Sequence-Based Analysis of Secondary-Metabolite Biosynthesis in Marine Actinobacteria^{∇‡}

Erin A. Gontang,[†] Susana P. Gaudêncio, William Fenical, and Paul R. Jensen*

Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography,
University of California, San Diego, La Jolla, California 92093-0204

Received 25 November 2009/Accepted 3 February 2010

A diverse collection of 60 marine-sediment-derived *Actinobacteria* representing 52 operational taxonomic units was screened by PCR for genes associated with secondary-metabolite biosynthesis. Three primer sets were employed to specifically target adenylation domains associated with nonribosomal peptide synthetases (NRPSs) and ketosynthase (KS) domains associated with type I modular, iterative, hybrid, and enediyne polyketide synthases (PKSs). In total, two-thirds of the strains yielded a sequence-verified PCR product for at least one of these biosynthetic types. Genes associated with enediyne biosynthesis were detected in only two genera, while 88% of the ketosynthase sequences shared greatest homology with modular PKSs. Positive strains included representatives of families not traditionally associated with secondary-metabolite production, including the *Corynebacteriaceae*, *Gordoniaceae*, *Intrasporangiaceae*, and *Micrococcaceae*. In four of five cases where phylogenetic analyses of KS sequences revealed close evolutionary relationships to genes associated with experimentally characterized biosynthetic pathways, secondary-metabolite production was accurately predicted. Sequence clustering patterns were used to provide an estimate of PKS pathway diversity and to assess the biosynthetic richness of individual strains. The detection of highly similar KS sequences in distantly related strains provided evidence of horizontal gene transfer, while control experiments designed to amplify KS sequences from *Salinispora arenicola* strain CNS-205, for which a genome sequence is available, led to the detection of 70% of the targeted PKS pathways. The results provide a bioinformatic assessment of secondary-metabolite biosynthetic potential that can be applied in the absence of fully assembled pathways or genome sequences. The rapid identification of strains that possess the greatest potential to produce new secondary metabolites along with those that produce known compounds can be used to improve the process of natural-product discovery by providing a method to prioritize strains for fermentation studies and chemical analysis.

Microbial natural products represent the primary resource from which new medicines are derived, accounting for approximately half of the antibiotics discovered as of 2002 (6). Over the past several decades, however, drug discovery efforts have moved away from microbial products (29), in part due to a reduction in the ratio of new chemical entities discovered relative to the isolation of known metabolites (3) and the challenges associated with developing effective “dereplication” methods to improve the efficiency of the discovery process. More recently, the rise in drug-resistant pathogens has left many current antibiotics obsolete, while the limited success of alternative discovery strategies, such as combinatorial chemistry, have created a void in the pipeline of new drug leads (39). The response to this need includes renewed interest in a group of actinobacteria commonly called actinomycetes (defined here as bacteria within the order *Actinomycetales*), which have long been recognized as a prolific source of natural products, including polyketides, nonribosomal peptides, and combina-

tions thereof (17). Although actinomycetes have been studied extensively in the past, it is estimated that only 3% of the natural-product potential of even the well-studied genus *Streptomyces* has been realized (46), thus leaving considerable opportunity for new discovery. In addition, strains derived from poorly studied environments, including marine samples (8, 15, 33), have proven to be a productive source of new compounds. This potential has helped drive calls for the development of new approaches to natural-product discovery that include minimizing the isolation of previously described compounds (4). Advances in our understanding of the molecular genetics of natural-product biosynthesis coupled with increasing access to DNA sequencing now create unparalleled opportunities to incorporate sequence-based approaches into the process of natural-product discovery.

Polyketides, nonribosomal peptides, and polyketide/nonribosomal peptide hybrids are synthesized by the coordinated actions of enzymatic assembly lines, which conduct the iterative chemical condensation of monomeric units, including carboxylic acid and/or amino acid monomers (10, 17, 45). The order and identity of each domain within a given assembly line specify the sequence of monomer activation and incorporation, the chemical reactions that occur at each step in the assembly process, and the length and functionality of the product released (17). In the case of type I polyketide synthases (PKSs), the condensation of a carboxylic acid monomer to a growing acyl chain is accomplished via ketosynthase (KS), acyltransferase, and acyl carrier protein domains that act either itera-

* Corresponding author. Mailing address: Scripps Institution of Oceanography, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0204. Phone: (858) 534-7322. Fax: (858) 534-1318. E-mail: pjensen@ucsd.edu.

[†] Present address: Department of Microbiology and Molecular Genetics, Harvard Medical School, 200 Longwood Ave., Boston, MA 02115.

[‡] Supplemental material for this article may be found at <http://aem.asm.org/>.

[∇] Published ahead of print on 12 February 2010.

tively, for the biosynthesis of aromatic polyketides, or, more commonly, in a nonredundant, assembly-line fashion. Following each condensation reaction, the presence of ketoreductase, dehydratase, and enoylreductase domains determines the oxidation state of the beta-carbonyl (25). The incorporation of amino acid monomers into peptides via nonribosomal peptide synthetases (NRPSs) is accomplished via the condensation, adenylation (A), and peptidyl carrier protein domains associated with the enzyme (17). When present, epimerization, methyltransferase, and oxidase domains tailor the amino acid being added. The immense structural complexity and functional diversity of polyketides, nonribosomal peptides, and compounds of hybrid biosynthetic origin are generated both by the intricately coordinated organization of these enzymatic assembly lines and by the specific tailoring enzymes responsible for postassembly modifications (10, 17).

In an effort to identify new sources of bioactive secondary metabolites, degenerate PCR primers have been used to screen for the presence of genes associated with PKS and NRPS pathways in DNA derived from soil (18, 47, 48), sponge tissues (16, 41), and a variety of cultured organisms, including cyanobacteria (9, 14, 37), dinoflagellates (37), and Gram-positive bacteria (1, 2, 21, 23, 28). When coupled with homology-based searches and phylogenetic analyses (18), sequence-based approaches offer an opportunity to predict which isolates or environments harbor the greatest potential to produce interesting new secondary metabolites. Phylogenetic analyses are particularly useful in that clustering patterns can be used to predict the potential number of biosynthetic pathways in a strain and if the biosynthetic logic of the PKS gene is modular, iterative, or typical of a PKS-NRPS hybrid (19, 25). In the case of modular organization, which is frequently the result of duplication events (22, 24), KS loci from the same type I PKS can often be recognized as a cluster of closely related sequences. Although KS domains within the same biosynthetic pathway are not always in the same cluster, those from unrelated strains that produce the same product generally have similar clustering patterns. By following this logic, a KS sequence from a chemically unknown strain that shows a high level of sequence identity to an experimentally characterized pathway can be used to predict that the unknown strain has the genetic potential to produce secondary metabolites related to those previously reported from that pathway (see, e.g., reference 28). This predictive capability provides a rapid method to avoid the isolation of known compounds or to identify strains that produce compounds within a desired structural class. Likewise, KS sequences that do not cluster with characterized biosynthetic pathways will have a greater probability of yielding new secondary metabolites, while the number of distinct KS sequence clusters provides an estimate of the maximum number of PKS pathways that an individual strain may possess.

By a PCR-based approach, cultured strains from 52 *Actinomyces* operational taxonomic units (OTUs) recovered from marine sediments collected in the Republic of Palau were screened for the presence of A domain sequences associated with NRPS pathways and KS loci associated with modular, iterative, hybrid (PKS-NRPS), and enediyne PKSs (type I). Strains from families that are well known to produce secondary metabolites were screened, and in addition, taxa not generally associated with the biosynthesis of these compounds were also

evaluated. Phylogenetic analyses were used to assess the similarity of KS sequences to those associated with experimentally characterized pathways and, in five cases where close matches were detected, to predict the types of products that would be produced. The results reveal that pathways associated with secondary-metabolite biosynthesis are widely distributed among the *Actinomycetales* and that bioinformatic analyses provide a powerful method of dereplication that has considerable potential to improve the efficiency with which secondary metabolites are discovered.

MATERIALS AND METHODS

Nucleic acid extraction, PCR amplification, cloning, and sequencing. A total of 60 marine-sediment-derived actinomycete strains, representing 52 OTUs cultured during a research expedition to the Republic of Palau (20), were examined. Genomic DNA was extracted from each isolate according to the DNeasy protocol (Qiagen Inc., Valencia, CA), with several modifications. After RNase A (2 mg/ml) was added to the enzymatic lysis buffer, the resuspended bacterial pellet was incubated for 2 h at 37°C. Following the addition of proteinase K and lysis buffer AL, the sample was held for 1 h at 70°C. Genomic DNA was eluted from the spin column with 100 µl of elution buffer AE for immediate use or storage at -20°C.

Ketosynthase (KS) domains of type I polyketide synthase (PKS) genes were PCR amplified from genomic DNA using the primers KS-F (5'-CCSCAGSAG CGCSTSYTCTSGA-3') and KS-R (5'-GTSCCSGTSCCGTSGYSTCSA-3') (18). These degenerate primers were designed to amplify KS domains associated with modular, iterative, and NRPS hybrid type I PKS genes. The 50-µl PCR mixture contained 20 to 50 ng of DNA, 800 pmol of each primer, 10× PCR buffer II (Applied Biosciences, Foster City, CA), 2.5 mM MgCl₂ (Applied Biosciences, Foster City, CA), 1.5 U of AmpliTaq Gold DNA polymerase (Applied Biosciences, Foster City, CA), 400 µM deoxynucleoside triphosphate mixture, and 7% dimethyl sulfoxide (DMSO). The PCR protocol consisted of a 15-min denaturation at 95°C; 1 cycle of 1 min at 95°C, 1 min at 65°C, and a 1-min extension at 72°C; 35 cycles of 1 min at 95°C, 1 min at 62°C, and a 1-min extension at 72°C; followed by 10 min at 72°C.

