APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Nov. 2005, p. 7401–7413 Vol. 71, No. 11
0099-2240/05/$08.00+0 doi:10.1128/AEM.71.11.7401–7413.2005
Copyright © 2005, American Society for Microbiology. All Rights Reserved.

Distribution and Diversity of Natural Product Genes in Marine and Freshwater Cyanobacterial Cultures and Genomes

Ian M. Ehrenreich, John B. Waterbury, and Eric A. Webb*

Biology Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

Received 25 March 2005/Accepted 16 July 2005

Natural products are a functionally diverse class of biochemically synthesized compounds, which include antibiotics, toxins, and siderophores. In this paper, we describe both the detection of natural product activities and the sequence identification of gene fragments from two molecular systems that have previously been implicated in natural product production, i.e., nonribosomal peptide synthetases (NRPSs) and modular polyketide synthases (PKSs), in diverse marine and freshwater cyanobacterial cultures. Using degenerate PCR and the sequencing of cloned products, we show that NRPSs and PKSs are common among the cyanobacteria tested. Our molecular data, when combined with genomic searches of finished and progressing cyanobacterial genomes, demonstrate that not all cyanobacteria contain NRPS and PKS genes and that the filamentous and heterocystous cyanobacteria are the richest sources of these genes and the most likely sources of novel natural products within the phylum. In addition to validating the use of degenerate primers for the identification of PKS and NRPS genes in cyanobacteria, this study also defines numerous gene fragments that will be useful as probes for future studies of the synthesis of natural products in cyanobacteria. Phylogenetic analyses of the cyanobacterial NRPS and PKS fragments sequenced in this study, as well as those from the cyanobacterial genome projects, demonstrate that there is remarkable diversity and likely novelty of these genes within the cyanobacteria. These results underscore the potential variety of novel products being produced by these ubiquitous organisms.

The cyanobacteria are an ecologically, morphologically, and physiologically diverse group of organisms whose primary productivity contributes to the bioenergetic foundation for higher trophic levels in both marine and freshwater environments. The ecological significance of the cyanobacteria extends beyond their productivity, though, as many of these organisms are capable of modifying their habitats through the synthesis of biologically active natural products. Characterizing these compounds, as well as the molecular mechanisms underlying their synthesis, may provide important ecological insights, in addition to opportunities for the practical application of these compounds.

In recent years, many microbiologists and biochemists have studied the distribution and functions of nonribosomal peptide synthetases (NRPSs) and modular polyketide synthases (PKSs), two similar molecular systems that are known to be involved in natural product synthesis in many bacteria, fungi, and plants (5, 16). Both NRPSs and PKSs are large (200- to 2,000-kDa), multifunctional enzymes that possess modular organization (4). NRPSs use amino acid monomers as substrates for synthesizing complex oligopeptides, whereas PKSs use acyl coenzyme A monomers to form elaborate chemical structures along a ketide backbone. NRPSs and PKSs have been found to synthesize a diverse array of biologically active compounds, including antibiotics, toxins (5), siderophores (9), and immunosuppressants (5).

Discrete modules in the NRPSs and PKSs mediate the synthesis of their products. Similar to an assembly line, each enzymatic module is responsible for the addition of a single monomer to an elongating chain. These modules, in turn, also possess modular organization, as they consist of a number of enzymatic active sites that can be partitioned into sequence domains at the level of primary structure.

Four core catalytic domains and a number of variable catalytic domains characterize individual NRPS modules, and within a single gene multiple modules may be present. An NRPS module usually contains an adenylation (A) domain, a peptidyl carrier protein domain, and a condensation domain (23). The final module involved in nonribosomal peptide synthesis ends with a thioesterase domain, which catalyzes the termination of chain elongation. Although A, peptidyl carrier protein, condensation, and thioesterase domains are thought to be present in most NRPS genes, other auxiliary domains that can have important effects on substrate processing have been discovered, including epimerization, N-methylation, cyclization, and oxidation domains (23). A well-known example of an NRPS-synthesized natural product is the immunosuppressant cyclosporin, which is synthesized by the fungi *Cylindrocarpon lucidium* Booth and *Tolypocladium inflatum* (4).

Analogous to NRPS genes, PKS genes are comprised of modules with distinct core and auxiliary catalytic domains. The essential organization of PKS modules, in order of domains, is ketosynthase (KS) domain, acyltransferase domain, and acyl carrier protein domain (23). As in NRPSs, the final PKS module in product synthesis ends with a thioesterase domain. Known auxiliary PKS domains are the ketoreductase, dehydratase, methyltransferase, and enoylreductase domains (23). Some of the most frequently dispensed antibiotics on a global scale, such as erythromycin A, are produced by PKSs (4).

Mixed NRPS-PKS pathways and genes also exist. One of the

* Corresponding author. Mailing address: Department of Biology, Woods Hole Oceanographic Institution, Woods Hole, MA 02543. Phone: (508) 289-3640. Fax: (508) 457-2134. E-mail: ewebb@whoi.edu.

TABLE 1. Strain list

Strain(s)	Reference	Characteristics[a]					Isolation
		CO	Hab	N-f	Mor	Het	
Chamaesiphon sp. strain PCC7430	22	Chro	F	N	U	NA	Stream water, Sarka Valley near Prague, Czechoslovakia, 1963
Crocosphaera watsonii WH8501 and WHO403	35	Chro	M	Y	U	NA	Western tropical Atlantic Ocean, off Brazil, 1985
Crocosphaera sp. strains WH0001 and WH0002	JW[b]	Chro	M	Y	U	NA	Mid-tropical Pacific Ocean
Crocosphaera sp. strains WH0401 and WHO402	EW[c]	Chro	M	Y	U	NA	Western tropical ocean, off Brazil, 2002
Cyanothece sp. strain WH8901	JW	Chro	M	Y	U	NA	Coastal mangroves, Florida
Cyanothece sp. strain WH8904	JW	Chro	M	Y	U	NA	Coastal mangroves, Florida
Gloeobacter sp. strain PCC7421	22	Chro	F	N	U	NA	Calcareous rock, near Kastanienbaum, Switzerland, 1972
Gloeothece sp. strain PCC6501	22	Chro	F	Y	U	NA	Freshwater, California, 1965
Anabaena sp. strain WH0404	JW	Nos	F	Y	F	Y	Small pond, School St., Woods Hole, Massachusetts
Calothrix sp. strain PCC7102	22	Nos	F	Y	F	Y	Fine desert sand, near La Portado, Antofaagasta, Chile, 1958
Nodularia sp. strain WH0405	BB[d]	Nos	M	Y	F	Y	Baltic Sea
Nostoc punctiforme ATCC 29133	1	Nos	F	Y	F	Y	Root section, *Macrozamia* sp., Australia, 1973
Nostoc sp. strain ATCC 53789	1	Nos	F	Y	F	Y	Lichen, Arron Island, Scotland
Scytonema sp. strain PCC7110	22	Nos	M	Y	F	Y	Crystal Cave (limestone), Bermuda, 1971
Lyngbya sp. strain PCC7419	22	Oscil	M	Y	F	N	Salt marsh, Woods Hole, Massachusetts, 1974
Microcoleus sp. strain PCC7420	22	Oscil	M	N	F	N	Salt marsh, Woods Hole, Massachusetts, 1974
Spirulina sp. strain PCC6313	22	Oscil	M	N	F	N	Brackish water, Berkeley, California, 1963
Trichodesmium thiebautii II-3	20	Oscil	M	Y	F	Y	Sargasso Sea, Atlantic Ocean
Pleurocapsa sp. strain PCC7319	22	Pl	M	N	U	N	Snail shell, intertidal zone, Arizona Marine Station, Mexico, 1971
Stanieria sp. strain PCC7302 (*"Dermocarpa"*)	22	Pl	M	N	U	N	Sea water tank, Arizona Marine Station, Mexico, 1971
Fischerella sp. strain PCC7414	22	Stig	F	Y	F	Y	Hot spring, New Zealand

[a] Cyanobacterial order (CO), basic habitat (Hab), ability to fix nitrogen (N-f) morphology (Mor), and ability to form heterocysts (Het) are shown. Chro, *Chroococcales* (section I); Pl, *Pleurocapsales* (section II); Oscil, *Oscillatoriales* (section III); Nos, *Nostocales* (section IV); Stig, *Stigonematales* (section V); F, freshwater; M, marine; Y, yes; N, no; U, unicellular; F, filamentous; NA, not applicable because only filamentous strains can form heterocysts.
[b] J. Waterbury, unpublished data.
[c] E. Webb, unpublished data.
[d] B. Bergman, unpublished data.

best examples of a product of a mixed NRPS-PKS pathway is the antibiotic rapamycin (23). Rapamycin synthesis proceeds first through three modular PKSs containing 14 modules in total and concludes with a single-module NRPS enzyme. Some mixed NRPS-PKS systems, such as the myxothiazol cluster in *Stigmatella aurantica* DW 4/3-1 and the antibiotic TA cluster in *Myxococcus xanthus*, have been found to actually contain both NRPS and PKS modules within the same gene (23). Numerous other mixed NRPS-PKS systems have been identified, including those that synthesize curacin A (7), barbamide (6), and jamaicamide (10). These findings suggest that multiple types of mixed NRPS-PKS systems exist.

Although the scientific community is generally aware of the prevalence of NRPSs and PKSs among the cyanobacteria (8, 17, 18), the distribution and diversity of these genes across the phylum have not been well characterized. NRPS A-domain fragments have been amplified, but not sequenced, from strains of each of the five major taxonomic sections of cyanobacteria (*Chroococcales*, *Pleurocapsales*, *Oscillitoriales*, *Nostocales*, and *Stigonematales*) (8, 25). For PKSs, even less information is currently available. To our knowledge, only one recent study has identified and sequenced PKSs from 10 cyanobacterial cultures isolated from stromatolites, representing two cyanobacterial sections (*Pleurocapsales* and *Oscillitoriales*) (3).

Using molecular and phylogenetic techniques in combination with the analysis of the sequenced cyanobacterial genomes, we show that NRPSs and PKSs are abundant across the cyanobacterial sections. In addition, we confirm previous reports of variability of NRPS and/or PKS presence within genera and provide new examples of such heterogeneity (8). Furthermore, we validate the use of degenerate probes for the detection of PKS and NRPS genes in diverse cyanobacterial strains for which genomes are not available by comparing our results for *Nostoc punctiforme* ATCC 29133 to its genome sequence. Lastly, to show that the presence of these genes correlates with natural product synthesis, we conducted biochemical assays to detect the presence of algicides and siderophores in cell extracts and supernatants. The results of these experiments suggest that both NRPS and PKS genes are widespread and genetically diverse even among very closely related cyanobacterial species. Based on our results, the undifferentiated filamentous and heterocystous strains appear to be the most likely sources of biochemically active natural products.

MATERIALS AND METHODS

Cyanobacterial culturing. The studied strains are listed and described in Table 1. All strains were grown in SN or SO (31), GN (30), or *Trichodesmium*-specific medium (32). Except for *Trichodesmium thiebautii* II-3 and *Crocosphaera* cultures, which were grown as described by Webb et al. (32), all cultures were maintained using the methods described by Waterbury et al. (31). The purity of the cyanobacterial cultures was monitored by aliquoting 1 ml of culture into 5 ml of sterile AC broth (Difco, Sparks, MD) (made with either double-distilled water [ddH_2O] or seawater, depending on the medium of the cyanobacterial culture), followed by microscopic observation for any sign of bacterial contamination. Not

all strains listed in Table 1 were tested in the bioassays of this study because of culture contamination or biomass limitations.

Algicide bioassay. Stationary-phase cultures (250 ml) were centrifuged at 30,074 × *g* for 5 min in a Beckman (Fullerton, CA) J2-21 M induction drive centrifuge. The supernatants were decanted from the samples, and the pellets were transferred into 2-ml screw cap tubes. Using a Branson (Stamford, CT) 545 Sonifier with a high-intensity microcup horn attachment, samples were sonicated for 6 min total with alternating 30-s on-off pulses. To keep the samples from overheating, ice water was pumped through the microcup horn by using a peristaltic pump. Following sonication, 1 ml of chloroform was added to each sample and the mixtures were vortexed horizontally for 2 min at maximum speed. The samples were then centrifuged for 1 min at 14,000 × *g* in an Eppendorf (Westbury, NY) MiniSpin Plus. Chloroform fractions were taken from each sample and transferred to 1.1-ml tapered glass ampoules. The samples were left unsealed in a fume hood for ~24 h until all chloroform had evaporated. The dried samples were resuspended in 250 μl of methanol.

Once cell extracts were generated, they were screened for algicides by using an adaptation of methods described by Mason et al. (14) with the test organisms *Synechococcus* sp. strain PCC7002 and *Synechocystis* sp. strain PCC6803 grown in SN supplemented with vitamin B_{12} (1.5-μg/liter final concentration) and in GN with no supplements, respectively. Algicidal test cultures were initiated with inoculums of 10 μl of stationary-phase *Synechococcus* sp. strain PCC7002 or *Synechocystis* sp. strain PCC6803 culture in 100-μl volumes, using a 96-well plate format. Various concentrations of the cyanobacterial extracts (0.5, 1, 2, and 3%) were added to the cultures prior to their inoculation with *Synechococcus* sp. strain PCC7002 or *Synechocystis* sp. strain PCC6803. Preliminary experiments showed that methanol concentrations of ≤3% permitted normal growth of the test strains. Plates were monitored by visual inspection over a 5-day period for growth inhibition relative to controls. At the end of the experiment, the absorbance at 670 nm (the maximal absorbance peak of chlorophyll *a*) of each culture was measured using a Versamax tunable microplate reader (Molecular Devices, Sunnyvale, CA). Two independent replicate extracts from each cyanobacterial strain were tested in a single plate, and the experiment was conducted in duplicate. Strains were scored as positive, negative, or variable for inhibiting growth if both replicate extracts consistently inhibited growth, no inhibition was observed from either extract, or a single extract consistently inhibited growth, respectively.

Siderophore bioassay. For many bacteria, prolonged culture growth in nutrient-depleted conditions can result in the synthesis of iron chelators known as siderophores. Hence, each tested cyanobacterium was grown under both Fe-replete (20 μM iron citrate, the standard iron concentration of GN or SN) and 1/10 Fe (2 μM) conditions. The latter condition was invariably iron limited as determined by biomass and optical density compared to a control. As described by Schwyn and Neilands (24), liquid chrome azurol S (CAS) assays were conducted to detect the presence of siderophores in supernatants from the cyanobacterial cultures. The CAS assay solution was made in two parts: (i) 6 ml 10 mM hexadecyltrimethylammonium bromide, 150 μl of an iron(III) solution (10 mM $FeCl_3$, 100 mM KCl), and 7.5 ml 2 mM CAS and (ii) 4.307 g piperazine (anhydrous) and 75 ml 1 M HCl. These two solutions were combined with ddH_2O to bring the total volume of the final solution to 100 ml. From each stationary-phase culture assayed, a 1-ml sample was taken and centrifuged at 14,000 × *g* in an Eppendorf (Westbury, NY) MiniSpin Plus. Next, 0.5 ml of each supernatant was taken and mixed with 0.5 ml of CAS assay solution in a 2-ml cuvette. The samples were allowed to equilibrate for 20 min at room temperature, and absorbance at 630 nm was measured using a Bio-Rad (Hercules, CA) SmartSpec3000 spectrophotometer. Because CAS is blue when complexed with hexadecyltrimethylammonium bromide and $FeCl_3$, a lowered absorbance at 630 nm suggests that other compounds in the assay solution, such as siderophores, have liberated CAS-bound iron and have complexed with it themselves. Visual inspections of samples at 20 min, 10 h, and 24 h were consistent at all time points. For cultures grown in GN, a GN-plus-CAS solution was used as a blank, whereas for SN-grown cultures, a SN-plus-CAS blank was used. For these experiments, *Synechococcus* sp. strain PCC7002, which is known to produce multiple siderophores (33), was used as a control. For each tested strain, two independent replicate cultures for each iron concentration were examined.

DNA extraction. An adaptation of the method described by Neilan et al. (19) for the extraction of PCR-quality DNA from cyanobacterial cultures was employed. A sample of cells (~150 mg of cell mass for filamentous strains; ~3 ml of stationary-phase culture for unicellular strains) was retrieved and pelleted by centrifugation for 5 min at 14,000 × *g* in an Eppendorf (Westbury, NY) MiniSpin Plus. The supernatants were decanted and the pellets washed with 500 μl of 5 M NaCl. The samples were then resuspended in 200 μl of well-mixed Instagene matrix (Bio-Rad, Hercules, CA), mixed with 20 μl of Triton X-100, and vortexed for 1 min. Next, the samples were heated for 30 min at 60°C while being mixed at 800 rpm in an Eppendorf Thermomixer R. Subsequent to this heating step, cells were centrifuged for 1 min at 14,000 × *g* and then heated for 15 min at 99°C while being mixed at 800 rpm. Following the extraction, the samples were centrifuged for 5 min at 14,000 × *g*, and the supernatants were transferred to 1.5-ml tubes for storage at −20°C. Prior to usage as a PCR template, aliquots from the Instagene preparations were diluted 10-fold into sterile ddH_2O.

PCRs. All extracts were checked for quality by PCR amplification of the 16S rRNA gene with the universal bacterial primers 8F (5′-AGAGTTTGATCCTG GCTCAG-3′) and 1492R (5′-GGTTACCTTGTTACGACTT-3′). After the quality of DNA extracts was confirmed, PCRs targeting NRPS and PKS genes were performed. The NRPS primers used for amplification were MTF2 (5′-GC NGGYGGYGCNTAYGTNCC-3′) and MTR (5′-CCNCGDATYTTNACYTG-3′) (18). The PKS primers used were DKF (5′-GTGCCGGTNCCRTGNGYY TC-3′) and DKR (5′-GCGATGGAYCCNCARCARYG-3′) (17). These NRPS and PKS primers target the A and KS domains, respectively. All PCRs were performed in either an Eppendorf (Westbury, NY) Mastercycler or a Bio-Rad (Hercules, CA) Icycler, and products were screened using 1% agarose gel electrophoresis. The following reaction mixtures and conditions were used for amplifying the 16S rRNA gene and the NRPS A and PKS KS domains. (i) For 16S rRNA gene PCR, the reaction mixture consisted of 30.8 μl of ddH_2O, 5 μl of 10× *Taq* buffer B, 3 μl of 15 mM $MgCl_2$, 1 U of *Taq* DNA polymerase (Promega, Madison, WI), 5 μl of 2 mM deoxynucleoside triphosphates, 1 μl of 50 μM 8F primer, 1 μl of 50 μM 1492R primer, and 5 μl of DNA template. The 16S rRNA gene PCR was run with the following holds and cycles: 94°C for 1 min; 30 cycles of 94°C for 45 s, 52.5°C for 45 s, and 72°C for 2 min; and 72°C for 7 min. (ii) For NRPS and PKS PCRs, the reaction mixture consisted of 33.8 μl of ddH_2O, 5 μl of *Taq* buffer A ($MgCl_2$ included), 1 U of *Taq* DNA polymerase (Fisher, Pittsburgh, PA), 5 μl of 2 mM deoxynucleoside triphosphates, 2 μl of 50 μM MTF2 or DKF primer, 2 μl of 50 μM MTR or DKR primer, and 5 μl of DNA template. The reactions were run with the following holds and cycles: 94°C for 5 min; 35 cycles of 94°C for 1 min, 50°C for 1 min for NRPS or 50.8°C for 1 min for PKS, and 72°C for 2 min for NRPS or 72°C for 1 min for PKS; and 72°C for 7 min. (iii) Annealing-temperature gradient NRPS and PKS PCRs were run to check for false negatives, as well as to confirm faint positives, in NRPS and PKS PCRs. The reaction mixtures used for the gradient PCR were the same as those used for initial NRPS and PKS PCRs. The NRPS and PKS annealing-temperature gradient PCRs had the same holds and cycles as the initial NRPS and PKS PCRs except that annealing-temperature gradients from 35 to 55°C and 40 to 55°C, respectively, were used.

Cloning and RFLP. Amplified NRPS and PKS fragments were purified from 1% agarose gels by using the QIAquick gel extraction kit (QIAGEN, Valencia, CA) and subsequently cloned with the TOPO TA cloning kit according to the manufacturer's instructions (Invitrogen, Carlsbad, CA). Ten clones were selected at random from the products of each cloning reaction for further analysis. Single colonies were placed into 100 μl of ddH_2O and heated on an Eppendorf (Westbury, NY) Mastercycler at 95°C for 10 min. These preparations were centrifuged at 2,750 × *g* for 5 min in an Eppendorf 5810R centrifuge, and the supernatant was used as a crude plasmid template for PCR. Inserts were amplified from the plasmid templates in 25-μl volumes, using standard M13 multiple-cloning site primers and PCR conditions. To eliminate identical clones, each PCR product was digested with a reaction mixture consisting of 0.5 μl of 20,000 U/ml TaqI restriction endonuclease (New England BioLabs, Beverly, MA), 1 μl of 10× TaqI buffer, 0.1 μl of 100 μg/ml bovine serum albumin, 3.4 μl of ddH_2O, and 5 μl of PCR product. The digests were incubated in an Eppendorf Mastercycler at 65°C for 1 h. After digestion, restriction fragment length polymorphism (RFLP) patterns were screened using 2% agarose gel electrophoresis.

Sequencing of NRPS and PKS fragments. The *Escherichia coli* strains possessing plasmids with unique fragments of the predicted sizes (~1 kb for NRPS; ~700 bp for PKS) were grown for 18 h in Super Broth at 37°C, and plasmids were extracted from these cultures by using the GeneMachines RevPrep Orbit (Genomic Solutions, Ann Arbor, MI) at the Josephine Bay Paul Center (JBPC) at the Marine Biological Laboratory. Sequencing was performed using the M13F and M13R reverse primers, the Big Dye Terminators volume 3.0 (ABI, Foster City, CA), and an ABI 3730 DNA analyzer at the JBPC. Sequencing reactions and cleanup were conducted according to the JBPC's instructions. For every cloned fragment, each strand was sequenced once entirely.

Sequence analyses. Sequences were assembled using Sequencher (Gene Codes Corporation, Ann Arbor, MI). Nucleotide sequences were translated into peptide sequences by using BioEdit version 5.0.7 (Tom Hall, Ibis Therapeutics, Carlsbad, CA). Peptide sequences were first aligned with ClustalX with the default settings (27) and then manually edited in BioEdit. Unrooted neighbor-joining phylogenies were constructed using MEGA version 2.1 with the Poisson

TABLE 2. PCR and bioassay results

Strain(s)	Axenic[a]	NRPS PCR	PKS PCR	Algicide against:		Siderophore
				PCC7002	PCC6803	
Chamaesiphon sp. strain PCC7430	N	+	+	−	−	−
Crocosphaera watsonii WH8501	Y	−	+	−	−	−
Crocosphaera sp. strains WH0001, WH0002, WH0401, WH0402, WH0403	Y	−	+	ND	ND	ND
Cyanothece sp. strain WH8901	Y	+	+	−	−	−
Cyanothece sp. strain WH8904	Y	−	−	+	+/−	−
Gloeobacter sp. strain PCC7421	Y	−	+	−	+/−	−
Gloeothece sp. strain PCC6501	Y	−	+	−	−	−
Anabaena sp. strain WH0404	N	+	+	ND	ND	ND
Calothrix sp. strain PCC7102	Y	+	+	+/−	−	+
Nodularia sp. strain WH0405	N	−	+	ND	ND	ND
Nostoc punctiforme ATCC 29133	Y	+	+	ND	ND	ND
Nostoc sp. strain ATCC 53789	Y	+	+	ND	ND	ND
Scytonema sp. strain PCC7110	Y	+	+	−	−	−
Lyngbya sp. strain PCC7419	Y	−	−	+/−	−	−
Microcoleus sp. strain PCC7420	Y	+	+	+	+	−
Spirulina sp. strain PCC6313	N	+	+	+	+	+
Trichodesmium thiebautii II-3	N	+	+	ND	ND	ND
Pleurocapsa sp. strain PCC7319	N	+	+	−	+	−
Stanieria sp. strain PCC7302	Y	+	+	−	+/−	−
Fischerella sp. strain PCC7414	Y	+	+	+	−	−

[a] Purities of cultures at time of study are also included. +, successful amplification for PCRs or activity observed for bioassays; −, no amplification for PCRs or no activity observed for bioassays; +/−, activity not observed in all replicate extracts of assay; ND, not determined; N, no; Y, yes.

correction model for amino acids, complete deletion handling of gaps, and a bootstrap consisting of 1,000 replications (12).

Cyanobacterial genome analysis. All cyanobacterial genome sequences available in GenBank were investigated for NRPS and PKS genes. Genome annotations, when available, were searched for predicted NRPSs and PKSs. For this work, positive NRPS annotations were included in the genomes as "nonribosomal peptide synthetase," "peptide synthetase," "microcystin synthetase B," "multifunctional peptide synthetase," or "nonribosomal peptide synthetase terminal component." The PKSs were annotated as "polyketide synthase," "polyketide synthase modular," "similar to polyketide synthase," or "probable modular polyketide synthase." To determine putative numbers of A and KS domains present in sequenced genomes, the A-domain *Stanieria* sp. strain PCC7302 fragment 3 and the KS-domain *Gloeothece* sp. strain PCC6501 fragment 2 were arbitrarily selected for use as query sequences for tBlastn (protein query sequence against translated nucleotide database) protocol BLAST (Basic Local Alignment Search Tool, National Center for Biotechnology Information [NCBI], Bethesda, MD) searches using all the available cyanobacterial genomes in NCBI's genome database. Hits that had E values of less than 1e−40 and had lengths within 20% of the complete query sequence (NRPS fragment length, 264 to 396 amino acids; PKS fragment length, 182 to 274 amino acids) were included as putative NRPS A and PKS KS domains. An unrooted neighbor-joining tree of one 16S rRNA gene from each of the analyzed cyanobacterial genomes was generated in MEGA version 2.1 (12), using the Kimura two-parameter model with complete deletion gap handling and 1,000-replication bootstrapping. One 16S rRNA gene copy was used from each genome, as all 16S rRNA gene copies from the same genome resulted in the same tree topology. In addition, comparison of our experimental results for *N. punctiforme* ATCC 29133 to its genome was conducted using the ERGO software package (Integrative Genomics, Chicago, IL).

For both the NRPS A domains and the PKS KS domains, one sequence representative of each distinct phylogenetic cluster was chosen arbitrarily from among the fragments sequenced in this study. Selected fragments were compared to the entire microbial genome database at NCBI, excluding the cyanobacteria (364 genomes in total) (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi). The two results with the lowest E values from each BLAST search were taken and used for additional phylogenetic analyses, which were conducted as described above. If the two most significant results came from the same genome, the output with the higher E value was ignored and the next best result from a different genome was taken. Building upon the initial NRPS and PKS analyses, fragments that were sequenced directly in this study or were obtained from the sequenced cyanobacterial genomes, in addition to the noncyanobacterial fragments most similar to cyanobacterial A and KS domains, were used for the construction of the phylogenies shown in the radial trees in Fig. 4 and 6.

Nucleotide sequence accession numbers. The sequences presented in this study have been deposited in GenBank and have accession numbers AY695274 through AY695303 (NRPS A domains) and AY695304 through AY695365 (PKS KS domains).

RESULTS

Screening for biologically active natural products. To show that the cyanobacteria in this study make natural products of pharmacological and ecological interest, we performed bioassays screening cell extracts for antialgal and iron binding compounds.

(i) Assaying cyanobacterial extracts for algicidal activity. With *Synechococcus* sp. strain PCC7002 and *Synechocystis* sp. strain PCC6803 as test cultures, the cyanobacterial extracts were assayed for growth-inhibitory effects (Table 2). Against *Synechococcus* sp. strain PCC7002, algicidal activity was observed with *Cyanothece* sp. strain WH8904, *Fischerella* sp. strain PCC7414, *Microcoleus* sp. strain PCC7420, and *Spirulina* sp. strain PCC6313 extracts. Variable growth inhibition of *Synechococcus* sp. strain PCC7002 was observed with *Calothrix* sp. strain PCC7102 and *Lyngbya* sp. strain PCC7419 extracts. In summary, 29% of the tested strains (4 of 14) inhibited the growth of *Synechococcus* sp. strain PCC7002, while 14% of the tested strains (2 of 14) gave variable results.

Against *Synechocystis* sp. strain PCC6803, the extracts from *Microcoleus* sp. strain PCC7420, *Pleurocapsa* sp. strain PCC7319, and *Spirulina* sp. strain PCC6313 exhibited algicidal activities. Furthermore, *Cyanothece* sp. strain WH8904, *Gloeobacter* sp. strain PCC7421, and *Stanieria* sp. strain PCC7302 extracts exhibited variable inhibitory effects on the growth of *Synechocystis* sp.

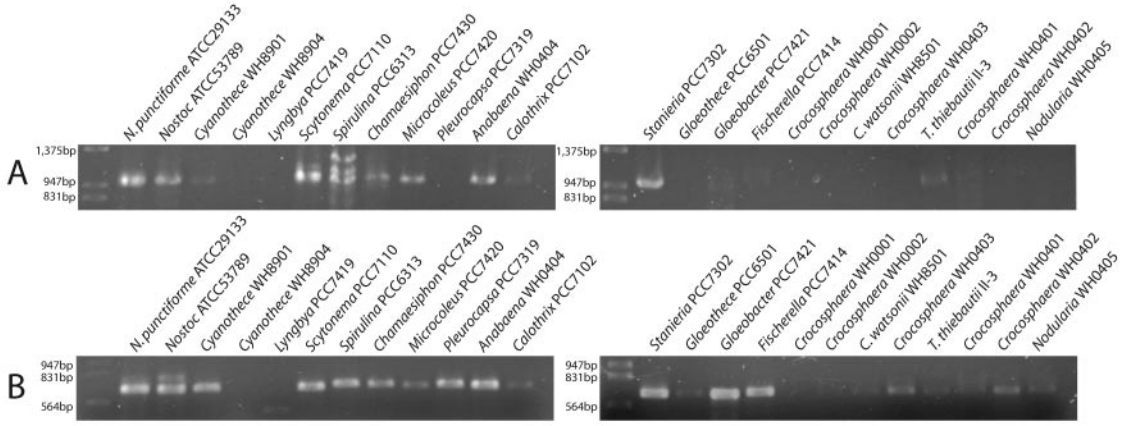

FIG. 1. Degenerate PCR results. NRPS PCR (A) and PKS PCR (B) results show amplification of the appropriate ~1-kb and ~700-bp fragments from many of the templates.

strain PCC6803. Thus, 21% of the tested strains (3 of 14) inhibited *Synechocystis* sp. strain PCC6803, while another 21% of the tested strains gave variable results.

Overall, two strains, *Microcoleus* sp. strain PCC7420 and *Spirulina* sp. strain PCC6313, consistently inhibited the growth of both *Synechococcus* sp. strain PCC7002 and *Synechocystis* sp. strain PCC6803.

(ii) Surveying cyanobacterial cultures for siderophores. Other than *Synechococcus* sp. strain PCC7002, which was used as the control culture and caused a detectable color change under both iron regimens, only two cultures had significantly different absorbance readings at 630 nm (indicative of siderophores production): *Calothrix* sp. strain PCC7102 (1/10 Fe) and *Spirulina* sp. strain PCC6313 (Fe replete) (Table 2). Visual inspection of these liquid CAS assay mixtures confirmed the spectrophotometric results, as these positive CAS assay mixtures had undergone color changes from blue to light purple or gray.

NRPS and PKS genetic screen. (i) Amplification of NRPS A domains. Putative NRPS A domains were successfully amplified from 54% (13 of 24) of the cyanobacterial DNA templates (Fig. 1; Table 2). NRPS gene fragments could not be amplified from *Crocosphaera* spp. (strains WH0001, WH0002, WH0401, WH0402, WH0403, and WH8501), *Cyanothece* sp. strain WH8904, *Gloeothece* sp. strain PCC6501, *Lyngbya* sp. strain PCC7419, and *Nodularia* sp. strain WH0405. The negative PCR results from the initial NRPS PCRs were confirmed through subsequent annealing-temperature gradient NRPS PCRs.

(ii) Amplification of PKS KS domains. Putative PKS KS domains were amplified from all cyanobacterial extracts except those of *Cyanothece* sp. strain WH8904 and *Lyngbya* sp. strain PCC7419 (Fig. 1; Table 2). Because initial PKS PCRs from *Crocosphaera* sp. strain WH0001 and WH0002 templates resulted in low yields of the desired product, annealing-temperature gradient PKS PCRs were conducted to confirm the presence of PKS KS domains in these organisms. These experiments confirmed the presence of PKS KS domains in *Crocosphaera* sp. strains WH0001 and WH0002 and confirmed the negative results obtained for *Cyanothece* sp. strain WH8904 and *Lyngbya* sp. strain PCC7419.

(iii) Cloning, RFLP, and sequencing of NRPS A and PKS KS domains. To verify that the amplified products were NRPS A and PKS KS domains, amplified fragments were cloned and sequenced. Up to six distinct clones were obtained from each cloning reaction mixture's 10-clone library by using RFLP. All unique clones of the appropriate size (~1 kb for NRPS A domains; ~700 bp for PKS KS domains) were sequenced. BLAST analysis in GenBank (tBlastn) confirmed that the sequenced clones were cyanobacterial NRPS A or PKS KS domains, as all sequences were most similar to cyanobacterial A or KS domains already present in GenBank.

(iv) Analysis of the sequenced cyanobacterial genomes. To determine how many NRPSs and PKSs most cyanobacteria possess, annotation and BLAST searches were performed on all available cyanobacterial genome sequences. Analysis of the sequenced cyanobacterial genomes showed that NRPSs and PKSs are absent in some cyanobacteria and numerous in others (Table 3). *Nostoc punctiforme* ATCC 29133 possessed the greatest number of NRPS genes (17 genes) and PKS genes (10 genes) in its annotation. Subsequent BLAST analyses found 42 NRPS A domains and 22 PKS KS domains in the *N. punctiforme* ATCC 29133 genome. Annotation and BLAST analyses also found that *Anabaena variabilis* ATCC 29413, *Crocosphaera watsonii* WH8501, *Nostoc* sp. strain PCC7120, and *Trichodesmium erythraeum* IMS-101 possess both NRPSs and PKSs. *Gloeobacter* sp. strain PCC7421 was annotated with PKSs but without NRPSs. BLAST results verified the *Gloeobacter* sp. strain PCC7421 annotation and found that *Synechococcus* sp. strain PCC7002 also possesses a single PKS gene. Furthermore, the absence of both NRPS and PKS genes from the genomes of *Prochlorococcus marinus* CCMP1375, *P. marinus* MED4, *P. marinus* MIT9313, *Synechococcus* sp. strain WH8102, *Synechococcus elongatus* PCC7942, *Synechocystis* sp. strain PCC6803, and *Thermosynechococcus elongatus* BP-1 was confirmed through BLAST analysis. A 16S rRNA-based phylogeny was generated for these cyanobacteria with sequenced genomes, and the presence of NRPS and PKS genes was mapped onto it (Fig. 2). This tree suggests that both NRPSs and PKSs are present in diverse cyanobacteria and that these genes are absent from the unicellular picoplankton.

TABLE 3. Genomic survey for NRPSs and PKSs

Genome	Facility[a]	CO[b]	Genome size (Mbp)	Genome status[c]	No. of:			
					Annotated NRPSs	Annotated PKSs	A-domain hits	KS-domain hits
Crocosphaera watsonii WH8501	DOE-JGI	Chro	6.2	D	9	4	18	5
Gloeobacter violaceus PCC7421	Kazusa	Chro	4.7	F	0	7	0	4
Synechococcus sp. strain PCC7002	PSU/HGC	Chro	3.4	F	NA[d]	NA	0	1
Synechococcus sp. strain WH8102	DOE-JGI	Chro	2.4	F	0	0	0	0
Synechococcus elongatus PCC7942	DOE-JGI	Chro	2.7	D	0	0	0	0
Synechocystis sp. strain PCC6803	Kazusa	Chro	3.6	F	0	0	0	0
Thermosynechococcus elongatus BP-1	Kazusa	Chro	2.6	F	0	0	0	0
Prochlorococcus marinus CCMP1375	CNRS	Chro	1.8	F	0	0	0	0
Prochlorococcus marinus MED4	DOE-JGI	Chro	1.7	F	0	0	0	0
Prochlorococcus marinus MIT9313	DOE-JGI	Chro	2.4	F	0	0	0	0
Anabaena variabilis ATCC 29413	DOE-JGI	Nos	7.1	D	6	1	20	7
Nostoc sp. strain PCC7120	Kazusa	Nos	6.4	F	7	6	11	6
Nostoc punctiforme ATCC 29133	DOE-JGI	Nos	9.1	D	17	10	42	22
Trichodesmium erythraeum IMS-101	DOE-JGI	Oscil	7.8	D	1	2	1	3

[a] DOE-JGI, U.S. Department of Energy Joint Genomes Institute; CNRS, Centre National de la Recerche Scientifique; Kazusa, Kazusa DNA Research Institute; PSU/HGC, Penn State University and Huada Genetics Center.
[b] Cyanobacterial order. Chro, *Chroococcales* (section I); Oscil, *Oscillatoriales* (section III); Nos, *Nostocales* (section IV).
[c] F, finished; D, draft.
[d] NA, annotation not available.

(v) Polypeptide alignment and phylogenetic analysis. Alignments of the NRPS A-domain and PKS KS-domain sequence sets obtained in this study showed that these domains are highly variable (data not shown). Of the sequenced A domains (~333 amino acids long) and KS domains (~233 amino acids long), only 36 and 25 conserved amino acids were present, respectively. This variation prohibited the prediction of end products synthesized by particular enzymes by the alignment of NRPS A and PKS KS domains to homologous fragments from characterized genes in GenBank. Hence, in an attempt to attribute functions to these identified genes and to further explore their novelty, phylogenies were constructed using both the NRPS A and the PKS KS domains from this study, genes with described function from GenBank, and all of the hits obtained in the BLAST searches of the cyanobacterial genomes.

In the NRPS A-domain phylogeny (Fig. 3), fragments from two proteins described in the literature, NcpA (13) and NosA (11), appeared to be most closely related to the sequenced fragments. NcpA and NosA are from *Nostoc* sp. strain ATCC 53789 and *Nostoc* sp. strain GSV224, respectively, and they synthesize 4-methylproline and nostopeptolide, respectively. Although these natural products have been fully characterized structurally, their biological activities are unknown. Furthermore, different A domains from each of these genes clustered in very different areas of the tree (e.g., NcpA and NosA fragments 1 through 3). While the BLAST genome analyses demonstrated that *N. punctiforme* contained the largest number of putative NRPS fragments, the tree shows that these fragments are quite diverse in sequence and uniformly distributed throughout the tree (Fig. 3).

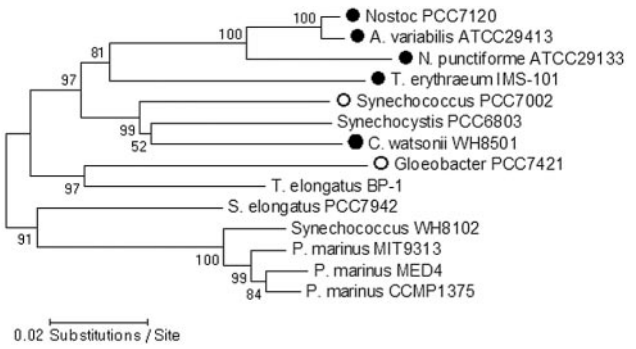

FIG. 2. 16S rRNA gene-based neighbor-joining phylogeny of the sequenced cyanobacterial genomes rooted with *Gloeobacter* sp. strain PCC7421. Bootstrap values above 50% are shown. Genomes preceded by open circles possess only PKSs, while genomes preceded by filled circles possess both NRPSs and PKSs. NRPSs and PKSs are absent from all other genomes.

Further analysis of the NRPS tree demonstrates that many of the sequences obtained in this study do not closely cluster with other cyanobacterial genes but instead form very tight clusters with other fragments obtained from the same organism (e.g., fragments 1 and 3 from *Stanieria* and fragments 1, 2, and 3 from *Microcoleus*). There are also examples of gene clusters entirely dominated by sequences from both filamentous and unicellular nitrogen-fixing isolates, suggesting a possible relationship between nitrogen fixation and the maintenance of these NRPS fragments (Fig. 3). Lastly, using the genome sequences as references allows for the easy identification of NRPS fragments that are not present in very closely related isolates of the same genus (e.g., *T. thiebautii* II-3 relative to *T. erythraeum* IMS-101).

A-domain fragments formed at least seven distinct phylogenetic clusters (Fig. 4). To determine if any of these clusters were unique to the cyanobacteria, representative sequences from each clade were selected for use in microbial genome BLAST searches. Because of the subjective nature of the clade designation above, additional sequences were screened against the microbial genomes to verify that our groupings were supported and that we analyzed the breadth of the sequence diversity contained in each clade. Phylogenies including the noncyanobacterial A domains most similar to the query se-

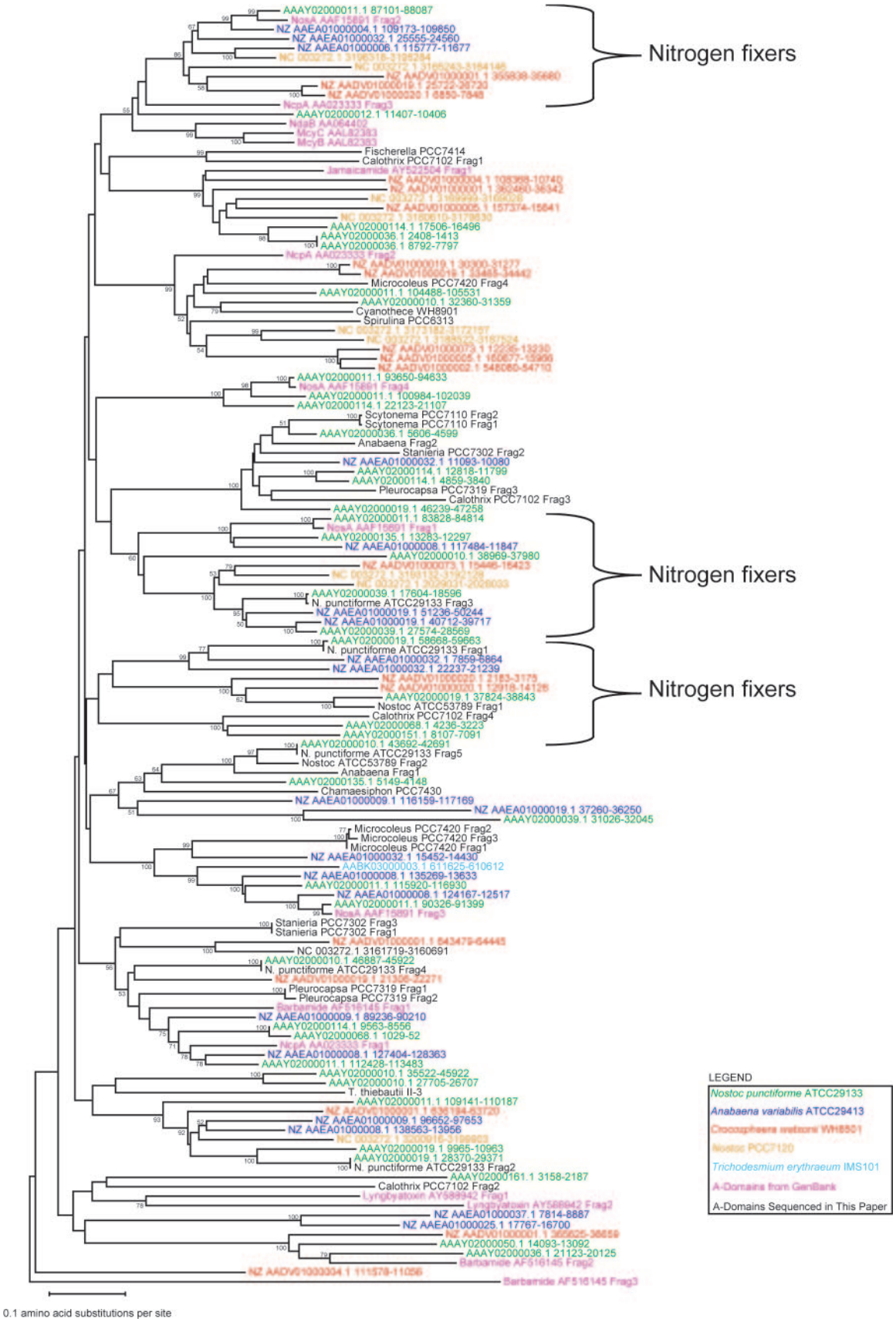

FIG. 3. Neighbor-joining phylogeny of translated NRPS A domains. Genomic sequences are followed by the GenBank accession numbers for the contigs containing them, as well as the fragment position (in base pairs) within the contig. All fragments from characterized genes are followed by their GenBank accession number. Only bootstrap values of ≥50% are shown.

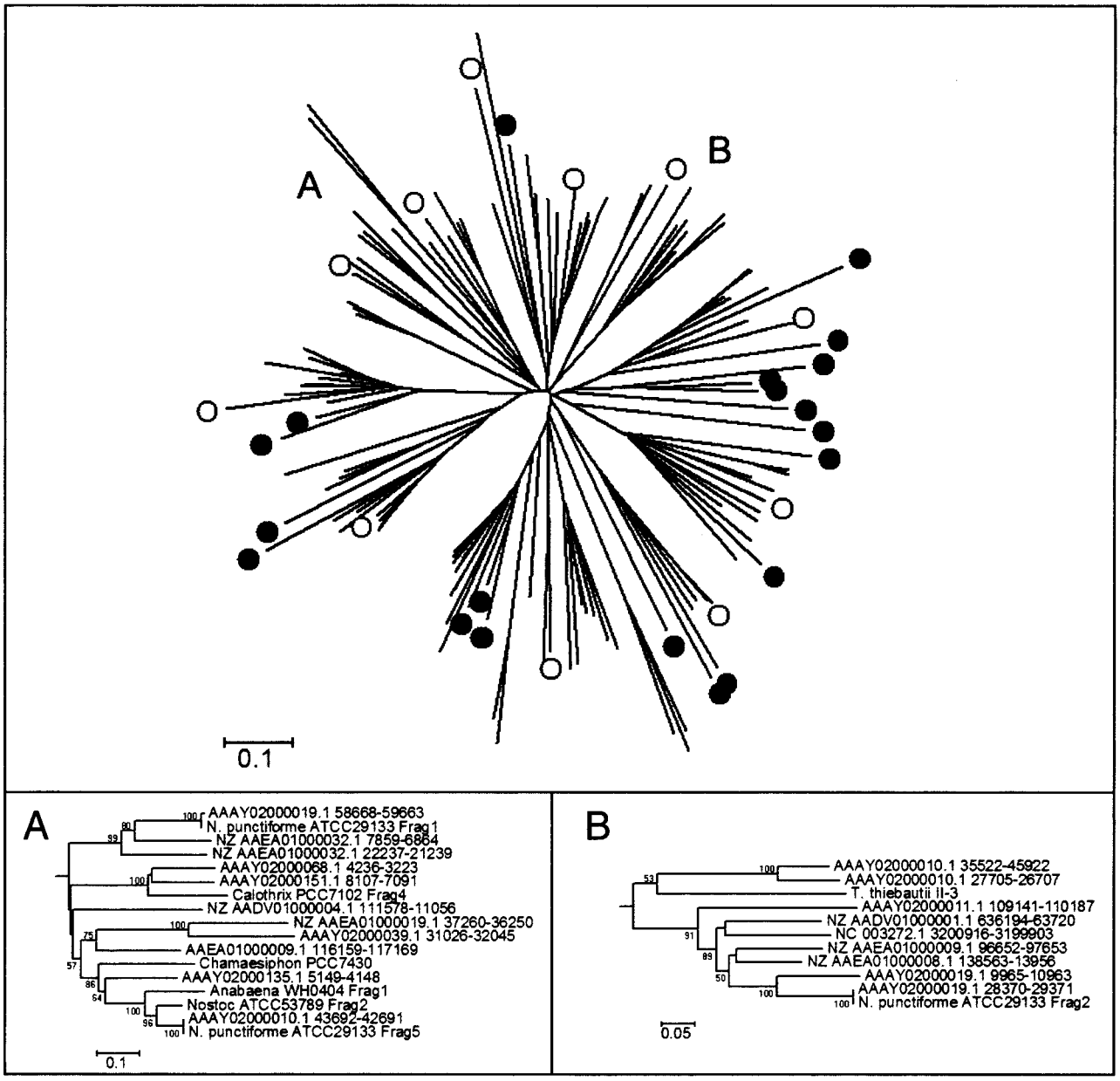

FIG. 4. Neighbor-joining phylogeny of all translated A domains sequenced in this study or taken from the sequenced cyanobacterial genomes. Open circles represent query sequences used for tBlastn BLAST searches against the entire NCBI microbial genome database, excluding the cyanobacteria. Black circles signify the two best hits from each BLAST search. Since certain hits were retrieved multiple times for different query sequences, the number of included hits is only 20. A and B, subtrees from the larger tree of A-domain clades which contain only cyanobacterial sequences. Only the bootstrap values of ≥50 are shown, and the bar is scaled to amino acid substitutions/site.

quences suggest that some A domains are common to diverse bacteria, whereas others are unique to the cyanobacteria.

Subtrees were constructed to identify the strains represented in the unique cyanobacterial A-domain clusters. As shown in Fig. 4, group A sequences were isolated from *Anabaena* sp. strain WH, A. *variabilis* ATCC 29413 (NZ_AAEA prefix), *Calothrix* sp. strain PCC7102, *Chamaesiphon* sp. strain PCC7430, *C. watsonii* WH8501 (NZ_AADV prefix), *Nostoc* sp. strain ATCC 53789, and *N. punctiforme* ATCC 29133 (AAAY_ prefix). In group B, sequences came from *A. variabilis* ATCC 29413, *C. watsonii* WH8501 (NC_ prefix), *Nostoc* sp. strain PCC7120 (NC_ prefix), *N. punctiforme* ATCC 29133, and *T. thiebautii* II-3.

In the phylogeny of the KS domains, many genes described in the literature clustered with the fragments identified in this study (Fig. 5), including the genes for EpoC (26), which is involved in epothilone synthesis in *Sorangium cellulosum*; Irp1 (GenBank accession no. AE013844, positions 75 to 9566), which is involved in yersiniabactin synthesis in *Yersinia pestis* KIM; and McyD (28), which is involved in microcystin synthe-

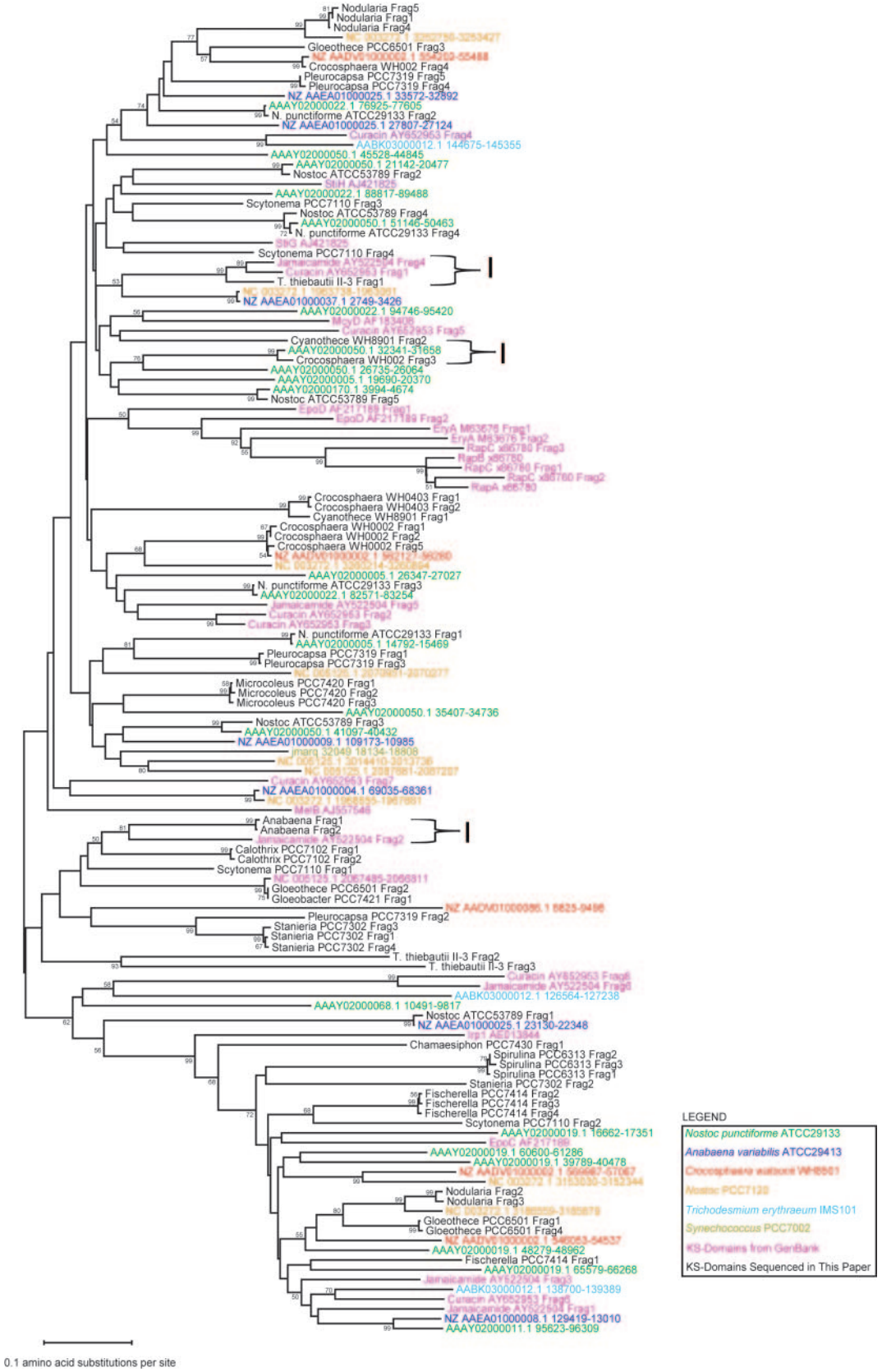

FIG. 5. Neighbor-joining phylogeny of translated PKS KS domains. Genomic sequences are followed by the GenBank accession numbers for the contigs containing them, as well as the fragment position (in base pairs) within the contig. All fragments from characterized genes are followed by their GenBank accession number. Only bootstrap values of ≥50% are shown.

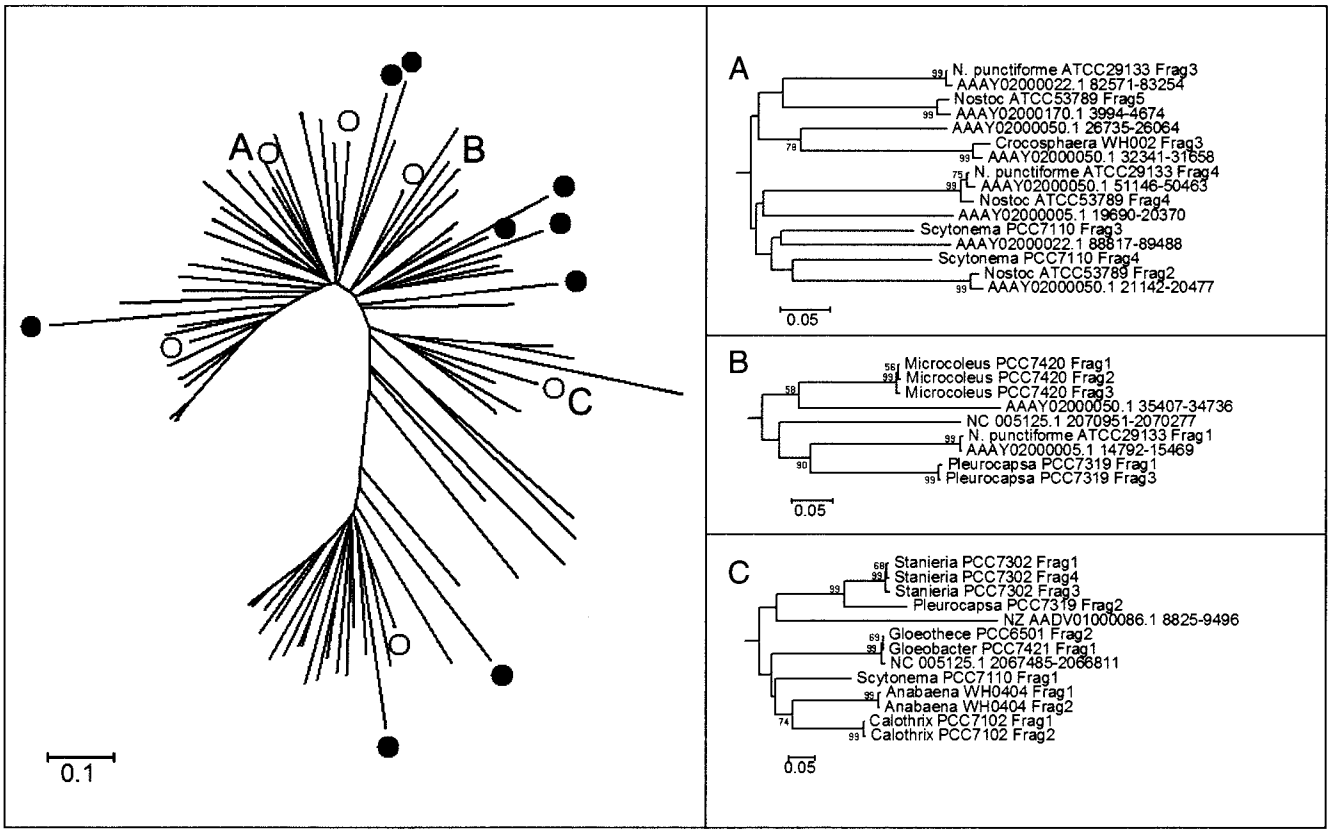

FIG. 6. Neighbor-joining phylogeny of all translated KS domains sequenced in this paper or taken from the sequenced cyanobacterial genomes. Open circles represent query sequences used for tBlastn BLAST searches against the entire NCBI microbial genome database, excluding the cyanobacteria. Since certain hits were retrieved multiple times for different query sequences, the number of included hits is only 9. Black circles signify the two best hits from each BLAST. A, B, and C, subtrees from the larger tree of KS-domain clades which contain only cyanobacterial sequences. Only bootstrap values of ≥50 are shown, and the bar is scaled to amino acid substitutions/site.

sis in *Microcystis aeruginosa* PCC7806. As was seen in the NRPS A-domain phylogeny, there were many examples of gene fragments obtained from one species forming a tightly clustering clade (e.g., *Stanieria*, *Fischerella*, *Crocosphaera*, and *Microcoleus*) in the KS-domain phylogeny. However, numerous KS domains cloned from strains closely related to cyanobacteria with sequenced genomes were not present in the relevant sequenced genomes (e.g., *T. thiebautii* II-3 fragments were very dissimiliar to *T. erythraeum* IMS-101 fragments, suggesting that these organisms do not possess orthologous PKSs [this and other examples are marked in Fig. 5 as grouping I]).

Although we were able to differentiate KS-domain phylogeny into at least six groups, the cluster relationships were less clear than was seen for the NRPS A domains (Fig. 6). As with the A domains, phylogenetic analysis including most similar noncyanobacterial KS domains suggests the presence of common bacterial and unique cyanobacterial KS domains within the cyanobacteria. Three groups that did not possess any noncyanobacterial sequences were identified. Group A consisted of *Crocosphaera* sp. strain WH0002, *Nostoc* sp. strain ATCC 53789, *N. punctiforme* ATCC 29133, and *Scytonema* sp. strain PCC7110 sequences. Group B contained *Microcoleus* sp. strain PCC7420, *Nostoc* sp. strain PCC7120, *N. punctiforme* ATCC 29133, and *Pleurocapsa* sp. strain PCC7319 fragments. *Anabaena* sp. strain WH, *Calothrix* sp. strain PCC7102, *C. watsonii* WH8501, *Gloeobacter* sp. strain PCC7421, *Gloeothece* sp. strain PCC6501, *Nostoc* sp. strain PCC7120, *Pleurocapsa* sp. strain PCC7319, *Scytonema* sp. strain PCC7110, and *Stanieria* sp. strain PCC7302 fragments comprised group C.

(vi) Mapping of *N. punctiforme* fragments. Detailed ERGO-based analysis of the genomic regions that contained our sequenced fragments from *N. punctiforme* found that A-domain fragments 1 and 2 mapped in different Nos biosynthetic clusters in contig 0606 (genes RNPU04187 and RNPU04176, respectively). Fragment 3 was found to be in an NRPS gene (RNPU04915) directly upstream of a large PKS gene representing a potential mixed NRPS-PKS cluster. Fragments 4 and 5 mapped in different A domains within the same gene (RNPU03419). For the KS domains, fragment 1 was present in a PKS (RNPU07181), fragment 2 was present in the *mtaE* myxothiazol biosynthesis gene (RNPU04192), fragment 3 was only partially represented at a contig junction and hence was not analyzed, and fragment 4 was in a PKS gene (RNPU05551).

DISCUSSION

Cyanobacteria are globally important primary producers that may alter components of their habitats, such as the densities of competitors or predators or the availability of critical

nutrients such as trace metals, by synthesizing natural products. Characterizing the biosynthetic capabilities of diverse cyanobacteria is crucial to understanding the full ecological impacts and niches of these organisms. In addition, the exploration of cyanobacterial natural products may result in the discovery of compounds with practical applications for society.

Cyanobacterial natural products. The results of this paper and others (2) suggest that the cyanobacteria are rich in natural products. In this study, half of the strains assayed for natural products (7/14) had both NRPS and/or PKS genes and biologically detectable activities. However, further experimentation is necessary to determine if NRPS and/or PKS genes underlie the detected activities. This point is underscored by the fact that both *Lyngbya* sp. strain PCC7419 and *Cyanothece* sp. strain WH8904 were negative for NRPS and PKS genes but both had detectable activities, suggesting that the activities detected were being produced by systems other than PKS and NRPS genes. Ultimate confirmation of the relationship between the existence of the NRPS and PKS genes and the biological activity would require gene knockouts in amenable strains and/or gene expression experiments (using the fragments defined in this study) to connect biochemical activities to specific genes, as well as to determine the signals that trigger expression.

NRPS and PKS genes are present in diverse cyanobacteria. This study's survey of the sequence and diversity of NRPS and PKS genes in laboratory cyanobacterial cultures demonstrated that the degenerate PCR screening, cloning, and sequencing of the fragments constitute a robust method to define potential gene targets in diverse cyanobacterial lineages. Although the method is quite effective, the NRPS PCR results for *C. watsonii* WH8501 (Fig. 1) suggest that the MTF2/MTR primer set and its standard reaction conditions are not optimal for all cyanobacterial A domains. Our data imply that the other studied *Crocosphaera* strains, as well as the other cyanobacterial genera that were found not to possess NRPSs, may actually have these genes.

Although we screened both axenic and nonaxenic strains of cyanobacteria, all of the fragments obtained were most similar to cyanobacterial sequences as determined using BLAST analyses, suggesting that the amplified products were cyanobacterial in origin. Furthermore, sequencing of the amplified products showed that there is remarkable diversity in the NRPS and PKS fragments obtained even in very closely related cyanobacteria (e.g., PKS fragments from *T. thiebautii* II-3 and *T. erythraeum* IMS-101).

Despite the distribution of these genes throughout the five major sections of cyanobacteria (8), NRPSs are noticeably absent in *Gloeobacter* sp. strain PCC7421 (Tables 2 and 3), which is thought to be the most basal of all extant cyanobacteria (29). The prevalence and diversity of the NRPS and PKS sequences described in this study suggest that NRPSs and PKSs may have been present in the cyanobacterial common ancestor. This also suggests that these genes may have been lost from *Gloeobacter* sp. strain PCC7421 during its evolution or introduced into the rest of the cyanobacterial lineage after its divergence from *Gloeobacter* sp. strain PCC7421.

The absence of NRPS and PKS genes from strains in genera known to commonly harbor these genes (e.g., *Cyanothece* sp. strain WH8904 and *Lyngbya* sp. strain PCC7419 (Fig. 1) (see reference 6 for an example of a mixed NRPS-PKS system in *Lyngbya majuscula*) and from strains from the genera *Arthrospira*, *Gloeocapsa*, *Microcystis*, and *Nostoc* (8) supports the assertion that these genes can be lost from genomes (2). An alternative interpretation of the variability of the presence of NRPSs and PKSs within cyanobacterial genera is the possibility that these genes are not ancestral to the cyanobacterial lineage but instead have been recently introduced into multiple genera. This argument is less parsimonious than gene loss because it requires numerous instances of genetic transfer, and it is not supported by the closer relationships of the sequences obtained in this project to each other than to noncyanobacterial NRPS A and PKS KS domains. Whether the observed heterogeneity in the presence of NRPS and PKS genes in closely related strains represents natural variation or loss of these genes during laboratory culture must be addressed by surveying more recent isolates from these genera.

Sequence and genome analyses. Analysis of the sequenced cyanobacterial genomes provides further insight into the distribution and abundance of NRPS and PKS genes in the cyanobacteria. Among the 14 sequenced cyanobacterial genomes, the numbers of NRPS and PKS genes present ranged from 0 for both genes to 17 for NRPSs and 10 for PKSs (Table 3). Why these genes are distributed this way within the cyanobacteria is an important question that has substantial bearing on our understanding of the ecology and evolution of the cyanobacterial NRPS and PKS genes. Although no obvious answer to this question can be determined from phylogenetic analysis of the cyanobacteria with sequenced genomes (Fig. 3), these data suggest that morphological and physiological complexity may be associated with the possession of these genes, as well as with the total number of these genes contained within a genome. This assertion is supported by a significant correlation ($r = 0.796$, df = 12, $P < 0.0001$) (34) between genome size and total number of NRPS A and PKS KS domains. Similar correlations exist between genome size and other measures of secondary metabolic diversity, such as annotated NRPSs, annotated PKSs, total A domains, and total KS domains.

Alignment and phylogenetic analyses found the sequenced NRPS A and PKS KS domains to be quite diverse. Inclusion of A and KS domains from genes of known function and all other homologues in the cyanobacterial genome projects suggested that many of the enzymes identified in this study might synthesize novel compounds. Few instances of close relationship between the newly sequenced fragments and the functionally characterized fragments were evident (Fig. 3 and 5). Additionally, in many instances A and KS domains taken from the same characterized gene or operon appeared to be quite divergent from one another (e.g., NcpA and NosA A domains and EpoC and EpoD KS domains), implying that domains in the same gene or pathway can differ substantially in primary structure. Together, these findings suggest that amino acid sequence identity to individual A and KS domains from characterized genes is not yet a useful estimator of function for genes newly detected by NRPS and PKS PCRs in cyanobacteria because of the dearth of characterized cyanobacterial NRPSs and PKSs.

The data presented also suggest that the diversification of the cyanobacterial NRPS A and PKS KS domains (Fig. 3 and 5) does not parallel the evolution of the cyanobacteria them-

selves, as understood through 16S rRNA data (29). However, interpretation of these A- and KS-domain phylogenies is complicated by limited knowledge about the functions of the identified genes. Figures 3 and 4 suggest that multiple distinct NRPS and PKS gene families may exist within the cyanobacteria and that it may be necessary to determine these functional classes to fully understand the evolution of these genes. Recent research using the 16S rRNA, *rpoC*, and *mcy* genes to interpret the evolution of microcystin synthetases has shown that it is likely that these genes are ancestral to a broad lineage of microcystin- and nodularin-synthesizing cyanobacteria (21). If other NRPS and PKS functional lineages exist in the cyanobacteria, similar approaches using multiple genes may be the most successful in parsing out the broader evolutionary history of NRPS and PKS genes in the cyanobacteria.

Genomic analyses showed that NRPS A and PKS KS domains were abundant in certain sequenced cyanobacteria, in particular *N. punctiforme* (others have noted this as well [15]), *A. variabilis*, and *C. watsonii*. The relationships of fragments within the same organism or genomic region were found to be highly variable, with some fragments being nearly identical (e.g., A domains from *N. punctiforme* contig AAAY02000036.1 [positions 2408 to 1413 and 8792 to 7797]) and others being highly diverged (e.g., A domains from *A. variabilis* contig NZ_AAEA01000032.1 [positions 25555 to 24560 and 22237 to 21239]). Phylogenies including the most closely related noncyanobacterial A and KS domains suggest that unique NRPSs and PKSs may exist within the cyanobacteria. Furthermore, the abundance of these cyanobacterium-specific fragments in diazotrophs may indicate a relationship between the ability to fix nitrogen and the capacity to synthesize unique natural products. However, this assertion does not hold true with the PKS groups B and C, as they include the unicellular nondiazotrophic *Pleurocapsa* sp. strain PCC7319 (Fig. 6). Furthermore, group B includes other nondiazotrophs, such as *Gloeobacter* sp. strain PCC7421 and *Stanieria* sp. strain PCC7302, making it difficult to hypothesize what the functional role for group C sequences might be. These results indicate that the cyanobacteria may be an untapped resource for novel NRPSs and PKSs, as well as the natural products that these genes synthesize.

Identity between our sequenced *N. punctiforme* A- and KS-domain fragments and those present in the sequenced genome validates the use of these degenerate primers for detection of NRPSs and PKSs in the cyanobacteria. Our findings suggest that if one were to generate a saturating clone library, this approach could be used to obtain most, if not all, of the NRPS and PKS genes present in the genome of a given cyanobacterium. This library combined with a fosmid or bacterial artificial chromosome clone library could allow for rapid and economical identification and localization of natural product gene clusters in the absence of a complete genome sequence.

The data presented here highlight the need for future studies to define the linkages between cyanobacterial ecology, the factors driving diversification and maintenance of NRPS and PKS genes, and the physiological roles of the compounds produced.

ACKNOWLEDGMENTS

We thank Frederica Valois and Nan Trowbridge for technical assistance, Stefan Sievert for certain reagents, and Michael Thomas and the members of the Waterbury and Webb laboratories for stimulating discussions and comments throughout this project.

I.M.E. was partially supported through a Woods Hole Oceanographic Institution Summer Fellowship. This work was funded by a WHOI internal grant, Woods Hole Center for Oceans and Human Health (OCE-0430724), and an NSF grant (OCE-0352241) to E.A.W.

REFERENCES

1. **American Type Culture Collection.** 19 August 2004, posting date. ATCC: the global bioresource center. American Type Culture Collection, Manassass, Va. [Online.] http://www.atcc.org.
2. **Burja, A. M., B. Banaigs, E. Abou-Mansour, J. G. Burgess, and P. C. Wright.** 2001. Marine cyanobacteria—a prolific source of natural products. Tetrahedron **57:**9347–9377.
3. **Burns, B. P., A. Seifert, F. Goh, F. Pomati, A. Jungblut, A. Serhat, and B. A. Neilan.** 2005. Genetic potential for secondary metabolite production in stromatolite communities. FEMS Microbiol. Lett. **243:**293–301.
4. **Cane, D. E., and C. T. Walsh.** 1999. The parallel and convergent universes of polyketide synthases and nonribosomal peptide synthetases. Chem. Biol. **6:**R319–R325.
5. **Cane, D. E., C. T. Walsh, and C. Khosla.** 1998. Harnessing the biosynthetic code: combinations, permutations, and mutations. Science **282:**63–68.
6. **Chang, Z., P. Flatt, W. H. Gerwick, V. A. Nguyen, C. L. Willis, and D. H. Sherman.** 2002. The barbamide biosynthetic gene cluster: a novel marine cyanobacterial system of mixed polyketide synthase (PKS)-non-ribosomal peptide synthetase (NRPS) origin involving an unusual trichloroleucyl starter unit. Gene **296:**235–247.
7. **Chang, Z., N. Sitachitta, J. V. Rossi, M. A. Roberts, P. M. Flatt, J. Jia, D. H. Sherman, and W. H. Gerwick.** 2004. Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacterium *Lyngbya majuscula*. J. Nat. Prod. **67:**1356–1367.
8. **Christiansen, G., E. Dittmann, L. Via Ordorika, R. Rippka, M. Herdman, and T. Borner.** 2001. Nonribosomal peptide synthetase genes occur in most cyanobacterial genera as evidenced by their distribution in axenic strains of the PCC. Arch. Microbiol. **176:**452–458.
9. **Crosa, J. H., and C. T. Walsh.** 2002. Genetics and assembly line enzymology of siderophore biosynthesis in bacteria. Microbiol. Mol. Biol. Rev. **66:**223–249.
10. **Edwards, D. J., B. L. Marquez, L. M. Nogle, K. McPhail, D. E. Goeger, M. A. Roberts, and W. H. Gerwick.** 2004. Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium *Lyngbya majuscula*. Chem. Biol. **11:**817–833.
11. **Hoffmann, D., J. M. Hevel, R. E. Moore, and B. S. Moore.** 2003. Sequence analysis and biochemical characterization of the nostopeptolide A biosynthetic gene cluster from *Nostoc* sp. GSV224. Gene **311:**171–180.
12. **Kumar, S., K. Tamura, I. B. Jakobsen, and M. Nei.** 2001. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics **17:**1244–1245.
13. **Luesch, H., D. Hoffmann, J. M. Hevel, J. E. Becker, T. Golakoti, and R. E. Moore.** 2003. Biosynthesis of 4-methylproline in cyanobacteria: cloning of *nosE* and *nosF* genes and biochemical characterization of the encoded dehydrogenase and reductase activities. J. Org. Chem. **68:**83–91.
14. **Mason, C. P., K. R. Edwards, R. E. Carlson, J. