based on the output of CRITICA (37) complemented with the output of Glimmer (38). Initial functional annotation was performed using similarity searches against TIGRFAM (39), PRIAM (40), Pfam (41), SMART (42), COGs (43), SwissProt/TrEMBL (44) and KEGG (45) databases using a set of rules for assigning a specific product description depending on the combination of the search results; product descriptions were further manually refined. Annotated genome sequences are available at the Microbial Portals at http://genome.ornl.gov/microbial/ and through the Integrated Microbial Genomes (IMG) system (46) at http://img.jgi.doe.gov/.

To more specifically identify and categorize natural product biosynthetic gene clusters, a custom Perl script was written to sequentially search each translated protein sequence via BLASTP alignment against a hand-constructed and expandable library of model natural product domains and genes. Hits were grouped by physical proximity within the chromosome, and putative clusters were further examined using software available at the NCBI website and visualized with Vector NTI (Invitrogen). Transposons and other mobile genetic element ORFs were identified in a similar manner. Annotations of all gene clusters discussed were deposited with GenBank (for accession numbers, see Table 2).

Salinilactam A isolation and characterization. Strain CNB-440/Δ*salD* was cultured in 9 x 1L Fernbach flasks containing A1bFe+C medium (10 g starch, 4 g yeast extract, 2 g peptone, 1 g $CaCO_3$, 5 mL $Fe_2(SO_4)_3 \cdot 4H_2O$ at 8 g/L, 5 mL KBr at 20 g/L, 1 liter seawater) at 27°C while shaking at 230 rpm. XAD-7 resin (20 g) was added to each flask after 24 hr, and the fermentation was continued for another six days. The resin was filtered, washed with water, and then extracted with acetone to afford 1.1 g of crude extract, which was fractionated by reversed-phase C18 vacuum liquid chromatography eluting with increasing amounts of methanol in water. The third fraction that eluted with 60% methanol (105 mg) was purified by RP-HPLC [Prep Nova-Pak HR C18, 6 μm, 60Å, 300 mm x 40 mm, flow rate 10 mL/min, detection at 210 nm, 25% for 10 min then a linear gradient up to 40% CH_3CN over 15 min and linear gradient to 100% over another 25 min] to afford salinilactam A (t_R = 25 min, 1.5 mg): light yellow solid; $[\alpha]_D$ –28.8 (*c* 0.06, MeOH); UV (MeOH) λ_{max} (log ε) 250 (4.01), 305 (4.51), 335 (4.26), 355 (4.18) nm; IR (NaCl) λ_{max} 3464, 2930, 1739, 1547, 1414, 1279, 1135, 1074, 998, and 743 cm^{-1}; ^1H NMR (500 MHz, CD_3OD) δ 7.02 (1H, dd, *J* = 15.6, 10.7 Hz, H-3), 6.76 (1H, d, *J* = 14.6 Hz, H-17), 6.51 (1H, dd, *J* = 15.0, 10.2 Hz, H-20), 6.49 (1H, dd, *J* = 14.8, 10.3 Hz, H-14), 6.30 (1H, dd, *J* = 15.0, 10.2 Hz, H-16), 6.29 (1H, dd, *J* = 14.8, 10.2 Hz, H-15), 6.27 (1H, dd, *J* = 14.9, 10.5 Hz, H-13), 6.25 (1H, dd, *J* = 15.1, 10.7 Hz, H-4), 6.10 (1H, dd, *J* = 14.6, 10.5 Hz, H-21), 6.05 (1H, dd, *J* = 15.0, 10.0 Hz, H-22), 6.02 (1H, dd, *J* = 15.1, 10.1 Hz, H-5), 5.95 (1H, d, *J* = 15.1 Hz, H-19), 5.87 (1H, d, *J* = 15.2 Hz, H-2), 5.77 (1H, dd, *J* =

