

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

Daniel W. Udway\*, Lisa Zeigler\*, Ratnakar N. 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–Lawrence Berkeley National Laboratory, Walnut Creek, CA 94598

Edited by Christopher T. Walsh, Harvard Medical School, Boston, MA, and approved May 7, 2007 (received for review February 1, 2007)

Recent fermentation studies have identified actinomycetes of the marine-dwelling genus *Salinispora* as prolific natural product producers. To further evaluate their biosynthetic potential, we sequenced the 5,183,331-bp *S. tropica* CNB-440 circular genome and analyzed all identifiable secondary natural product gene clusters. Our analysis shows that *S. tropica* dedicates a large percentage of its genome (~9.9%) to natural product assembly, which is greater than previous *Streptomyces* genome sequences as well as other natural product-producing actinomycetes. The *S. tropica* genome features polyketide synthase systems of every known formally classified family, nonribosomal peptide synthetases, and several hybrid clusters. Although a few clusters appear to encode molecules previously identified in *Streptomyces* species, the majority of the 17 biosynthetic loci are novel. Specific chemical information about putative and observed natural product molecules is presented and discussed. In addition, our bioinformatic analysis not only was critical for the structure elucidation of the polyene macrolactam salinilactam A, but its structural analysis aided the genome assembly of the highly repetitive *slm* loci. This study firmly establishes the genus *Salinispora* as a rich source of drug-like molecules and importantly reveals the powerful interplay between genomic analysis and traditional natural product isolation studies.

genome mining | marine bacteria | natural products | polyketide synthase

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 *Streptomyces coelicolor* (2) and *Streptomyces 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 (11). 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 colinearity rule (12) that, 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 (13). Among the strains cultured from marine samples is the genus *Salinispora*, which was recently described as the first seawater-requiring marine actinomycete (14). This genus is widely distributed in tropical and subtropical ocean sediments and currently comprises the two formally described species, *Salinispora tropica* and *Salinispora arenicola*, and a third species for which the name *Salinispora pacifica* has been proposed (15). These actinomycetes are proving to be an exceptionally rich source of structurally diverse secondary metabolites, which are produced in species-specific patterns (16). In the case of *S. tropica*, the compounds observed to date from this bacterium include the potent proteasome inhibitor salinosporamide A (17), which is currently in phase I human clinical trials for the treatment of cancer; the unprecedented halogenated macrolides sporolides A and B (18); lymphostin, which was observed by scientists at Nereus Pharmaceuticals (San Diego, CA) during salinosporamide A fermentation development (R. Lam, personal communication) and was first reported from a *Streptomyces* species (19); and salinilactam, the structure of which was solved as part of the present study.

Here we report all identified secondary metabolic biosynthetic gene clusters derived from the complete genome sequence of the marine actinomycete *S. tropica* strain CNB-440. To our knowledge, 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. Bioinformatic analysis was used to facilitate the structure elucidation of the 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

Author contributions: D.W.U., V.S., A.L., P.R.J., and B.S.M. designed research; D.W.U., R.N.A., V.S., and A.L. performed research; D.W.U., L.Z., R.N.A., V.S., A.L., W.F., P.R.J., and B.S.M. analyzed data; and D.W.U., L.Z., P.R.J., and B.S.M. wrote the paper.

The authors declare no conflict of interest.

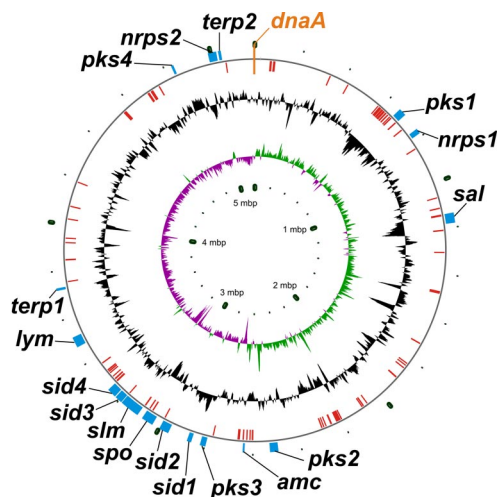
This article is a PNAS Direct Submission.

Abbreviations: KR, ketoreductase; NRPS, nonribosomal peptide synthetase; PKS, polyketide synthase.

Data deposition: The sequence described in this paper has been deposited in the GenBank database (accession no. CP000667).

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

© 2007 by The National Academy of Sciences of the USA



**Fig. 1.** Circular chromosome of *S. tropica* CNB-440, oriented to the *dnaA* gene. The outside outer ring shows the locations of secondary metabolic gene clusters. The inside outer ring shows the locations of putative mobile genetic elements. The center ring shows a normalized plot of GC content (maximum, 75.5%; minimum, 60.3%; average, 69.5%). The inner ring shows a normalized plot of GC skew (maximum, 0.2346; minimum, -0.2504; average, -0.0020).

recent horizontal gene transfer, which has important implications for the 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 composed of 5,183,331 bp, with no plasmids, and an average G+C content of 69.5% (Fig. 1). *S. tropica* has 4,536 predicted protein-coding sequences and is similar in size to other actinomycetes that harbor circular chromosomes [*Mycobacterium tuberculosis* (20), *Frankia* sp. CcI3 (GenBank accession CP000249), and *Nocardia farcinica* (21)] yet is substantially smaller than those with linear chromosomes [*S. coelicolor* (11) and *Rhodococcus* sp. RHA1 (22)] (Table 1).

Biosynthetic gene clusters were initially identified from a draft genome sequence provided by the Joint Genome Institute and checked against a single contiguous sequence provided in advance of public release. Each putative 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 synthetases and cyclases, prenyl transferases, methyl transferases, NAD(P)H-dependent oxidoreductases, and CoA/AMP ligases].

These analyses revealed 17 secondary metabolic biosynthesis

gene clusters that we predict to be involved in siderophore, melanin, polyketide, nonribosomal peptide, terpenoid, and aminocyclitol production (Table 1). The combined length of these gene clusters is estimated at 518 kb, establishing that, among bacteria sequenced to date, *S. tropica* devotes the largest percentage of its genome (~9.9%) 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, 23).

The majority 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 of which reside in close proximity to secondary metabolic clusters (*pks1*, *sal*, *spo*, *amc*, *sid3*, and *slm*). Several clusters also have strong similarity in ORF and domain organization to existing gene clusters (*sid1*, *sid2*, *sid4*, and *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 use 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 cross-talk with other primary or secondary metabolic pathways. *S. tropica* harbors five 4'-phosphopantetheinyl transferase-encoding genes (24) whose products transfer the 4'-phosphopantetheine moiety from CoA to the approximate 35 conserved serine residues in fatty acid synthase-, PKS-, and NRPS-associated carrier proteins. Three transferases are found in identified carrier protein-containing biosynthetic clusters (Stro0685 with *nrps1*, Stro2496 with *pks3*, and Stro2822 with *sid4*) and thus may be specific to their pathways.

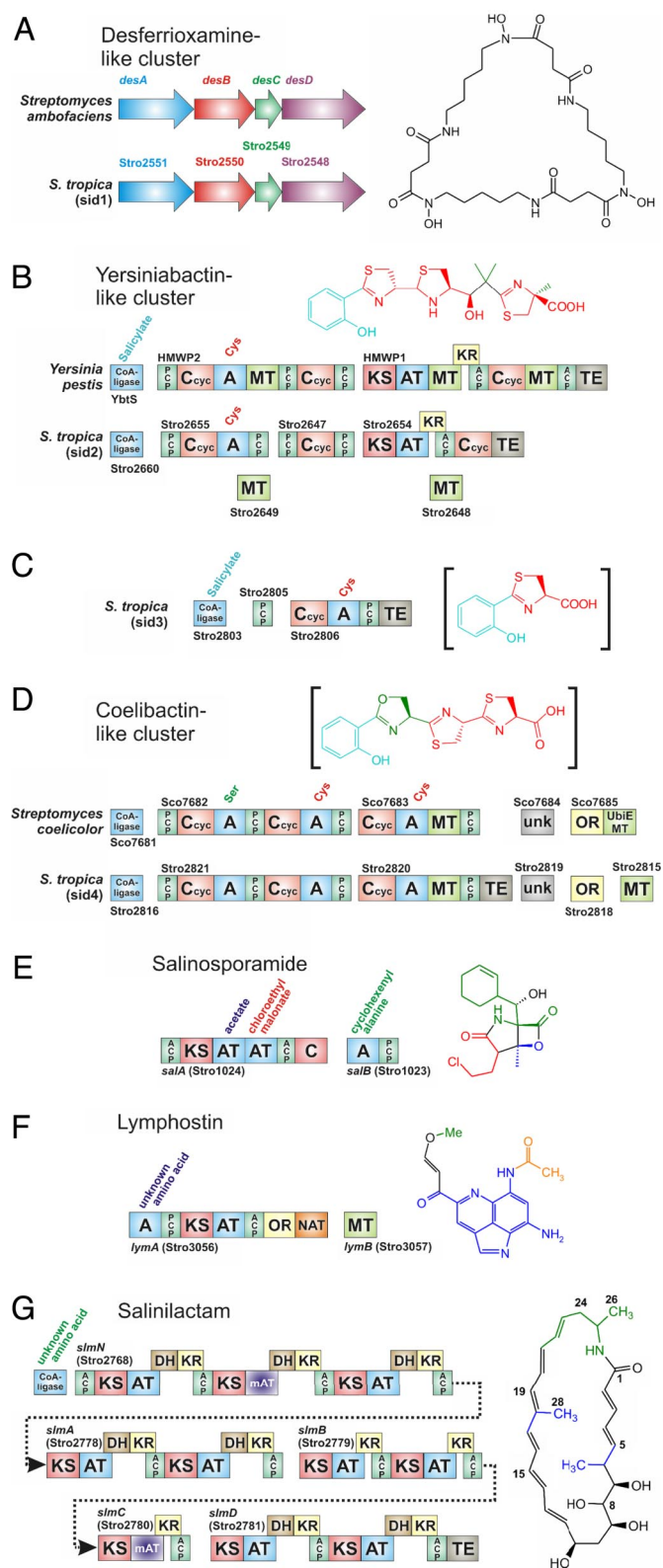
A striking feature of the *S. tropica* genome is its remarkable diversity of polyketide biosynthetic pathways, which is greater in variety than in other sequenced bacterial genomes (Table 1). Pathways include those for modular type I PKSs (*slm*), iterative enediyne type I PKSs (*spo* and *pks1*), hybrid 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 15 type I PKS-associated modules contains the full set of reductive domains required to completely reduce the transient  $\beta$ -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).

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

Organism	Size, Mb	Chr. org.	%G+C content	%GDSM	Major NP clusters, n							
					PKS				Mixed PKS/NRPS	NRPS	Non-NRPS siderophore	
					Modular type I	Enediyne	Type II	Type III				
<i>S. tropica</i> CNB-440	5.18	C	69.5	≈9.9	1	2	2	1	4	3	1	
<i>S. coelicolor</i> A3(2)	8.72	L	72	≈8	2	—	2	3	—	3	1	
<i>M. tuberculosis</i> H5N1	4.41	C	67	ND	7	—	1	3*	—	2	—	
<i>Frankia</i> sp. CcI3	5.43	C	70	ND	4	—	2	—	1	3	1	
<i>N. farcinica</i> IFM 10152	6.01	C	70	ND	4	—	1	1	1	7	—	

NP, nonribosomal peptide; Mb, megabases; Chr. org., chromosome organization; %GDSM, percentage of the genome dedicated to secondary metabolism; C, circular; L, linear; ND, not determined; —, not applicable.