Due to a high level of sequence divergence, a distinct set of type I PKS primers were used to screen for sequences associated with the biosynthesis of the core ("warhead") portion of enediyne secondary metabolites. The primers EdyA (5'-CCCCGVCACATCACSGSCCTCGCSGTGAACATGCT-3') and EdyE (5'-GCAGGCKCCGTCSACSGTGTABCCGCCGCC-3') (31) targeted the N-terminal region of the enediyne PKS gene, inclusive of the KS domain. The 50-µl PCR mixture contained 20 to 50 ng of DNA, 500 pmol of each primer, 10× PCR buffer II (Applied Biosciences, Foster City, CA), 1.5 mM MgCl₂ (Applied Biosciences, Foster City, CA), 1 U of AmpliTaq Gold DNA polymerase (Applied Biosciences, Foster City, CA), 200 µM deoxynucleoside triphosphate mixture, and 5% DMSO. PCR amplification was performed according to the following protocol: a 5-min denaturation at 94°C, 30 cycles of 45 s at 94°C, 90 s at 62°C, and a 2-min extension at 72°C, followed by 7 min at 72°C.

Adenylation domains of NRPS pathways were PCR amplified using the primers NRPS-A3 (5'-GCSTACSYSATSTACACSTCSGG-3') and NRPS-A7R (5'-SASGTCVCCSGTSCGGTAS-3') (2). The 50-µl PCR mixture contained 20 to 50 ng of DNA, 400 pmol of each primer, 10× PCR buffer II (Applied Biosciences, Foster City, CA), 2.5 mM MgCl₂ (Applied Biosciences, Foster City, CA), 1.5 U of AmpliTaq Gold DNA polymerase (Applied Biosciences, Foster City, CA), 400 µM deoxynucleoside triphosphate mixture, and 10% DMSO. PCR amplification of NRPS adenylation domains was performed according to the following PCR protocol: a 5-min denaturation at 95°C, 35 cycles of 30 s at 95°C, 90 s at 59°C, and a 1-min extension at 72°C, followed by 10 min at 72°C. All PCR mixtures were examined by gel electrophoresis, and bands corresponding to KS domains (expected size, ~670 bp), N-terminal regions of enediyne PKS pathways (expected size, ~1.4 kb), and adenylation domains (expected size, ~700 bp) were excised using a sterile scalpel and purified using Qiagen's QIAquick gel extraction kit according to the manufacturer's suggested protocol (Qiagen Inc., Valencia, CA).

All correctly sized PCR products were immediately cloned using the TOPO TA cloning kit (Invitrogen Co., Carlsbad, CA) according to the manufacturer's suggested protocol with the following modifications. The 6-µl TOPO cloning reaction mixture consisted of 1 µl fresh PCR product, 1 µl salt solution, 3 µl sterile water, and 1 µl TOPO vector. The reactions were mixed gently and allowed to incubate at room temperature for 15 min. Once the ligation reaction

TABLE 1. PCR screening of marine actinobacteria

Family	Genus	OTU	Strain (NCBI accession no.)	PKS		NRPS
				Type I ^a	Enediyne	
<i>Brevibacteriaceae</i>	<i>Brevibacterium</i>	1	CNJ737 (DQ448693)	—	—	—
<i>Corynebacteriaceae</i>	<i>Corynebacterium</i>	2	CNJ954 (DQ448694)	—	—	+
<i>Dermacoccaceae</i>	<i>Kytococcus</i>	3	CNJ855 (DQ448695)	—	—	—
<i>Dietziaceae</i>	<i>Dietzia</i>	4	CNJ898 (DQ448696)	—	—	—
<i>Geodermatophilaceae</i>	<i>Blastococcus</i>	5	CNJ868 (DQ448697)	—	—	—
	<i>Modestobacter</i>	6	CNJ793 (DQ448698)	—	—	—
	<i>Modestobacter</i>	6	CNJ794 (DQ448774)	—	—	—
<i>Gordoniaceae</i>	<i>Gordonia</i>	7	CNJ756 (DQ448699)	—	—	+
	<i>Gordonia</i>	8	CNJ863 (DQ448700)	+	—	+
	<i>Gordonia</i>	9	CNJ754 (DQ448701)	—	—	+
	<i>Gordonia</i>	10	CNJ752 (DQ448702)	—	—	+
<i>Intrasporangiaceae</i>	<i>Ornithinimicrobium</i>	11	CNJ824 (DQ448703)	—	—	—
	<i>Serinicoccus</i>	12	CNJ927 (DQ448704)	+	—	—
<i>Microbacteriaceae</i>	<i>Agromyces</i>	13	CNJ745 (DQ448705)	—	—	—
	<i>Microbacterium</i>	14	CNJ930 (DQ448706)	—	—	—
	<i>Microbacterium</i>	15	CNJ743 (DQ448707)	—	—	—
	<i>Microbacterium</i>	16	CNJ797 (DQ448708)	—	—	—
<i>Micrococcaceae</i>	<i>Kocuria</i>	17	CNJ723 (DQ448709)	—	—	—
	<i>Kocuria</i>	17	CNJ928 (DQ448783)	—	—	—
	<i>Kocuria</i>	18	CNJ900 (DQ448710)	—	—	—
	<i>Kocuria</i>	19	CNJ770 (DQ448711)	—	—	+
	<i>Kocuria</i>	19	CNJ787 (DQ448773)	—	—	+
	<i>Micrococcus</i>	20	CNJ719 (DQ448712)	—	—	—
<i>Micromonosporaceae</i>	<i>Micromonospora</i>	21	CNS326 (DQ448713)	—	—	+
	<i>Micromonospora</i>	22	CNJ878 (DQ448714)	+	+	+
	<i>Salinispora</i>	23	CNS051 (DQ448715)	+	+	+
	<i>Salinispora</i>	23	CNS205 (CP000850)	+	+	+
	<i>Salinispora</i>	24	CNS055 (DQ224159)	+	—	+
	<i>Salinispora</i>	24	CNS143 (DQ092624)	+	+	+
<i>Mycobacteriaceae</i>	<i>Mycobacterium</i>	25	CNJ859 (DQ448716)	+	—	+
	<i>Mycobacterium</i>	26	CNJ823 (DQ448717)	+	—	+
<i>Nocardiaceae</i>	<i>Nocardia</i>	27	CNS044 (DQ448718)	+	—	+
<i>Nocardiodiaceae</i>	<i>Aeromicrobium</i>	28	CNJ889 (DQ448719)	—	—	—
	<i>Marmoricola</i>	29	CNJ780 (DQ448720)	—	—	—
	<i>Marmoricola</i>	30	CNJ872 (DQ448721)	—	—	+
	<i>Nocardioiodes</i>	31	CNJ892 (DQ448722)	—	—	—
<i>Nocardiopsaceae</i>	<i>Nocardiopsis</i>	32	CNR923 (DQ448723)	+	—	+
<i>Promicromonosporaceae</i>	<i>Promicromonospora</i>	33	CNJ734 (DQ448724)	—	—	—
<i>Pseudonocardiaceae</i>	<i>Pseudonocardia</i>	34	CNJ888 (DQ448725)	—	—	+
	<i>Pseudonocardia</i>	35	CNS139 (DQ448726)	—	—	+
	<i>Pseudonocardia</i>	36	CNS004 (DQ448727)	—	—	+
<i>Streptomycetaceae</i>	<i>Streptomyces</i>	37	CNJ962 (DQ448737)	+	—	+
	<i>Streptomyces</i>	38	CNR872 (DQ448734)	+	—	+
	<i>Streptomyces</i>	39	CNR877 (DQ448740)	—	—	+
	<i>Streptomyces</i>	40	CNR880 (DQ448735)	—	—	—
	<i>Streptomyces</i>	41	CNR881 (DQ448730)	+	—	+
	<i>Streptomyces</i>	42	CNR884 (DQ448728)	—	—	+
	<i>Streptomyces</i>	43	CNR879 (DQ448785)	+	—	+
	<i>Streptomyces</i>	43	CNR885 (DQ448739)	+	—	+
	<i>Streptomyces</i>	44	CNR887 (DQ448738)	+	—	+
	<i>Streptomyces</i>	45	CNR918 (DQ448731)	+	—	+
	<i>Streptomyces</i>	46	CNR924 (DQ448732)	+	—	+
	<i>Streptomyces</i>	47	CNR925 (DQ448742)	+	—	+
	<i>Streptomyces</i>	48	CNR926 (DQ448729)	—	—	+
	<i>Streptomyces</i>	49	CNR927 (DQ448786)	+	—	+
	<i>Streptomyces</i>	49	CNR940 (DQ448741)	+	—	+
	<i>Streptomyces</i>	50	CNS177 (DQ448736)	+	—	+
	<i>Streptomyces</i>	51	CNR876 (DQ448784)	+	—	+
<i>Thermomonosporaceae</i>	<i>Actinomadura</i>	52	CNU125 (DQ448743)	+	—	—
Total				26	5	38

^a Primers targeted modular, iterative, and hybrid but not enediyne PKSs.

TABLE 2. KS amino acid sequences^a

Strain	Nearest type strain	No. of clones sequenced	No. of unique clones	NCBI accession no.	Top BLAST match (source organism), clone identifier ^a	BLAST match pathway product	% identity ^b	Cluster	No. of clones of PKS type ^c :	
									Mod.	Hyb.
CNU-125	<i>Actinomyces crema</i>	2	1	FJ844611	<i>chlB1</i> (<i>Streptomyces antibioticus</i>)	Chlorothricin	58	1	0	1
CNJ-863	<i>Gordonia nitida</i>	26	1	FJ844612	<i>mcyG</i> KS1 (<i>Microcystis aeruginosa</i>), 1	Microcystin	45	1	0	1
CNJ-878	<i>Micromonospora endolithica</i>	2	2	FJ844613	<i>miaB</i> (<i>Stigmatella aurantiaca</i>)	Myxothiazol	60	1	1	0
		1	1	FJ844614	<i>nysI</i> (<i>Streptomyces noursei</i>)	Nystatin	63	2	1	0
CNJ-859	<i>Mycobacterium brisbanense</i>	1	1	FJ844615	<i>jamL</i> (<i>Lyngbya majuscula</i>)	Jamaicamide	49	1	1	0
CNJ-823	<i>Mycobacterium poriferarum</i>	2	1	FJ844616	<i>chlA2</i> (<i>S. antibioticus</i>)	Chlorothricin	75	1	1	0
CNS-044	<i>Nocardia arthritidis</i>	13	4	FJ844617	<i>amphC</i> (<i>Streptomyces nodosus</i>)	Amphotericin	71	1	1	0
				FJ844618	<i>fecC</i> (<i>Streptomyces</i> sp. strain FR-008)	Candidin	66	2	1	0
				FJ844619	<i>jamJ</i> (<i>Lyngbya majuscula</i>)	Jamaicamide	51	3	1	0
CNR-923	<i>Nocardiosis lucentensis</i>	3	1	FJ844620	<i>mcyD</i> (<i>Microcystis aeruginosa</i>)	Microcystin	52	4	1	0
CNS-051	<i>Salinispora arenicola</i>	2	1	FJ844621	<i>chlB1</i> KS1 (<i>S. antibioticus</i>), 1	Chlorothricin	53	1	0	1
CNS-205	<i>Salinispora arenicola</i>	98	7	FJ844622	<i>chlB1</i> (<i>S. antibioticus</i>)	Chlorothricin	61	1	0	1
				FJ844623	<i>rifB</i> KS2 (<i>Amicrolatopsis mediterranei</i>), 1247	Rifamycin	87	1	1	0
				FJ844624	<i>rifB</i> KS3 (<i>Amicrolatopsis mediterranei</i>), 1247	Rifamycin	89	1	1	0
				FJ844625	<i>rifE</i> KS1 (<i>Amicrolatopsis mediterranei</i>), 1250	Rifamycin	86	1	1	0
				FJ844626	<i>merC</i> KS1 (<i>Streptomyces violaceusniger</i>), 3152	Meridamycin	68	2	1	0
				FJ844627	<i>chlB1</i> KS1 (<i>S. antibioticus</i>), 2029	Chlorothricin	75	3	0	1
				FJ844628	<i>chlB1</i> KS1 (<i>S. antibioticus</i>), 2407	Chlorothricin	55	4	0	1
				FJ844629	<i>chlB1</i> KS1 (<i>S. antibioticus</i>), 4951	Chlorothricin	60	5	0	1
CNS-143	<i>Salinispora pacifica</i>	45	5	FJ844630	<i>vica</i> (<i>Streptomyces halstedii</i>)	Vicenistatin	75	1	1	0
				FJ844631	<i>vica</i> (<i>S. halstedii</i>)	Vicenistatin	77	1	1	0
				FJ844632	<i>vicB</i> (<i>S. halstedii</i>)	Vicenistatin	72	1	1	0
				FJ844633	<i>vicC</i> (<i>S. halstedii</i>)	Vicenistatin	71	1	1	0
				FJ844634	<i>vicD</i> (<i>S. halstedii</i>)	Vicenistatin	72	1	1	0
CNS-055	<i>Salinispora pacifica</i> A	40	8	FJ844635	<i>vica</i> (<i>S. halstedii</i>)	Vicenistatin	73	1	1	0
				FJ844636	<i>vica</i> (<i>S. halstedii</i>)	Vicenistatin	72	1	1	0
				FJ844637	<i>vicB</i> (<i>S. halstedii</i>)	Vicenistatin	72	1	1	0
				FJ844638	<i>vicD</i> (<i>S. halstedii</i>)	Vicenistatin	74	1	1	0
				FJ844639	<i>vicD</i> (<i>S. halstedii</i>)	Vicenistatin	70	1	1	0
				FJ844640	<i>merC</i> (<i>S. violaceusniger</i>)	Meridamycin	68	2	1	0
				FJ844641	<i>mcyE</i> (<i>Microcystis aeruginosa</i>)	Meridamycin	50	3	1	0
				FJ844642	<i>curG</i> (<i>Lyngbya majuscula</i>)	Curacin	50	4	0	1
CNS-237	<i>Salinispora pacifica</i> B	40	11	FJ844643	<i>curI</i> (<i>Lyngbya majuscula</i>)	Curacin	57	1	1	0
				FJ844644	<i>ecoC</i> (<i>Streptomyces aizunensis</i>)	ECO-02301	71	2	1	0
				FJ844645	<i>ecoC</i> (<i>S. aizunensis</i>)	ECO-02301	69	2	1	0
				FJ844646	<i>ecoC</i> (<i>S. aizunensis</i>)	ECO-02301	71	2	1	0
				FJ844647	<i>ecoC</i> (<i>S. aizunensis</i>)	ECO-02301	71	2	1	0
				FJ844648	<i>ecoC</i> (<i>S. aizunensis</i>)	ECO-02301	68	2	1	0
				FJ844649	<i>ecoC</i> (<i>S. aizunensis</i>)	ECO-02301	69	2	1	0
				FJ844650	<i>ecoG</i> (<i>S. aizunensis</i>)	ECO-02301	68	2	1	0
				FJ844651	<i>spnD</i> (<i>Saccharopolyspora spinosa</i>)	Spinosa	75	3	1	0
				FJ844652	<i>spnD</i> (<i>Saccharopolyspora spinosa</i>)	Spinosa	73	3	1	0
CNJ-927	<i>Serinicoccus marinus</i>	1	1	FJ844653	<i>mcyE</i> (<i>Microcystis aeruginosa</i>)	Microcystin	50	4	1	0
CNR-876	<i>Streptomyces aureofaciens</i>	1	1	FJ844654	<i>chlB1</i> KS1 (<i>S. antibioticus</i>), 1	Chlorothricin	53	1	0	1
CNR-881	<i>Streptomyces bikiniensis</i>	2	1	FJ844655	<i>jamK</i> (<i>Lyngbya majuscula</i>)	Jamaicamide	56	1	1	0
				FJ844656	<i>jamP</i> (<i>Lyngbya majuscula</i>)	Jamaicamide	54	1	0	1
CNR-918	<i>Streptomyces caviscabies</i>	3	2	FJ844657	<i>jamK</i> (<i>Lyngbya majuscula</i>)	Jamaicamide	54	1	1	0
				FJ844658	<i>chlB1</i> (<i>S. antibioticus</i>)	Chlorothricin	52	2	0	1

CNR-924	<i>Streptomyces chartreusis</i>	3	2	FJ844659	<i>ecoA</i> (<i>S. aizunensis</i>)	ECO-02301	75	1	1	0	0
				FJ844660	<i>epoD</i> (<i>Polyangium cellulosum</i>)	Epothilone	68	2	1	0	0
CNR-872	<i>Streptomyces hebeiensis</i>	2	1	FJ844661	<i>jamL</i> (<i>Lyngbya majuscula</i>)	Jamaicamide	59	1	1	0	0
CNS-177	<i>Streptomyces lydicus</i>	3	2	FJ844662	<i>fsuC</i> (<i>Streptomyces</i> sp. FR-008)	Candicidin	74	1	1	0	0
				FJ844663	<i>merC</i> (<i>S. violaceusniger</i>)	Meridamycin	72	2	1	0	0
CNJ-962	<i>Streptomyces sampsonii</i>	1	1	FJ844664	<i>lipC</i> (<i>S. aureofaciens</i>)	Lipomycin	73	1	1	0	0
CNR-887	<i>S. sampsonii</i>	1	1	FJ844665	<i>conB</i> (<i>Streptomyces neyagawaensis</i>)	Concanamycin	71	1	1	0	0
CNR-879	<i>Streptomyces tendae</i>	2	2	FJ844666	<i>merB</i> (<i>S. violaceusniger</i>)	Meridamycin	70	1	1	0	0
				FJ844667	<i>vicA</i> (<i>S. halstedii</i>)	Vicenistatin	77	2	1	0	0
CNR-885	<i>S. tendae</i>	41	14	FJ844668	<i>kjID</i> KS2 (<i>Actinomadura kijaniata</i>), 1	Kijanicin	89	1	1	0	0
				FJ844669	<i>kjIA</i> KS4 (<i>A. kijaniata</i>), 2	Kijanicin	81	2	1	0	0
				FJ844670	<i>kjIA</i> KS4 (<i>A. kijaniata</i>), 3	Kijanicin	84	2	1	0	0
				FJ844671	<i>kjIC</i> KS1 (<i>A. kijaniata</i>), 4	Kijanicin	85	2	1	0	0
				FJ844672	<i>kjIC</i> KS1 (<i>A. kijaniata</i>), 5	Kijanicin	84	2	1	0	0
				FJ844673	<i>kjIA</i> KS4 (<i>Actinomadura kijaniata</i>), 6	Kijanicin	80	2	1	0	0
				FJ844674	<i>kjIB</i> KS2 (<i>A. kijaniata</i>), 7	Kijanicin	91	2	1	0	0
				FJ844675	<i>merA</i> KS3 (<i>S. violaceusniger</i>), 8	Meridamycin	76	3	1	0	0
				FJ844676	<i>merA</i> KS3 (<i>S. violaceusniger</i>), 9	Meridamycin	74	3	1	0	0
				FJ844677	<i>spnA</i> KS1 (<i>Saccharopolyspora spinosa</i>), 12	Spinosa	76	4	1	0	0
				FJ844678	<i>spnA</i> KS1 (<i>Saccharopolyspora spinosa</i>), 13	Spinosa	72	4	1	0	0
				FJ844679	<i>tmcC</i> KS1 (<i>Streptomyces spirovercillatus</i>), 10	Tautomycin	93	5	1	0	0
				FJ844680	<i>tmcC</i> KS2 (<i>S. spirovercillatus</i>), 11	Tautomycin	95	5	1	0	0
				FJ844681	<i>kjIA</i> KS2 (<i>A. kijaniata</i>), 14	Kijanicin	72	6	1	0	0
CNR-925	<i>Streptomyces thermocaprophilus</i>	14	6	FJ844682	<i>furA</i> KS3 (<i>S. aculeolatus</i>), 1	Actinofuranone	96	1	1	0	0
				FJ844683	<i>furA</i> KS2 (<i>S. aculeolatus</i>), 2	Actinofuranone	92	2	1	0	0
				FJ844684	<i>furD</i> KS1 (<i>S. aculeolatus</i>), 3	Actinofuranone	97	2	1	0	0
				FJ844685	<i>tmnA</i> KS3 (<i>Streptomyces</i> sp. strain NRRL 11266), 4	Tetronomycin	97	3	1	0	0
				FJ844686	<i>tmnA</i> KS1 (<i>Streptomyces</i> sp. NRRL 11266), 5	Tetronomycin	97	4	1	0	0
				FJ844687	<i>tmnA</i> KS2 (<i>Streptomyces</i> sp. NRRL 11266), 6	Tetronomycin	98	4	1	0	0
CNR-927	<i>S. thermocaprophilus</i>	8	7	FJ844688	<i>curA</i> (<i>Lyngbya majuscula</i>)	Curacin	52	1	1	0	0
				FJ844689	<i>curA</i> (<i>Lyngbya majuscula</i>)	Curacin	53	1	1	0	0
				FJ844690	<i>ecoI</i> (<i>S. aizunensis</i>)	ECO-02301	75	2	1	0	0
				FJ844691	<i>ecoH</i> (<i>S. aizunensis</i>)	ECO-02301	76	2	1	0	0
				FJ844692	<i>ecoH</i> (<i>S. aizunensis</i>)	ECO-02301	83	2	1	0	0
				FJ844693	<i>jamL</i> (<i>Lyngbya majuscula</i>)	Jamaicamide	61	3	1	0	0
				FJ844694	<i>vicD</i> (<i>S. halstedii</i>)	Vicenistatin	76	4	1	0	0
CNR-940	<i>S. thermocaprophilus</i>	35	6	FJ844695	<i>ecoA</i> (<i>S. aizunensis</i>)	ECO-02301	75	1	1	0	0
				FJ844696	<i>ecoA</i> (<i>S. aizunensis</i>)	ECO-02301	74	1	1	0	0
				FJ844697	<i>ecoA</i> (<i>S. aizunensis</i>)	ECO-02301	78	1	1	0	0
				FJ844698	<i>spnE</i> (<i>Saccharopolyspora spinosa</i>)	Spinosa	64	2	1	0	0
				FJ844699	<i>spnE</i> (<i>Saccharopolyspora spinosa</i>)	Spinosa	73	2	1	0	0
				FJ844700	<i>epoD</i> (<i>Polyangium cellulosum</i>)	Epothilone	59	3	1	0	0
Total								79	8	3	3

^a Top BLAST matches are to KS domains associated with experimentally characterized biosynthetic pathways. Clone identifiers correspond to those presented in Fig. 1 and 2.

^b KS amino acid sequences with $\geq 85\%$ amino acid identity to the top BLAST matches are in bold.

^c Mod., modular; iter., iterative; hyb., hybrid.

mixture was added to the competent cells, the cells were incubated on ice for 30 min, heat shocked at 42°C for 30 s, and then allowed to recover during 1 h of shaking at 37°C in 250 μ l of SOC medium (Invitrogen). Following recovery, 25 μ l of the transformed cells were plated onto LB agar containing X-Gal (5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside) and kanamycin (50 μ g/ml). Following overnight incubation at 37°C, vectors were isolated using Qiagen's QIAprep Spin Miniprep kit according to the manufacturer's suggested protocol (Qiagen Inc., Valencia, CA). Following isolation, the vector inserts were sized using the restriction enzyme EcoRI (New England Biolabs, Inc., Beverly, MA) and sequenced using the M13F primer on an ABI 3100 DNA sequencer at the DNA Sequencing Shared Resource, UCSD Cancer Center (funded in part by NCI Cancer Center support grant 2 P30CA23100-18).

Sequence analyses. From 1 to 98 clones were sequenced from each PCR that yielded a correctly sized product. All nucleotide sequences were analyzed and manually edited using the Sequencer software package (version 4.5; Gene Codes Co., Ann Arbor, MI), and sequences were verified using the NCBI (<http://www.ncbi.nlm.nih.gov/>) Basic Local Alignment Search Tool (BLAST). Type I KS domain sequences were translated using the Sequence Manipulation Suite (<http://www.bioinformatics.org/sms2/>) and compared to sequences within the NCBI protein database using BLASTp. The top BLASTp matches reported are for KS sequences associated with experimentally characterized biosynthetic pathways.

A reference data set of KS sequences was assembled from PKS genes associated with 36 experimentally characterized pathways (22 were described in previous studies [18, 19], and 14 are described in Table S1 in the supplemental material). The 368 KS sequences in this reference data set were aligned using ClustalX version 1.83 (34) and manually edited using MacClade version 4.07 (28). To estimate the number of within-strain KS sequence clusters, all unique KS amino acid sequences obtained in this study were aligned with the reference data set and used to create phylogenetic trees (data not shown) constructed using the neighbor-joining, unweighted-pair group method using average linkages (UPGMA), and maximum-parsimony methods in PAUP (33). Clustering patterns within these trees were examined, and in cases where the KS domains from an individual strain were in the same clade and had top BLAST hits to the same pathway, those sequences were assigned to the same cluster. Stand-alone KS sequences were also assigned a distinct cluster number. KS sequences were identified as being associated with either modular, iterative, or hybrid PKSs based on the characterized function of the top BLAST match in the reference data set. The hybrid designation was applied only in cases where the KS domain was directly downstream of an NRPS module in the gene cluster from which the reference data set was derived.

Fermentation and chemical analysis. Strain CNR-925 was inoculated from a frozen stock into 25 ml of medium A1 (10 g soluble starch, 4 g yeast extract, 2 g peptone, 750 ml natural seawater, and 250 ml deionized [DI] water). The culture was incubated at 27°C with shaking at 230 rpm for 3 days and then transferred to a 2.8-liter Fernbach flask containing 1 liter of the same medium. Following another three days of incubation, 25 ml of culture was added to 1 liter of medium E1 (10 g glucose, 10 g soluble starch, 50 ml corn steep liquor, 3 g NaCl, 1 g $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 5 g CaCO_3 , 1 liter of DI water, pH 7.0), which is known to support tetracycline (Tm) production (27). On days 3, 5, and 7, 25 ml was removed and extracted with 50 ml of ethyl acetate. The organic layers were separated, dried over anhydrous sodium sulfate, decanted, and concentrated under vacuum. The resulting crude extracts were dissolved in methanol and analyzed at 254 and 350 nm in the positive mode using an Agilent 1100 liquid chromatography-mass spectrometer (LC-MS) and a gradient from 10% acetonitrile in water to 100% acetonitrile. Fractions containing masses characteristic of tetracycline were examined by high-resolution (HR) electrospray ionization (ESI)-mass spectrometry (MS) in positive mode using a Thermo Scientific LTQ Orbitrap XL mass spectrometer. To test for actinofuranone production, strain CNR-925 was additionally inoculated into 5 liters of medium A1BFe+C [10 g starch, 4 g yeast extract, 2 g peptone, 1 g CaCO_3 , 5 g $\text{Fe}_2(\text{SO}_4) \cdot 4\text{H}_2\text{O}$, 5 g potassium bromide (KBr), 1 liter natural seawater, pH 8.0] by following a similar scale-up procedure. Sterilized Amberlite XAD-7 resin (20 g/liter) was added on day 7 and the resulting mixture shaken for an additional 4 h. The resin was then collected by filtration through cheesecloth, washed with DI water, and extracted overnight with acetone. The acetone was then separated from the resin, concentrated *in vacuo*, and partitioned between water and ethyl acetate, and the organic fraction was dried *in vacuo*. The crude extract was then separated using a silica gel flash column, eluting with a step gradient of isooctane to ethyl acetate. All fractions were analyzed by LC-MS, and the fraction containing a compound with the molecular weight and UV properties consistent with actinofuranone B was identified. This fraction was then subjected to further chromatographic purification using semipreparative reversed-phase (C_{18}) high-performance liquid chro-

matography (HPLC), resulting in purified actinofuranone B, which was analyzed by HR ESI-MS.

To test for the production of compounds in the kijanimicin class, strain CNR-885 was cultured in 20 2.8-liter Fernbach flasks, each containing 1 liter of medium A1BFe+C by following the previously described scale-up procedures. The culture was extracted with Amberlite XAD-16 resin on day 7 as previously described, and following removal of the acetone, the resulting aqueous layer was freeze-dried. The resulting material was fractionated using silica gel column chromatography with a step gradient of isooctane, ethyl acetate, and methanol. The methanol fraction was further purified using C_{18} chromatography with a step gradient of water and acetonitrile followed by reversed-phase HPLC using 55% aqueous acetonitrile. This resulted in a partially purified compound that possessed LC-MS characteristics indicative of the kijanimicin class. This compound was further analyzed by HR ESI-MS and by carbon nuclear magnetic resonance (NMR) (CDCl_3) using a 300-MHz Varian Inova NMR spectrometer.

Method validation. KS sequences from type I PKS pathways were obtained from the genome sequence of *Salinispora arenicola* strain CNS-205 (NCBI accession no. NC009953) by using KS domains from the rifamycin (CAA11035) and calicheamicin (AAM94794) biosynthetic pathways to query the genome sequence using the Joint Genome Institute (JGI) BLAST search engine (<http://genome.ornl.gov/microbial/sare/>). The resulting sequences were aligned using ClustalX version 1.83 (42) and manually edited using MacClade version 4.07 (32), and a distance neighbor-joining tree was constructed using PAUP (33). To test the effectiveness of the type I PKS primers, KS domains from strain CNS-205 were amplified, cloned, and sequenced as described above. To assess the effect of PCR bias, the cloning experiments were performed following a single PCR and after three separate PCR mixtures were pooled. In addition, nondegenerate primers specific for the KS domains of Sare1248, Sare1249, Sare3151, and the third module of Sare3156, which were not detected in the cloning experiments, were also tested (see Table S2 in the supplemental material).

RESULTS

Distribution of biosynthetic genes. Sixty diverse actinomycete isolates representing 18 family level groupings, 25 genera, and 52 OTUs cultured as part of a prior study (20) were screened by PCR for the presence of KS domains associated with type I PKS (including enediyne) biosynthetic pathways and A domains associated with NRPS pathways. In total, all but 20 of the strains yielded sequence-verified products associated with at least one of the three targeted pathway types (Table 1). Targeted loci were observed from representatives of 12 (66.7%) of the families and 13 (52.0%) of the genera screened. More specifically, type I PKS loci were detected in 26 strains representing 22 of 52 OTUs, enediyne PKS loci were detected in five strains representing three OTUs, while NRPS loci were detected in 38 strains representing 32 OTUs. Enediyne PKS loci were detected only in strains belonging to the family *Micromonosporaceae* and represented the least common of the three biosynthetic types targeted in this study.

Sequence analysis. For each correctly sized insert, the top BLAST match was to a KS or A domain, confirming that the correct loci had been amplified. Additional sequencing (8 to 98 clones) of the KS libraries was performed for 10 of the 26 PCR-positive strains in an effort to assess genome-level sequence diversity (biosynthetic richness). These strains were selected to include a mix of genera that are either recognized (e.g., *Streptomyces*, *Salinispora*) or not recognized (e.g., *Gordonia*) as a rich source of secondary metabolites. Attempts were made to sequence these libraries until saturation, although in some cases (e.g., CNR-927), this may not have been achieved. The deeply sequenced libraries revealed a range of 1 (CNJ-863) to 14 (CNR-885) unique KS alleles (Table 2). Phylogenetic analyses (data not shown) placed these sequences into as many as six distinct clusters, indicating that some strains

may possess as many as six different PKS biosynthetic pathways.

The vast majority (88%) of the KS sequences are most closely associated with modular type I PKSs, based on BLAST and phylogenetic analyses (Table 2). Of the 90 unique KS sequences that were identified, 14 shared $\geq 85\%$ amino acid identity with their top BLAST matches (Table 2). These 14 sequences were derived from strains CNS-205 (3 sequences), CNR-885 (5 sequences), and CNR-925 (6 sequences). The top BLAST matches for these 14 sequences are to genes responsible for the biosynthesis of five distinct secondary metabolites (Table 2). CNS-205 was included largely for method validation (see below), as a genome sequence is available for this strain (40) and it has previously been shown to produce PKS-derived metabolites in the rifamycin class (26). As expected, KS sequences with a high percent identity (86 to 89%) to those associated with rifamycin biosynthesis were detected from this strain (Table 2).

The six KS sequences obtained from strain CNR-925 all shared $>90\%$ amino acid identity with homologous sequences from experimentally characterized pathways. Three of these shared $>95\%$ identity to sequences associated with the tetronomycin (Tmn) biosynthetic gene cluster (13), while the remaining three had an equally high level of identity to sequences associated with the biosynthesis of the metabolite E-837, which was concurrently reported from both *Streptomyces aculeolatus* (5) and a marine-sample-derived *Streptomyces* sp. (strain CNQ-766) under the name actinofuranone (12). Interestingly, the E-837-producing *S. aculeolatus* strain (NRRL 18422) was also reported to produce tetronomycin. A phylogenetic analysis revealed the close evolutionary relationships of the CNR-925 KS sequences with the two experimentally characterized pathways (Fig. 1) and led to the hypothesis that this strain had the genetic potential to produce secondary metabolites related to both actinofuranone and tetronomycin (see "Fermentation studies," below). For both biosynthetic pathways, the KS sequences fell into two separate clusters (Fig. 1; Table 2), providing examples in which the KS domains from one pathway do not all have similar evolutionary histories. For this reason, KS clustering patterns provide an estimate of the maximum as opposed to the absolute number of distinct biosynthetic pathways that may be present in a strain.

Strain CNR-885 yielded the largest number of KS sequence clusters (six) (Fig. 1) and distinct loci (14), of which five shared $\geq 85\%$ amino acid identity with top BLAST hits from experimentally characterized biosynthetic pathways (Table 2). Of these five sequences, three had top BLAST matches to genes associated with the biosynthesis of the polyketide-derived macrolide kijanimicin (34). Two additional KS sequences formed a separate cluster and had high levels of sequence identity with KSs associated with the biosynthesis of the polyketide tautomycin (30). These close affiliations led to the hypothesis that CNR-885 has the genetic potential to produce secondary metabolites in both the kijanimicin and tautomycin classes. Despite sequencing of 41 clones from this strain, 12 of the alleles were detected only once or twice, while two were detected 12 times each (Fig. 1), indicating clear PCR bias. The detection of six KS lineages in CNR-885, with closest BLAST matches to four different PKS pathways, provides evidence that this strain

may be a rich source of diverse, polyketide-derived secondary metabolites.

The remaining KS sequences detected in this study had top BLAST matches that ranged from 45 to 84% amino acid identity to KS domains from 15 different biosynthetic pathways. The most common top BLAST matches were to KS sequences associated with chlorothricin and jamaicamide biosynthesis, which was observed in each of seven strains. Interestingly, strains CNJ-927 (*Serinicoccus marinus*) and CNR-923 (*Nocardioopsis lucentensis*) belong to different families within the *Actinomycetales* yet possess KS sequences that cluster closely together (Fig. 1), suggesting that they have undergone a recent horizontal gene transfer (HGT) event. The pathways associated with these two sequences likely produce the same metabolites, although they are not likely to be closely related to chlorothricin given the low level of sequence identity (53%) to this pathway (Table 2). It is also of interest that there was little overlap in the top BLAST hits for *Salinispora pacifica* strains CNS-143 and CNS-055 with those obtained for the "B" phylogroup of this species (CNS-237), suggesting that different phylogenotypes within the same species may produce distinct sets of polyketide-derived secondary metabolites.

In addition to being detected among members of the *Streptomycetaceae*, type I PKS sequences were detected among actinomycete taxa that are not traditionally recognized for the production of secondary metabolites. These included strain CNJ-863, which is most closely related to the unicellular actinomycete *Gordonia nitida*, a genus for which no prior reports of secondary-metabolite production could be found. The KS sequences associated with this strain were largely identical (sharing $>99\%$ amino acid identity), thus indicating that it may possess only one type I PKS pathway consisting of a single module. BLAST (Table 2) and phylogenetic (Fig. 1) analyses revealed that these KSs were most closely related to the hybrid PKS-NRPS pathway responsible for microcystin biosynthesis (43). These sequences, however, shared only 45% amino acid identity, indicating that a closely related biosynthetic pathway has yet to be characterized. Chemical studies of this strain have yet to yield any hybrid PKS-NRPS secondary metabolites.

Fermentation studies. Fermentation studies were performed to test the hypothesis that the production of specific secondary metabolites could be predicted in cases where KS sequences shared $\geq 85\%$ amino acid identity with experimentally characterized pathways. In the case of CNR-925, for which the production of tetronomycin and actinofuranone were predicted, the strain was first cultured in the medium from which tetronomycin was originally isolated (27). On day 5, a molecular ion that matched the mass of tetronomycin was detected in the ethyl acetate extract of the culture. The UV absorbance maxima associated with this compound also matched those of tetronomycin (wavelengths of maximum absorption [λ_{max}], 252 and 301) (27) and subsequent high-resolution mass spectral data confirmed the presence of the tetronomycin sodium salt in the extract (Fig. 2). To test for actinofuranone production (12), LC-MS analyses of CNR-925 extracts revealed a compound having a molecular ion and UV profile diagnostic for actinofuranone (λ_{max} , 241 and 282) (12). Subsequent high-resolution mass spectral analysis of the purified compound confirmed the presence of actinofuranone B in the extract (Fig. 2).

FIG. 2. Tetronomycin and actinofuranone structure confirmation. (A and B) High-resolution mass (A) and structure (B) of tetronomycin (sodium salt); (C and D) high-resolution mass (C) and structure (D) of actinofuranone B. MW, molecular weight; MF, molecular formula; mmu, milli-mass units.

In the case of CNR-885, the production of compounds in the kijanimicin and tautomycin classes was predicted based on the KS analyses. Following cultivation in medium A1BFe+C, the acetone extract of organic materials sequestered by XAD-16 resin was subjected to column chromatography and then

HPLC, yielding a pure compound. The high-resolution mass of this compound was 1,209.5958 (m/z [M+Na]⁺), and its molecular formula was C₆₁H₉₀N₂O₂₁ (Fig. 3), which corresponds precisely to that reported for kijanimicin 53 (34). A further analysis of the carbon NMR spectrum for this compound in

FIG. 1. Actinobacterial KS phylogeny. Neighbor-joining distance tree constructed using aligned KS domain sequences (223 amino acid positions) from type I PKS pathways cloned from strains CNJ-863, CNJ-927, CNR-885, CNR-923, and CNR-925 (in bold). Strain numbers are followed by the clone's identification number (Table 2), the number of identical clones sequenced (in brackets), and the accession number. In the case of CNR-885, KS clusters are designated with a "C" number, the predicted product of the pathway, and the percentage of amino acid identity to KS sequences from that pathway. The four sequences comprising clusters C3 and C4 were separated into two clusters because they have top BLAST matches to different biosynthetic pathways (Table 2). For reference, the KS domain sequence corresponding to the top BLAST match for each clone is included in the tree, along with any additional KS sequences in that gene. Also included for reference are all of the KS sequences from the actinofuranone (Fur), kijanimicin (Kij), meridamycin (Mer), microcystin (Mcy), nostopeptolide (Nos), rapamycin (Rap), spinosad (Spn), tautomycin (Ttm), and tetronomycin (Tmn) biosynthetic pathways. Reference sequences are identified by the name of the associated product (in bold if this product is predicted based on a high level of sequence identity), protein name, KS domain number in that gene, and NCBI accession number (in parentheses). A KS domain from the type I fatty acid synthase involved in mycocerosic acid biosynthesis was used to position the root. Bootstrap values (in percentages) calculated from 1,000 resamplings using the parsimony fully heuristic search are shown at their respective nodes for values of ≥60%. Although some within-pathway branching patterns changed when neighbor-joining, UPGMA, and maximum-parsimony treeing methods were applied, the overall tree topology was maintained.

FIG. 3. Kijanimicin 53 structure confirmation. High-resolution mass (A) and structure (B) of kijanimicin 53.

comparison to literature values (34) confirms the identity of this compound (see Table S3 in the supplemental material). Strain CNR-885 was additionally cultured in the medium from which tautomycin was originally reported (11); however, there was no evidence for the production of this or related compounds over the course of the fermentation. In the case of strain CNS-205, where KS sequences sharing $\geq 85\%$ amino acid identity with those associated with the biosynthesis of rifamycin were observed, the production of this compound has already been reported (26), so no further studies were performed. Thus, in four of five cases with three different strains, when KS sequences shared $\geq 85\%$ amino acid identity with homologous loci from experimentally characterized pathways, compounds identical or closely related to those produced by the characterized pathways were detected. No efforts were made to test for the products of PKS genes that shared $< 85\%$ sequence identity to the reference set of experimentally characterized pathways.

Method validation. A total of 27 type I KS domains were obtained following a bioinformatic analysis of the genome sequence of *Salinispora arenicola* strain CNS-205. A subsequent analysis placed these KS sequences into nine biosynthetic pathways, of which two are modular, three are iterative, two are hybrid (PKS-NRPS), and two are associated with enediyne biosynthesis (40). The type I KS primers employed for method validation were designed to detect all but those sequences associated with enediyne biosynthesis (25 in total). To test their effectiveness, a single PCR and cloning experiment was performed using genomic DNA from strain CNS-205. Despite sequencing 47 clones, only seven of the expected KS domains were amplified (see Fig. S1 and Table S2 in the supplemental material). However, these sequences represented five of the seven biosynthetic pathways targeted by the primers, including both modular and all three iterative pathway types. With the exceptions of Sare3282 and Sare3156 (KS1), the phylogenetic analysis reveals a clear association of the KS sequences with reference sequences of the predicted PKS type (see Fig. S1 in the supplemental material). There was a clear bias for certain KS domains that was not easily explained by primer comple-

mentarity. For example, KS1 of Sare4951 was amplified despite a total of five primer/template mismatches, while KS1 of Sare3154 was not amplified even though there were only two mismatches (see Table S2 in the supplemental material). However, specific primers designed for the KS domains present in Sare1248, Sare1249, Sare3151, and KS3 of Sare3156 (Table S2), which were not observed among the original 47 clones sequenced, readily amplified these sequences, indicating that PCR bias was not due to template access. Cloning the combined products of three individual PCRs did not increase the number of KS domains detected.

DISCUSSION

Microbial natural products represent a remarkable source of small-molecule diversity from which new drugs and other useful products have been developed. At present, contemporary approaches to natural-product discovery include methods such as genome scanning (49, 50) and phylogenetic prediction (28), which incorporate DNA sequencing and bioinformatic analyses into the discovery process. The research described here extends these concepts by providing an assessment of the biosynthetic capabilities of a taxonomically diverse collection of marine-sediment-derived actinomycetes (20). The goals were to screen this collection for genes associated with the production of specific types of secondary metabolites and to test the hypothesis that a relatively small amount of sequence data from the appropriate genetic loci can be used to predict secondary-metabolite production in cases where the sequences have a high level of identity to experimentally characterized biosynthetic pathways.

PCR screening of 60 marine-sediment-derived actinomycetes revealed that genes associated with secondary-metabolite biosynthesis were common yet unevenly distributed among different taxa. The detection of PKS and NRPS genes among strains related to genera such as *Gordonia* and *Kocuria* suggests that these poorly studied genera represent an underexplored resource for natural-product discovery. Our initial chemical analyses of these strains, however, have

failed to yield secondary metabolites in the predicted structure classes, which may be due to the products of these pathways having low molecular weights and thus going undetected in the LC-MS screening (data not shown). Conversely, the nearly ubiquitous detection of PKS and NRPS genes in *Streptomyces* and *Micromonospora* spp. creates challenges in selecting strains that produce new metabolites as opposed to the many well-known compounds that have already been discovered from these genera. To help overcome these challenges, PCR-amplified KS domains from PKS genes were cloned, sequenced, and analyzed. KS domains were selected, as they are highly conserved and tend to cluster phylogenetically based on the secondary metabolites they produce (19). Although the sequencing of complete genomes (44) and biosynthetic pathways (35) has been used to make effective predictions about the structures of secondary metabolites, these methods require extensive sequencing efforts and a high level of expertise in the molecular genetics of natural-product biosynthesis. The results presented here revealed that in four of five cases where a high level of KS amino acid sequence identity (i.e., $\geq 85\%$) to previously characterized biosynthetic pathways was observed, the production of compounds related to those pathways could be accurately predicted and subsequently confirmed by chemical analyses. These predictions could be made in the absence of completely sequenced biosynthetic pathways and instead based simply on the analysis of 669-bp regions of KS domains from single PKS genes.

Interestingly, two of these cases were observed in strain CNR-925, where KS domains sharing 97 and 98% amino acid identity to those associated with actinofuranone and tetrone-mycin biosynthesis, respectively, were detected. Once the appropriate growth conditions were applied and the molecular weights and UV absorption properties of the target molecules incorporated into the search parameters, these metabolites were readily identified in the fermentation extracts. Strain CNR-885 provides another interesting example, as 14 unique KS sequences were detected among six sequence clusters, the maximum number observed in any of the strains examined (Table 2). BLAST analyses indicated that these sequences were most closely related to genes involved in the biosynthesis of four different secondary metabolites. In two cases (kijanimitin and tautomycin), the level of sequence identity was high ($\geq 85\%$), and thus this strain was tested for the ability to produce compounds in these classes. In the case of kijanimitin, the compound kijanimitin 53 (27) was isolated and its identity confirmed by HR MS and NMR studies. In the case of tautomycin, compounds in this class were not detected even when the strain was cultured in a medium known to support production. These results suggest that the tautomycin-like pathway may be nonfunctional or that the growth conditions required for production in this strain were not met. We did not test for the production of compounds related to meridamycin and spinosad due to the relatively low levels of sequence identity (72 to 76%) to KS domains associated with their biosynthesis. Although the number of sequence clusters identified is not directly correlated to the number of distinct types of secondary metabolites that will be produced, it does provide a mechanism by which strains can be identified as attractive targets for detailed fermentation and chemical studies, es-

pecially in cases where the sequence identities to characterized pathways are low.

The observation that only 14 of 90 KS sequences displayed $\geq 85\%$ amino acid identity to experimentally characterized pathways is likely as much a reflection on the relatively small number of pathways that have been characterized to date as it is on the possibility that some of the pathways detected may be associated with the production of new compounds. Nonetheless, it is clear that the analysis of the appropriate genetic loci can provide an effective method for natural-product dereplication and that the effectiveness of this approach will increase in parallel with the number of biosynthetic pathways that are experimentally characterized. The potentially large number of fermentation conditions required to induce the expression of the complete secondary-metabolite repertoire of an individual strain underscores the utility of incorporating DNA sequence analyses into the early stages of the discovery process so that strains with the greatest genetic potential to produce new secondary metabolites can be identified and targeted for study. Although there is no way to predict that new metabolites will have interesting pharmacological properties, and thus this is not a direct approach to drug discovery, it is nonetheless a rational method to avoid the isolation of previously described metabolites in favor of new compounds that can subsequently be tested for biological activity.

The levels of KS sequence identity that are predictive of secondary-metabolite production remain to be determined; however, it is unlikely that a consistent level will apply to all strains, as it depends on the rate of sequence evolution and the amount of time that the pathways have been isolated in the respective genomes. It is also unlikely that this approach will work in all cases, as HGT events may blur relationships between individual domains and the products of entire biosynthetic pathways. However, the utility of this approach is clearly linked to the observation that complete biosynthetic pathways are exchanged among dissimilar organisms via HGT. If HGT were not occurring, the predictions made here would simply be a form of chemotaxonomy, with convergent evolution accounting for scenarios in which distantly related strains produce the same metabolite. Clear evidence of recent-pathway HGT is observed in an analysis of the rifamycin biosynthetic pathways in *Amycolatopsis mediterranei* and the marine actinomycete *S. arenicola* (26, 28). In this case, both strains produce compounds in the rifamycin class with the KS domains cloned from *S. arenicola* (CNS-205) sharing $\geq 86\%$ amino acid identity with those from *A. mediterranei* (Table 2). Thus, identities of $\geq 85\%$ that are also supported by phylogenetic methods may be a good starting point for the level of sequence similarity that can be used to make predictions about the products of secondary metabolism. This level of sequence similarity is further supported by the observation that accurate predictions could be made in four of five cases where it was achieved.

Strong conservation of the enediyne PKS across several actinomycete genera suggests that the PCR primers used during this study might effectively identify diverse actinomycetes in possession of this gene cassette. However, the percentage of marine-sediment-derived actinomycetes that tested positive for enediyne PKS genes (8.3%) was lower than that reported for soil-derived actinomycetes (49). Nonetheless, enediyne bio-

synthesis in *Salinispora* spp. has been linked to a number of unique secondary metabolites (7, 36, 38), suggesting that marine strains may represent an important source of novel enediyne and related metabolites. It is also of interest that the eight KS sequences associated with iterative type I PKS pathways (i.e., ChlB) were widely distributed among diverse taxonomic groups. Two of these KS domains display clear evidence of recent HGT (e.g., Fig. 1 and Table 2, strains CNJ-927 and CNR-923), suggesting that iterative PKS pathways may be particularly promiscuous. The fact that all of these KS sequences display relatively low levels of identity to the iterative portion of the chlorothricin pathway (ChlB) suggests they are not associated with the production of this compound and that the top matches reflect a paucity of characterized pathways in this biosynthetic class.

The type I PKS primers used during this study did not amplify 100% of the KS loci detected in the genome sequence of *Salinispora arenicola* strain CNS-205, suggesting that these primers will underestimate the KS sequence diversity within individual strains. This limitation appears to be at least in part due to insufficient primer complementarity, as specific (non-degenerate) primers readily amplified KS domains that were not observed within libraries created using the degenerate primers (see Table S2 in the supplemental material). Nonetheless, a single primer set captured representative KSs from the majority of the PKS gene clusters in the CNS-205 genome, indicating that this simple approach could be used to provide a reasonable estimate of within-strain biosynthetic diversity. It is also likely that highly divergent sequences will go undetected in the PCR screens, as the primers were designed largely based on known sequences from *Streptomyces* and *Micromonospora* species, and thus some of the most interesting pathways may be missed with this approach. Despite these limitations, a large number of strains from families other than the *Streptomycetaceae* and *Micromonosporaceae* yielded sequence-verified PCR products, indicating that this approach can be successfully applied to diverse actinobacteria and that this method can be used as a rapid screen to identify strains (or environments) that represent logical targets for natural-product discovery. In future experiments, the use of multiplex PCR may further improve the detection capabilities of this approach.

The future of natural-product drug discovery will undoubtedly continue to be influenced by advances in our understanding of the molecular genetics of secondary-metabolite biosynthesis. Sequence-based approaches will likely continue to be incorporated into discovery paradigms and provide perspective on the biosynthetic potential of an organism prior to fermentation, extraction, and chemical analysis. As these methods are developed and refined, it is conceivable that the stochastic approaches by which microorganisms have traditionally been cultured and screened for secondary-metabolite production will be abandoned in favor of sequence-based approaches that provide insight into the organism's genetic potential to produce secondary metabolites prior to fermentation and chemical studies. As access to genome sequencing becomes more readily available, there is also a growing need for facile methods to make predictions about the products of secondary metabolism that can be applied by those working outside the field of natural-product discovery and biosynthesis. In the case of PKS pathways, the phylogenetic analysis of KS domains rep-

resents a promising approach to obtain a more in-depth interpretation of secondary metabolism from genome sequence data.

ACKNOWLEDGMENTS

We acknowledge funding from the Leger Benbough Foundation, the NOAA California Sea Grant College Program, U.S. Department of Commerce grant NA04OAR4170038, and National Institutes of Health grant GM085770 (to P.R.J.) and National Cancer Institute grant CA44848 (to W.F.). S.P.G. acknowledges postdoctoral funding from the Fundação para a Ciência e Tecnologia, Portugal.

Brian Murphy is acknowledged for assisting with the analysis of actinofuranone.

The statements, findings, conclusions, and recommendations are ours and do not necessarily reflect the views of California Sea Grant or the U.S. Department of Commerce.

REFERENCES

1. Ayuso, A., D. Clark, I. Gonzalez, O. Salazar, A. Anderson, and O. Genilloud. 2005. A novel actinomycete strain de-replication approach based on the diversity of polyketide synthase and nonribosomal peptide synthetase biosynthetic pathways. *Appl. Microbiol. Biotechnol.* **67**:795–806.
2. Ayuso-Sacido, A., and O. Genilloud. 2005. New PCR primers for the screening of NRPS and PKS-I systems in actinomycetes: detection and distribution of these biosynthetic gene sequences in major taxonomic groups. *Microb. Ecol.* **49**:10–24.
3. Baltz, R. H. 2005. Antibiotic discovery from actinomycetes: will a renaissance follow the decline and fall? *SIM News* **55**:186–196.
4. Baltz, R. H. 2006. Marcel Faber Roundtable: is our antibiotic pipeline unproductive because of starvation, constipation or lack of inspiration? *J. Ind. Microbiol. Biotechnol.* **33**:507–513.
5. Banskota, A. H., J. B. McAlpine, D. Sorensen, M. Aoudate, M. Pirae, A. M. Alarco, S. Omura, K. Shiom, C. M. Farnet, and E. Zazopoulos. 2006. Isolation and identification of three new 5-alkenyl-3,3(2H)-furanones from two streptomycetes species using a genomic screening approach. *J. Antibiot. (Tokyo)* **59**:168–176.
6. Berdy, J. 2005. Bioactive microbial metabolites—a personal view. *J. Antibiot.* **58**:1–26.
7. Buchanan, G. O., P. G. Williams, R. H. Felting, C. A. Kauffman, P. R. Jensen, and W. Fenical. 2005. Sporolides A and B: structurally unprecedented halogenated macrolides from the marine actinomycete *Salinispora tropica*. *Org. Lett.* **7**:2731–2734.
8. Bull, A. T., and J. E. M. Stach. 2007. Marine actinobacteria: new opportunities for natural product search and discovery. *Trends Microbiol.* **15**:491–499.
9. Burns, B. P., A. Seifert, F. Goh, F. Pomati, A. D. Jungblut, A. Serhat, and B. A. Neilan. 2005. Genetic potential for secondary metabolite production in stromatolite communities. *FEMS Microbiol. Lett.* **243**:293–301.
10. Cane, D. E., C. T. Walsh, and C. Khosla. 1998. Harnessing the biosynthetic code: combinations, permutations, and mutations. *Science* **282**:63–68.
11. Cheng, X. C., T. Kihara, H. Kusakabe, J. Magae, Y. Kobayashi, R. P. Fang, Z. F. Ni, Y. C. Shen, K. Ko, I. Yamaguchi, and K. Isono. 1987. A new antibiotic, tautomycin. *J. Antibiot. (Tokyo)* **40**:907–909.
12. Cho, J. Y., H. C. Kwon, P. G. Williams, C. A. Kauffman, P. R. Jensen, and W. Fenical. 2006. Actinofuranones A and B, polyketides from a marine-derived bacterium related to the genus *Streptomyces* (actinomycetales). *J. Nat. Prod.* **69**:425–428.
13. Demydchuk, Y., Y. H. Sun, H. Hong, J. Staunton, J. B. Spencer, and P. F. Leadlay. 2008. Analysis of the tetracycline gene cluster: insights into the biosynthesis of a polyether tetracycline antibiotic. *ChemBiochem* **9**:1136–1145.
14. Ehrenreich, I. M., J. B. Waterbury, and E. A. Webb. 2005. Distribution and diversity of natural product genes in marine and freshwater cyanobacterial cultures and genomes. *Appl. Environ. Microbiol.* **71**:7401–7413.
15. Fenical, W., and P. R. Jensen. 2006. Developing a new resource for drug discovery: marine actinomycete bacteria. *Nat. Chem. Biol.* **2**:666–673.
16. Fieseler, L., U. Hentschel, L. Grodzanov, A. Schirmer, G. Wen, M. Platzer, S. Hrvatin, D. Butzke, K. Zimmermann, and J. Piel. 2007. Widespread occurrence and genomic context of unusually small polyketide synthase genes in microbial consortia associated with marine sponges. *Appl. Environ. Microbiol.* **73**:2144–2155.
17. Fischbach, M. A., and C. T. Walsh. 2006. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. *Chem. Rev.* **106**:3468–3496.
18. Ginolhac, A., C. Jarrin, B. Gillet, P. Robe, P. Pujic, K. Tuphile, H. Bertrand, T. M. Vogel, G. Perriere, P. Simonet, and R. Nalin. 2004. Phylogenetic analysis of polyketide synthase I domains from soil metagenomic libraries allows selection of promising clones. *Appl. Environ. Microbiol.* **70**:5522–5527.

19. Ginolhac, A., C. Jarrin, P. Robe, G. Perriere, T. M. Vogel, P. Simonet, and R. Nalin. 2005. Type I polyketide synthases may have evolved through horizontal gene transfer. *J. Mol. Evol.* **60**:716–725.
20. Gontang, E. A., W. Fenical, and P. R. Jensen. 2007. Phylogenetic diversity of Gram-positive bacteria cultured from marine sediments. *Appl. Environ. Microbiol.* **73**:3272–3282.
21. Haydock, S. F., T. Mironenko, H. I. Ghoorahoo, and P. F. Leadlay. 2004. The putative elaiophyllin biosynthetic gene cluster in *Streptomyces* sp. DSM4137 is adjacent to genes encoding adenosylcobalamin-dependent methylmalonyl CoA mutase and to genes for synthesis of cobalamin. *J. Biotechnol.* **113**:55–68.
22. Hopwood, D. A. 1997. Genetic contributions to understanding polyketide synthases. *Chem. Rev.* **97**:2465–2498.
23. Izumikawa, M., M. Murata, K. Tachibana, Y. Ebizuka, and I. Fujii. 2003. Cloning of modular type I polyketide synthase genes from salinomycin producing strain of *Streptomyces albus*. *Bioorg. Med. Chem.* **11**:3401–3405.
24. Jenke-Kodama, H., T. Borner, and E. Dittmann. 2006. Natural biocombinatorics in the polyketide synthase genes of the actinobacterium *Streptomyces avermitilis*. *PLoS Comput. Biol.* **2**:1210–1218.
25. Jenke-Kodama, H., A. Sandmann, R. Muller, and E. Dittmann. 2005. Evolutionary implications of bacterial polyketide synthases. *Mol. Biol. Evol.* **22**:2027–2039.
26. Jensen, P. R., P. G. Williams, D. C. Oh, L. Zeigler, and W. Fenical. 2007. Species-specific secondary metabolite production in marine actinomycetes of the genus *Salinispora*. *Appl. Environ. Microbiol.* **73**:1146–1152.
27. Keller-Juslen, C., H. D. King, M. Kuhn, H. R. Loosli, W. Pache, T. J. Petcher, H. P. Weber, and A. von Wartburg. 1982. Tetronomycin, a novel polyether of unusual structure. *J. Antibiot. (Tokyo)* **35**:142–150.
28. Kim, T. K., A. K. Hewavitharana, P. N. Shaw, and J. A. Fuerst. 2006. Discovery of a new source of rifamycin antibiotics in marine sponge actinobacteria by phylogenetic prediction. *Appl. Environ. Microbiol.* **72**:2118–2125.
29. Li, J. W. H., and J. C. Vederas. 2009. Drug discovery and natural products: end of an era or an endless frontier? *Science* **325**:161–165.
30. Li, W. L., Y. G. Luo, J. H. Ju, S. R. Rajski, H. Osada, and B. Shent. 2009. Characterization of the tautomycin biosynthetic gene cluster from *Streptomyces griseochromogenes* provides new insight into dialkylmaleic anhydride biosynthesis. *J. Nat. Prod.* **72**:450–459.
31. Liu, W., J. Ahlert, Q. Gao, E. Wendt-Pienkowski, B. Shen, and J. S. Thorson. 2003. Rapid PCR amplification of minimal enediynes polyketide synthase cassettes leads to a predictive familial classification model. *Proc. Natl. Acad. Sci. U. S. A.* **100**:11959–11963.
32. Maddison, D. R., and W. P. Maddison. 2001. *MacClade*, 4th ed. Sinauer Associates, Sunderland, MA.
33. Magarvey, N. A., J. M. Keller, V. Bernan, M. Dworkin, and D. H. Sherman. 2004. Isolation and characterization of novel marine-derived actinomycete taxa rich in bioactive metabolites. *Appl. Environ. Microbiol.* **70**:7520–7529.
34. Mallams, A. K., M. S. Puar, R. R. Rossman, A. T. McPhail, and R. D. Macfarlane. 1981. Kijanimicin 2. Structure and absolute stereochemistry of kijanimicin. *J. Am. Chem. Soc.* **103**:3940–3943.
35. McAlpine, J. B., B. O. Bachmann, M. Pirae, S. Tremblay, A. M. Alarco, E. Zazopoulos, and C. M. Farnet. 2005. Microbial genomics as a guide to drug discovery and structural elucidation: ECO-02301, a novel antifungal agent, as an example. *J. Nat. Prod.* **68**:493–496.
36. McGlinchey, R. P., M. Nett, and B. S. Moore. 2008. Unraveling the biosynthesis of the sporolide cyclohexenone building block. *J. Am. Chem. Soc.* **130**:2406–2407.
37. Moffitt, M. C., and B. A. Neilan. 2003. Evolutionary affiliations within the superfamily of ketosynthases reflect complex pathway associations. *J. Mol. Evol.* **56**:446–457.
38. Oh, D. C., P. G. Williams, C. A. Kauffman, P. R. Jensen, and W. Fenical. 2006. Cyanosporasides A and B, chloro- and cyano-cyclopenta[a]indene glycosides from the marine actinomycete “*Salinispora pacifica*.” *Org. Lett.* **8**:1021–1024.
39. Pelaez, F. 2006. The historical delivery of antibiotics from microbial natural products—can history repeat? *Biochem. Pharmacol.* **71**:981–990.
40. Penn, K., C. Jenkins, M. Nett, D. W. Udway, E. A. Gontang, R. P. McGlinchey, B. Foster, A. Lapidus, S. Podell, E. E. Allen, B. S. Moore, and P. R. Jensen. 2009. Genomic islands link secondary metabolism to functional adaptation in marine Actinobacteria. *ISME J.* **1751**:1–11.
41. Schirmer, A., R. Gadkari, C. D. Reeves, F. Ibrahim, E. F. DeLong, and C. R. Hutchinson. 2005. Metagenomic analysis reveals diverse polyketide synthase gene clusters in microorganisms associated with the marine sponge *Discodermia dissoluta*. *Appl. Environ. Microbiol.* **71**:4840–4849.
42. Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The CLUSTAL_X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**:4876–4882.
43. Tillett, D., E. Dittmann, M. Erhard, H. von Dohren, T. Borner, and B. A. Neilan. 2000. Structural organization of microcystin biosynthesis in *Microcystis aeruginosa* PCC7806: an integrated peptide-polyketide synthetase system. *Chem. Biol.* **7**:753–764.
44. Udway, D. W., L. Zeigler, R. N. Asolkar, V. Singan, A. Lapidus, W. Fenical, P. R. Jensen, and B. S. Moore. 2007. Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*. *Proc. Natl. Acad. Sci. U. S. A.* **104**:10376–10381.
45. Walsh, C. T. 2004. Polyketide and nonribosomal peptide antibiotics: modularity and versatility. *Science* **303**:1805–1810.
46. Watve, M. G., R. Tickoo, M. M. Jog, and B. D. Bhole. 2001. How many antibiotics are produced by the genus *Streptomyces*? *Arch. Microbiol.* **176**:386–390.
47. Wawrik, B., L. Kerkhof, G. J. Zylstra, and J. J. Kukor. 2005. Identification of unique type II polyketide synthase genes in soil. *Appl. Environ. Microbiol.* **71**:2232–2238.
48. Wawrik, B., D. Kudiev, U. A. Abdivasieva, J. J. Kukor, G. J. Zylstra, and L. Kerkhof. 2007. Biogeography of actinomycete communities and type II polyketide synthase genes in soils collected in New Jersey and Central Asia. *Appl. Environ. Microbiol.* **73**:2982–2989.
49. Zazopoulos, E., K. Huang, A. Staffa, W. Liu, B. O. Bachmann, K. Nonaka, J. Ahlert, J. S. Thorson, B. Shen, and C. M. Farnet. 2003. A genomics-guided approach for discovering and expressing cryptic metabolic pathways. *Nat. Biotechnol.* **21**:187–190.
50. Zerikly, M., and G. L. Challis. 2009. Strategies for the discovery of new natural products by genome mining. *Chembiochem* **10**:625–633.