14.8, 10.5 Hz, H-12), 5.67 (1H, ddd, *J* = 15.0, 10.1, 7.0 Hz, H-23), 4.36 (1H, q, *J* = 11.2, 4.1 Hz, H-11), 3.97 (1H, dt, *J* = 11.2, 6.8, 4.1 Hz, H-9), 3.86 (1H, m, H-25), 3.60 (1H, dd, *J* = 7.7, 4.1 Hz, H-8), 3.50 (1H, dd, *J* = 7.7, 5.0 Hz, H-7), 2.59 (1H, m, H-6), 2.33 (1H, m, H-24a), 2.19 (1H, ddd, *J* = 14.1, 7.0, 6.5 Hz, H-24b), 1.98 (1H, m, H-10a), 1.84 (3H, s, H_3-28), 1.78 (1H, dd, *J* = 14.8, 6.8 Hz, H-10b), 1.27 (3H, d, *J* = 7.1 Hz, H_3-26), 1.08 (3H, d, *J* = 7.2 Hz, H_3-27); ^{13}C NMR (125 MHz, CD_3OD) δ 168.2 (C-1), 142.1 (C-5), 139.4 (C-3), 138.9 (C-20), 136.9 (C-12), 134.4 (C-22), 133.1 (C-18), 131.2 (C-15), 130.5 (C-19), 130.3 (C-23), 129.6 (C-17), 129.4 (C-21), 128.9 (C-16), 128.8 (C-14), 128.7 (C-4), 128.6 (C-13), 123.1 (C-3), 75.6 (C-7), 75.2 (C-8), 71.3 (C-11), 69.3 (C-9), 46.7 (C-25), 39.5 (C-10), 39.4 (C-24), 38.7 (C-6), 19.7 (C-25), 19.4 (C-28), 13.8 (C-27); ESIMS *m/z* 452 (M – H_2O + H), 470 (M + H), 492 (M + Na), 961 (2M + Na); HRESITOFMS *m/z* 492.2720 [calcd for $C_{28}H_{39}NO_5Na^+$, 492.2726].

Acknowledgments. We thank Dr. G. L. Challis (University of Warwick) for valuable discussions, Dr. D. C. Oh for preliminary characterization of salinilactam A, and Dr. A. Eustaquio for strain CNB-440/Δ*salD*. Funding was provided from NOAA through the Oceans and Human Initiative (NA05NOS4781249) to B.S.M. and P.R.J. and the NIH (R37 CA44848) to W.F. Genome sequencing was provided through a grant from the Community Sequencing Program of JGI to P.R.J. and B.S.M.

This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231 and Los Alamos National Laboratory under contract No. DE-AC52-06NA25396.

References

1. Berdy, J. (2005) *J. Antibiot.* **58,** 1-26.
2. Bentley, S. D., Chater, K. F., Cerdeño-Tárraga, A.-M., Challis, G. L., Thomson, N. R., James, K. D., Harris, D. E., Quail, M. A., Kieser, H., Harper, D.*, et al.* (2002) *Nature* **417,** 141-147.
3. Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T.*, et al.* (2001) *Proc. Natl. Acad. Sci. USA* **98,** 12215-12220.
4. Lautru, S., Deeth, R. J., Bailey, L. M., & Challis, G. L. (2005) *Nat. Chem. Biol.* **1,** 244-245.
5. Song, L., Barona-Gomez, F., Corre, C., Xiang, L., Udwary, D. W., Austin, M. B., Noel, J. P., Moore, B. S., & Challis, G. L. (2006) *J. Am. Chem. Soc.* **128,** 14754-14755.
6. Bok, J. W., Hoffmeister, D., Maggio-Hall, L. A., Murillo, R., Glasner, J. D., & Keller, N. P. (2006) *Chem. Biol.* **13,** 31-37.
7. Gross, H., Stockwell, V. O., Henkels, M. D., Nowak-Thompson, B., Loper, J. E., & Gerwick, W. H. (2007) *Chem. Biol.*.**14**, 53-63
8. Zazopoulos, E., Huang, K. X., Staffa, A., Liu, W., Bachmann, B. O., Nonaka, K., Ahlert, J., Thorson, J. S., Shen, B., & Farnet, C. M. (2003) *Nat. Biotechnol.* **21,** 187-190.
9. McAlpine, J. B., Bachmann, B. O., Piraee, M., Tremblay, S., Alarco, A. M., Zazopoulos, E., & Farnet, C. M. (2005) *J. Nat. Prod.* **68,** 493-496.
10. Fischbach, M. A. & Walsh, C. T. (2006) *Chem. Rev.* **106,** 3468-3496.
11. Staunton, J. & Weissman, K. J. (2001) *Nat. Prod. Rep.* **18,** 380-416.
12. Fenical, W. & Jensen, P. R. (2006) *Nat. Chem. Biol.* **2,** 666-673.
13. Maldonado, L., Fenical, W., Goodfellow, M., Jensen, P. R., & Ward, A. C. (2005) *Int. J. System. Appl. Microbiol.* **55,** 1759-1766.
14. Jensen, P. R. & Mafnas, C. (2006) *Environ. Microbiol.* **8,** 1881-1888.
15. Jensen, P. R., Williams, P. G., Oh, D.-C., Zeigler, L., & Fenical, W. (2007) *Appl. Envir. Microbiol.*, published online: doi:10.1128/AEM.01891-06.
16. Feling, R. H., Buchanan, G. O., Mincer, T. J., Kauffman, C. A., Jensen, P. R., & Fenical, W. (2003) *Angew. Chem. Int. Ed.* **115,** 369-371.
17. Buchanan, G. O., Williams, P. G., Feling, R. H., Kauffman, C. A., Jensen, P. R., & Fenical, W. (2005) *Org. Lett.* **7,** 2731-2734.
18. Aotani, Y., Nagata, H., & Yoshida, M. (1997) *J. Antibiot.* **50,** 543-545.
19. Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E.*, et al.* (1998) *Nature* **393,** 537-544.
20. Ishikawa, J., Yamashita, A., Mikami, Y., Hoshino, Y., Kurita, H., Hotta, K., Shiba, T., & Hattori, M. (2004) *Proc. Natl. Acad. Sci. U.S.A.* **101,** 14925-14930.
21. McLeod, M. P., Warren, R. L., Hsiano, W. W. L., Araki, N., Myhre, M., Fernandes, C., Miyazawa, D., Wong, W., Lillquist, A. L., Wang, D.*, et al.* (2006) *Proc. Natl. Acad. Sci. U.S.A.* **103,** 15582-15587.
22. Lambalot, R. H., Gehring, A. M., Flugel, R. S., Zuber, P., LaCelle, M., Marahiel, M. A., Reid, R., Khosla, C., & Walsh, C. T. (1996) *Chem. Biol.* **3,** 923-936.
23. Eustáquio, A. S., Beer, L., Udwary, D. W., Asolkar, R., Fenical, W., & Moore, B. S. (2007), submitted for publication.
24. Gehring, A. M., DeMoll, E., Fetherston, J. D., Mori, I., Mayhew, G. F., Blattner, F. R., Walsh, C. T., & Perry, R. D. (1998) *Chem. Biol.* **5,** 573-586.

25. Shen, B., Liu, W., & Nonaka, K. (2003) *Curr. Med. Chem.* **10,** 2317-2325.
26. Oh, D. C., Williams, P. G., Kauffman, C. A., Jensen, P. R., & Fenical, W. (2006) *Org. Lett.* **8,** 1021-1024.
27. Stachelhaus, T., Mootz, H. D., & Marahiel, M. A. (1999) *Chem. Biol.* **6,** 493-505.
28. Challis, G. L., Ravel, J., & Townsend, C. A. (2000) *Chem. Biol.* **7,** 211-224.
29. Barona-Gomez, F., Wong, U., Giannakopulos, A. E., Derrick, P. J., & Challis, G. L. (2004) *J. Am. Chem. Soc.* **126,** 16282-16283.
30. Crosa, J. H. & Walsh, C. T. (2002) *Microbiol. Mol. Biol. Rev.* **66,** 223-249.
31. Ogasawara, Y., Katayama, K., Minami, A., Otsuka, M., Eguchi, T., & Kakinuma, K. (2004) *Chem. Biol.* **11,** 79-86.
32. Siskos, A. P., Baerga-Ortiz, A., Bali, S., Stein, V., Mamdani, H., Spiteller, D., Popovic, B., Spencer, J. B., Staunton, J., Weissman, K. J.*, et al.* (2005) *Chem. Biol.* **12,** 1145-1153.
33. Clardy, J. & Walsh, C. (2004) *Nature* **432,** 829-837.
34. Fischbach, M. A. & Walsh, C. T. (2006) *Chem. Rev.* **106,** 3468-3496.
35. Choulet, F., Aigle, B., Gallois, A., Mangenot, S., Gerbaud, C., Truong, C., Francou, F. X., Fourrier, C., Guerineau, M., Decaris, B.*, et al.* (2006) *Mol. Biol. Evol.* **23,** 2361-2369.
36. Hauser, L., Larimer, F., Land, M., Shah, M., & Uberbacher, E. (2004) *Genet. Eng.* **26,** 225-238.
37. Badger, J. H. & Olsen, G. J. (1999) *Mol. Biol. Evol.* **16,** 512-524.
38. Delcher, A. L., Harmon, D., Kasif, S., White, O., & Salzberg, S. L. (1999) *Nucleic Acids Res.* **27,** 4636-4641.
39. Haft, D. H., Loftus, B. J., Richardson, D. L., Yang, F., Eisen, J. A., Paulsen, I. T., & White, O. (2001) *Nucleic Acids Res.* **29,** 41-43.
40. Claudel-Renard, C., Chevalet, C., Faraut, T., & Kahn, D. (2003) *Nucleic Acids Res.* **31,** 6633-6639.
41. Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S. R., Griffiths-Jones, S., Howe, K. L., Marshall, M., & Sonnhammer, E. L. (2002) *Nucleic Acids Res.* **30,** 276-280.
42. Letunic, I., Copley, R. R., Pils, B., Pinkert, S., Schultz, J., & Bork, P. (2006) *Nucleic Acids Res.* **34,** D257-260.
43. Tatusov, R. L., Galperin, M. Y., Natale, D. A., & Koonin, E. V. (2000) *Nucleic Acids Res.* **28,** 33-36.
44. Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E., Martin, M. J., Michoud, K., O'Donovan, C., Phan, I.*, et al.* (2003) *Nucleic Acids Res.* **31,** 365-370.
45. Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., & Hattori, M. (2004) *Nucleic Acids Res.* **32,** D277-280.
46. Markowitz, V. M., Korzeniewski, F., Palaniappan, K., Szeto, E., Werner, G., Padki, A., Zhao, X., Dubchak, I., Hugenholtz, P., Anderson, I.*, et al.* (2006) *Nucleic Acids Res.* **34,** D344-348.

Captions

Fig. 1. Circular chromosome of *Salinispora tropica* CNB-440, oriented to *dnaA* gene. Outer ring, locations of secondary metabolic gene clusters. Inner ring, locations of identified mobile genetic elements.

Fig. 2. Selected genes from *S. tropica* modular biosynthetic enzyme systems and their associated natural products. (A-D) putative siderophore clusters: (A) *sid1* cluster in comparison to the desferrioxamine cluster from *Streptomyces coelicolor* A3(2). (B) *sid2* cluster in comparison to the yersiniabactin cluster from *Yersinia pestis*. (C) *sid3* cluster with putative product intermediate. (D) *sid4* cluster in comparison to proposed *S. coelicolor* 'coelibactin' cluster (unconfirmed structure). (E-G) Novel modular enzyme systems: (E) *salAB* from proposed salinosporamide cluster. (F) *lymAB* from proposed lymphostin cluster. (G) modular PKS system from salinilactam cluster (stereochemistry proposed, green portion is derived from the proposed lysine-based starter unit). Domain notation: A, adenylation (amino acid substrate noted); C, condensation; Ccyc, condensation with cyclization; PCP, peptidyl carrier protein; KS, β-ketoacyl synthase; AT, acyl-CoA/ACP transacylase (activates malonyl-CoA unless otherwise noted); mAT, methylmalonyl-CoA/ACP transacylase; DH, dehydratase; ACP, acyl carrier protein; KR, ketoreductase; TE, thioesterase; OR, oxidoreductase; MT, SAM-dependent methyltransferase; NAT, N-acetyl transferase; unk, unknown.

Fig. 3. (A) Representative enediyne structures (B) phylogenetic tree of representative enediyne PKSs. GenBank accessions: MadE, AAQ17110; NcsE, AAM78012; SgcE, AAL06699; CalE8, AAM94794; EspE, AAP92148; DynE, AAN79725. (C) Proposed biosynthesis of sporolide A.

Table 1 - *S. tropica* CNB-440 genome data in comparison to other actinomycete natural product producers

Organism	*S. tropica* CNB-440	*S. coelicolor* A3(2)	*M. tuberculosis* H5N1	*Frankia* sp. CcI3	*N. farcinica* IFM 10152
Size	5.18 Mb	8.72 Mb	4.41 Mb	5.43 Mb	6.01 Mb
Chromosome organization	circular	linear	circular	circular	circular
%G+C content	69.5%	72%	67%	70%	70%
% of genome dedicated to secondary metabolism	~9.5%	~8%	Not determined	Not determined	Not determined
Major NP clusters					
Modular Type I PKS	1	2	7	4	4
Enediyne PKS	2	-	-	-	-
Type II PKS	2	2	1	2	1
Type III PKS	1	3	3*	-	1
Mixed PKS/NRPS	4	-	-	1	1
NRPS	3	3	2	3	7
non-NRPS siderophore	1	1	-	1	-

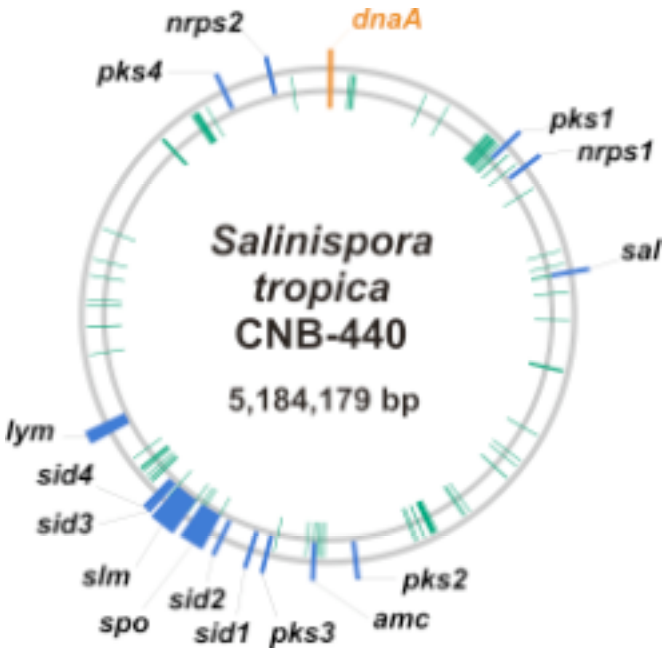
* Two type III PKS enzymes are associated with the modular type I PKSs *pks7*, *pks8*, *pks9*, *pks17*.

Table 2 - *S. tropica* CNB-440 biosynthetic gene clusters

Cluster designation	**Actual (*) or predicted product**	**Type**	**%G +C**	**Size**	**GenBank Accession**
pks1	10-membered enediyne	Enediyne PKS	67	30 kb	xx
pks2	Glycosylated decaketide	Type II PKS	69	35 kb	xx
pks3	Spore pigment	Type II PKS	71	23 kb	xx
spo	Sporolide*	Enediyne PKS	68	71 kb	xx
slm	Salinilactam*	Type I PKS	70	80 kb	xx
pks4	Aromatic polyketide	Type III PKS	73	4 kb	xx
nrps1	Unknown dipeptide	NRPS	68	31 kb	xx
nrps2	Reductively offloaded tetrapeptide	NRPS	70	33 kb	xx
sal	Salinosporamide*	PKS-NRPS	67	41 kb	xx
lym	Lymphostin*	PKS-NRPS	71	33 kb	xx
sid1	Desferrioxamine-like siderophore	NRPS-independent	72	38 kb	xx
sid2	Yersiniabactin-like siderophore	PKS-NRPS	71	28 kb	xx
sid3	Unknown siderophore	PKS-NRPS	72	30 kb	xx

sid4	'Coelichelin'-like siderophore	NRPS	70	10 kb	xx
amc	Unknown aminocyclitol	Aminocyclitol	70	8 kb	xx

Figure 1


nrps2
dnaA
pks4
pks1
nrps1
Salinispora
tropica
CNB-440
5,184,179 bp
sal
lym
sid4
sid3
slm
sid2
spo
sid1
pks3
amc
pks2

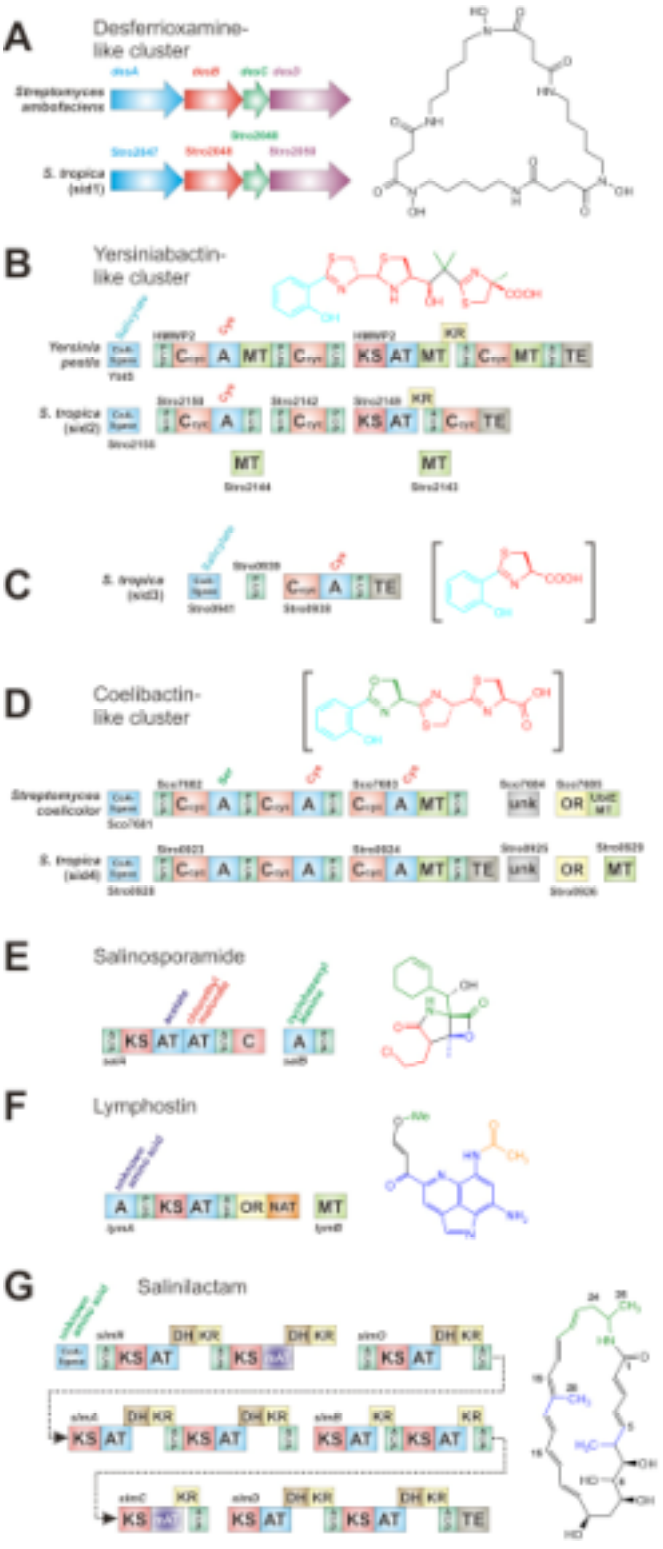
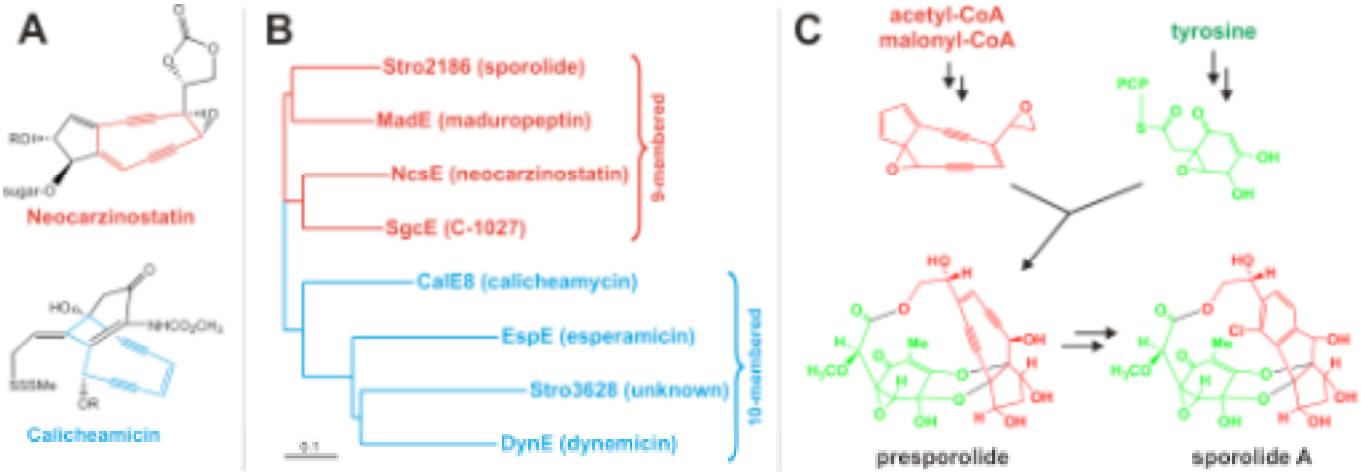

A
Desferrioxamine-like cluster
Streptomyces ambofaciens
S. tropica (sid1)
B
Yersiniabactin-like cluster
Yersinia pestis
S. tropica (sid2)
C
S. tropica (sid3)
D
Coelibactin-like cluster
Streptomyces coelicolor
S. tropica (sid4)
E
Salinosporamide
F
Lymphostin
G
Salinilactam

Figure 2

Figure 3


A
Neocarzinostatin
Calicheamicin
B
Stro2186 (sporolide)
MadE (maduropeptin)
NcsE (neocarzinostatin)
SgcE (C-1027)
9-membered
CalE8 (calicheamycin)
EspE (esperamicin)
Stro3628 (unknown)
DynE (dynemicin)
10-membered
C
acetyl-CoA
malonyl-CoA
tyrosine
PCP
presporolide
sporolide A