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

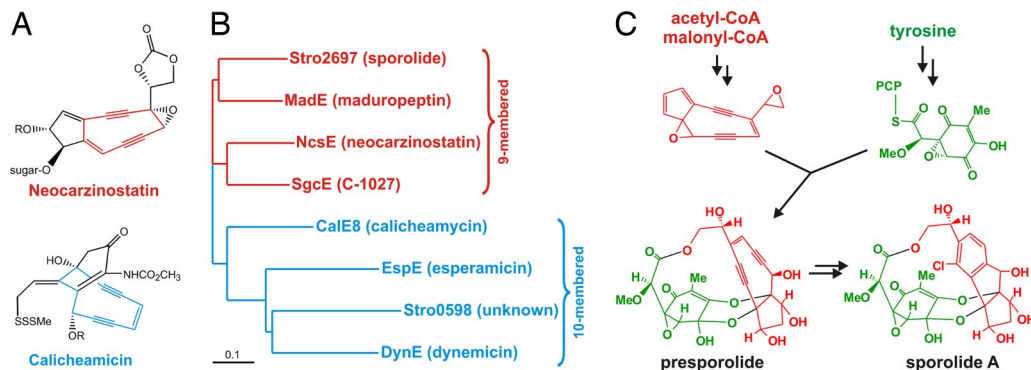


**Fig. 2.** Selected genes from *S. tropica* modular biosynthetic enzyme systems and their associated natural products. (A–D) Putative siderophore clusters. (A) *sid1* cluster in comparison to the desferrioxamine cluster from *Streptomyces coelicolor* A3 (2). (B) *sid2* cluster in comparison to the yersiniabactin cluster from *Y. 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) *slmNABCD* modular PKS system from proposed salinilactam cluster (stereo-

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 (17). Its biosynthetic gene cluster (*sal*) encodes a 29-ORF, 41-kb hybrid PKS–NRPS pathway (Fig. 2E) that involves mechanisms of chlorination and  $\beta$ -lactone synthesis as well as 20S proteasome resistance (A. S. Eustáquio, L. Beer, D.W.U., and B.S.M., unpublished data). 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 (19). We propose that the uniquely organized two-module synthetase LymA (Stro3056) catalyzes a single polyketide extension followed by reductive offloading and *N*-acetylation of a nonproteinogenic tricyclic planar amino acid derived from tryptophan, with subsequent *O*-methylation catalyzed by the adjacent SAM-dependent methyl transferase LymB (Stro3057) (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 (although lacks direct sequence similarity) to the yersiniabactin cluster (*ybt*) from *Yersinia pestis* (25) 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/nonribosomal peptide core with possible alterations in sites of methylation and/or oxidation.

Two independent enediynes 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* (26). The enediyne polyketide structural motif falls into two classes that possess either nine- 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 nine- and 10-membered enediyne polyketides, respectively (Fig. 3). Although no enediyne natural products have yet been directly characterized from *S. tropica*, sporolides A and B (18) are putatively derived from a nine-membered enediyne PKS product that is uniquely halogenated with chloride during the aromatization of the enediyne unit to yield the tricyclic hydrocarbon nucleus, as proposed for the related cyanosporasides from “*S. pacifica*” (27) and recently evaluated synthetically with model substrates (28). The  $\approx$ 50-kb *spo* cluster harbors as many as 39 ORFs, which include other conserved enediyne PKS-associated genes, as well as those that encode the postulated biosynthesis and coupling of the sporolide cyclohexenone unit derived from the amino acid tyrosine (Fig. 3). 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 or 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.

chemistry 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,  $\beta$ -ketoacyl synthase; AT, acyl-CoA/ACP transacylase (activates malonyl-CoA unless otherwise noted); mAT, methylmalonyl-CoA/ACP transacylase; DH, dehydratase; ACP, acyl carrier protein; TE, thioesterase; OR, oxidoreductase; MT, SAM-dependent methyltransferase; NAT, *N*-acetyl transferase; unk, unknown.



**Fig. 3.** Enediynes polyketide biochemistry in *S. tropica*. (A) Representative enediynes structures. (B) Phylogenetic tree of representative enediynes PKSs. GenBank accession nos.: MadE, AAK17110; NcsE, AAM78012; SgcE, AAL06699; CalE8, AAM94794; EspE, AAP92148; DynE, AAN79725. (C) Proposed biosynthesis of sporolide A.

In addition to the abundance of diverse PKS pathways, *S. tropica* harbors four NRPS operons (*nrps1*, *nrps2*, *sid3*, and *sid4*) (Fig. 1). Although the products of the dimodular *nrps1* and tetramodular *nrps2* gene clusters are unknown and could not be accurately predicted on the basis of bioinformatics alone (29, 30), *sid3* and *sid4* are homologous to gene sets that encode known siderophore pathway biosynthetic enzymes. When combined with the highly homologous *S. coelicolor* NRPS-independent siderophore desferrioxamine (31) 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*–*sid4* 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 (32). The NRPS from the *sid4* gene cluster shares unmistakable domain organization 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 extends and cyclizes an aromatic acid with cysteine, tentatively producing a siderophore-like intermediate (Fig. 2C). It is unclear whether *sid3* is fully functional or is instead biosynthetically linked with *sid4*. Although the latter three gene clusters are traditionally found in  $\gamma$ -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 that encode a 10-module PKS with 49 domains. Because of 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 compounds with characteristic UV chromophores associated with polyene units, which led to the isolation of a series of polyene macrolactams exemplified by salinilactam A (Fig. 2G). NMR and MS characterization of salinilactam A quickly revealed a polyene macrolactam framework that was consistent only 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 by the chemical instability of the compound and the presence of eight conjugated olefins with similar NMR properties, which impeded their accurate assignment.

Inspection of the structure fragments, together with the molecular formula  $C_{28}H_{39}NO_5$ , revealed by high-resolution MS, suggested that salinilactam A was derived from a PKS with at least 10 extension modules. This information was useful to help resolve and properly assemble the repetitive DNA sequences associated with the *slm* PKS into two operons (Stro2778–2781 and Stro2768) separated by nine accessory genes presumed to be involved in starter unit biosynthesis and macrolactam hydroxylation (Fig. 2G). On the basis of the colinearity rule of modular polyketide biosynthesis, the relationship of *slm* to the terminal region of the vicenistatin PKS (33), and partial NMR-based structural fragments we were able to accurately predict the gross chemical structure of salinilactam A. Initial assignment of the C-28 methyl group in the middle of the hexaene unit at C-18 was complicated by overlapping olefinic NMR signals and was guided by bioinformatics based on the associated methylmalonyl-CoA specific acyl transferase (AT) domain in SlmN (Stro2768) and the absence of  $\beta$ -methylation machinery (34). The natural product would thus be produced by chain extension of the lysine-derived PKS starter unit 5-aminohex-2-enoate with eight malonate and two methylmalonate units, followed by macrolactamization and cytochrome P450 hydroxylation at C-8.

The salinilactam A structure was verified by comprehensive NMR analyses of the purified natural product to confirm the bioinformatics-based total 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, on the basis of the strong homology of all *slm* ketoreductase (KR) domains to other “A-type” KRs (35). The intimate interplay between microbial genomics, biosynthetic logic, and natural product chemistry was critical not only in the structural 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 (36). 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 by using small quantities of material. A relatively new concept that can be added to this list is the molecular genetics of natural product biosynthesis (37). Ad-

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

Cluster designation	Actual (*) or predicted product	Type	%G+C	Size, kb	Gene cluster location
<i>pks1</i>	10-membered enediyne	Enediyne PKS	67	30	Stro0586–0610
<i>pks2</i>	Glycosylated decaketide	Type II PKS	69	35	Stro2204–2235
<i>pks3</i>	Spore pigment	Type II PKS	71	23	Stro2486–2509
<i>spo</i>	Sporolide*	Enediyne PKS	68	50	Stro2695–2733
<i>slm</i>	Salinilactam*	Type I PKS	70	80	Stro2757–2781
<i>pks4</i>	Aromatic polyketide	Type III PKS	73	10	Stro4262–4271
<i>nrps1</i>	Unknown dipeptide	NRPS	68	31	Stro0672–0699
<i>nrps2</i>	Reductively offloaded tetrapeptide	NRPS	70	33	Stro4409–4429
<i>sal</i>	Salinosporamide*	PKS–NRPS	67	41	Stro1012–1043
<i>lym</i>	Lymphostin*	PKS–NRPS	71	33	Stro3042–3066
<i>sid1</i>	Desferrioxamine-like siderophore	NRPS-independent	72	18	Stro2539–2554
<i>sid2</i>	Yersiniabactin-like siderophore	PKS–NRPS	71	28	Stro2636–2660
<i>sid3</i>	Unknown siderophore	PKS–NRPS	72	30	Stro2786–2813
<i>sid4</i>	"Coelibactin"-like siderophore	NRPS	70	44	Stro2814–2842
<i>amc</i>	Unknown aminocyclitol	Aminocyclitol	70	8	Stro2340–2346
<i>terp1</i>	Unknown terpene	Terpene	72	10	Stro3244–3253
<i>terp2</i>	Unknown carotenoid	Terpene	70	12	Stro4437–4445

vances 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 latter, complete genome sequencing provides unparalleled access to the genes involved in secondary metabolism; the means of their assembly; and, in some cases, what products they may yield (2, 3). Armed with these data, it becomes possible to compare the compounds observed by using traditional fermentation procedures with those 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 (8, 9). The genus *Salinispora* has become a model organism by which to address questions about species-specific patterns of natural product production (16) 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* revealed an abundance of novel biosynthetic gene clusters, the majority of which were unexpected on the basis of 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, because our analysis clearly confirms the value of the *Salinispora* genus as a source of novel drug-like molecules. In the absence of other *Micromonosporacea* genome sequences, it is unclear whether this natural product diversity is general to this family or whether the diversity in *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 marine actinomycete genus and address marine adaptation features in these sediment-dwelling actinomycetes.

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 polyene. We expect that the genome information 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 (A. S. Eustáquio, L. Beer, D.W.U., and B.S.M., unpublished data).

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 (38) and makes clear that the instability found at the terminal ends of *Streptomyces* linear chromosomes has been used 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 with *S. tropica* CNB-440.

## Materials and Methods

**Genome Sequencing, Annotation, and Analysis.** Draft sequencing and automated annotation were provided by the Department of Energy, Joint Genome Institute, under the Community Sequencing Program, and completed sequence and annotations have been deposited with GenBank (accession no. CP000667). Automated gene prediction and functional annotation was performed by the Genome Analysis Pipeline (39). Automated prediction of protein-coding genes was based on the output of CRITICA (40) complemented with the output of Glimmer (41). Initial functional annotation was performed by means of similarity searches against the TIGRFAM (42), PRIAM (43), Pfam (44), SMART (45), COGs (46), SwissProt/TrEMBL (47), and KEGG (48) databases by 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 also available at the Microbial Portals (<http://genome.ornl.gov/microbial>) and through the Integrated Microbial Genomes system (<http://img.jgi.doe.gov>) (49).

To more specifically identify and categorize natural product biosynthetic gene clusters, we wrote a custom Perl script 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 with software available at the National Center for Biotechnology Information web site and visualized with Vector NTI (Invitrogen). Transposons and other mobile genetic element ORFs were identified in a similar manner. Suggested boundaries of all gene clusters discussed are cataloged in Table 2.

**Salinilactam A Isolation and Characterization.** Strain CNB-440/ $\Delta$ salD was cultured in nine 1-liter Fernbach flasks containing A1bFe+C medium [10 g of starch/4 g of yeast extract/2 g of peptone/1 g of CaCO<sub>3</sub>/5 ml of 8 g/liter Fe<sub>2</sub>(SO<sub>4</sub>)<sub>3</sub>·4H<sub>2</sub>O/5 ml of 20 g/liter KBr/1 liter of seawater] at 27°C with shaking at 230 rpm. XAD-7 resin (20 g) was added to each flask after 24 h, and the fermentation was continued for another 6 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 reversed phase-HPLC (Prep Nova-Pak HR C18; 6  $\mu$ m; 60 Å; 300 mm  $\times$  40 mm; flow rate of 10 ml/min; detection at 210 nm; 25% for 10 min then a linear gradient up to 40% CH<sub>3</sub>CN 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)  $\lambda_{max}$  (log  $\epsilon$ ) 250 (4.01), 305 (4.51), 335 (4.26), 355 (4.18) nm; IR (NaCl)  $\lambda_{max}$  3,464, 2,930, 1,739, 1,547, 1,414, 1,279, 1,135, 1,074, 998, and 743 cm<sup>-1</sup>; <sup>1</sup>H NMR (500 MHz, CD<sub>3</sub>OD)  $\delta$  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, dd,  $J$  = 11.2, 4.1 Hz, H-11), 3.97 (1H, ddd,  $J$  = 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<sub>3</sub>-28), 1.78 (1H, dd,  $J$  = 14.8, 6.8 Hz, H-10b), 1.27 (3H, d,  $J$  = 7.1 Hz, H<sub>3</sub>-26), 1.08 (3H, d,  $J$  = 7.2 Hz, H<sub>3</sub>-27) [coupling constants derived from analysis of homo- $J$ -resolved <sup>1</sup>H NMR spectral data]; <sup>13</sup>C NMR (125 MHz, CD<sub>3</sub>OD)  $\delta$  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<sub>2</sub>O + H), 470 (M + H), 492 (M + Na), 961 (2M + Na); HRESITOFMS  $m/z$  492.2720 [calculated for C<sub>28</sub>H<sub>39</sub>NO<sub>5</sub>Na<sup>+</sup>, 492.2726].

We thank G. L. Challis (University of Warwick) and R. P. McGlinchey for valuable discussions, R. Lam and G. Tsung (Nereus Pharmaceuticals) for providing the initial material that was used for the preliminary isolation and structural elucidation of salinilactam A, D. C. Oh for preliminary characterization of salinilactam A, and A. S. Eustáquio (University of California at San Diego) for strain CNB-440/ $\Delta$ salD. This work was supported by the National Oceanic and Atmospheric Administration through Oceans and Human Health Initiative Grant NA05NOS4781249 (to B.S.M. and P.R.J.) and National Institutes of Health Grants R01 CA127622 (to B.S.M.) and R37 CA44848 (to W.F.). Genome sequencing was provided through a grant from the Community Sequencing Program of the Joint Genome Institute (to P.R.J. and B.S.M.).

- Berdy J (2005) *J Antibiot* 58:1–26.
- Bentley SD, Chater KF, Cerdeno-Tarraga AM, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D, et al. (2002) *Nature* 417:141–147.
- 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.
- Lautru S, Deeth RJ, Bailey LM, Challis GL (2005) *Nat Chem Biol* 1:244–245.
- Song L, Barona-Gomez F, Corre C, Xiang L, Udway DW, Austin MB, Noel JP, Moore BS, Challis GL (2006) *J Am Chem Soc* 128:14754–14755.
- Bok JW, Hoffmeister D, Maggio-Hall LA, Murillo R, Glasner JD, Keller NP (2006) *Chem Biol* 13:31–37.
- Gross H, Stockwell VO, Henkels MD, Nowak-Thompson B, Loper JE, Gerwick WH (2007) *Chem Biol* 14:53–63.
- Zazopoulos E, Huang K, Staffa A, Liu W, Bachmann BO, Nonaka K, Ahlert J, Thorson JS, Shen B, Farnet CM (2003) *Nat Biotechnol* 21:187–190.
- McAlpine JB, Bachmann BO, Piraece M, Tremblay S, Alarco AM, Zazopoulos E, Farnet CM (2005) *J Nat Prod* 68:493–496.
- Fischbach MA, Walsh CT (2006) *Chem Rev* 106:3468–3496.
- Martin JF (1992) *J Ind Microbiol* 9:73–90.
- Staunton J, Weissman KJ (2001) *Nat Prod Rep* 18:380–416.
- Fenical W, Jensen PR (2006) *Nat Chem Biol* 2:666–673.
- Maldonado L, Fenical W, Goodfellow M, Jensen PR, Ward AC (2005) *Int J System Appl Microbiol* 55:1759–1766.
- Jensen PR, Mafnas C (2006) *Environ Microbiol* 8:1881–1888.
- Jensen PR, Williams PG, Oh D-C, Zeigler L, Fenical W (2007) *Appl Environ Microbiol* 73:1146–1152.
- Feling RH, Buchanan GO, Mincer TJ, Kauffman CA, Jensen PR, Fenical W (2003) *Angew Chem Int Ed* 115:369–371.
- Buchanan GO, Williams PG, Feling RH, Kauffman CA, Jensen PR, Fenical W (2005) *Org Lett* 7:2731–2734.
- Aotani Y, Nagata H, Yoshida M (1997) *J Antibiot* 50:543–545.
- Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE, et al. (1998) *Nature* 393:537–544.
- Ishikawa J, Yamashita A, Mikami Y, Hoshino Y, Kurita H, Hotta K, Shiba T, Hattori M (2004) *Proc Natl Acad Sci USA* 101:14925–14930.
- McLeod MP, Warren RL, Hsiano WWL, Araki N, Myhre M, Fernandes C, Miyazawa D, Wong W, Lillquist AL, Wang D, et al. (2006) *Proc Natl Acad Sci USA* 103:15582–15587.
- 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.
- Lambalot RH, Gehring AM, Flugel RS, Zuber P, LaCelle M, Marahiel MA, Reid R, Khosla C, Walsh CT (1996) *Chem Biol* 3:923–936.
- Gehring AM, DeMoll E, Fetherston JD, Mori I, Mayhew GF, Blattner FR, Walsh CT, Perry RD (1998) *Chem Biol* 5:573–586.
- Shen B, Liu W, Nonaka K (2003) *Curr Med Chem* 10:2317–2325.
- Oh DC, Williams PG, Kauffman CA, Jensen PR, Fenical W (2006) *Org Lett* 8:1021–1024.
- Perrin CL, Rodgers BL, O'Connor JM (2007) *J Am Chem Soc* 129:4795–4799.
- Stachelhaus T, Mootz HD, Marahiel MA (1999) *Chem Biol* 6:493–505.
- Challis GL, Ravel J, Townsend CA (2000) *Chem Biol* 7:211–224.
- Barona-Gomez F, Wong U, Giannakopoulos AE, Derrick PJ, Challis GL (2004) *J Am Chem Soc* 126:16282–16283.
- Crosa JH, Walsh CT (2002) *Microbiol Mol Biol Rev* 66:223–249.
- Ogasawara Y, Katayama K, Minami A, Otsuka M, Eguchi T, Kakinuma K (2004) *Chem Biol* 11:79–86.
- Calderone CT, Kowtoniuk WE, Kelleher NL, Walsh CT, Dorrestein PC (2006) *Proc Natl Acad Sci USA* 103:8977–8982.
- Siskos AP, Baerga-Ortiz A, Bali S, Stein V, Mamdani H, Spitteller D, Popovic B, Spencer JB, Staunton J, Weissman KJ, et al. (2005) *Chem Biol* 12:1145–1153.
- Fenical W, Walsh C (2004) *Nature* 432:829–837.
- Fischbach MA, Walsh CT (2006) *Chem Rev* 106:3468–3496.
- Choulet F, Aigle B, Gallois A, Mangenot S, Gerbaud C, Truong C, Francou FX, Fourrier C, Guerineau M, Decaris B, et al. (2006) *Mol Biol Evol* 23:2361–2369.
- Hauser L, Larimer F, Land M, Shah M, Uberbacher E (2004) *Gen Engineer* 26:225–238.
- Badger JH, Olsen GJ (1999) *Mol Biol Evol* 16:512–524.
- Delcher AL, Harmon D, Kasif S, White O, Salzberg SL (1999) *Nucleic Acids Res* 27:4636–4641.
- Haft DH, Loftus BJ, Richardson DL, Yang F, Eisen JA, Paulsen IT, White O (2001) *Nucleic Acids Res* 29:41–43.
- Claudet-Renard C, Chevalet C, Faraut T, Kahn D (2003) *Nucleic Acids Res* 31:6633–6639.
- Bateman A, Birney E, Cerruti L, Durbin R, Ewinger L, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M, Sonnhammer EL (2002) *Nucleic Acids Res* 30:276–280.
- Letunic I, Copley RR, Pils B, Pinkert S, Schultz J, Bork P (2006) *Nucleic Acids Res* 34:D257–D260.
- Tatusov RL, Galperin MY, Natale DA, Koonin EV (2000) *Nucleic Acids Res* 28:33–36.
- Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I, et al. (2003) *Nucleic Acids Res* 31:365–370.
- Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M (2004) *Nucleic Acids Res* 32:D277–D280.
- Markowitz VM, 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–D348.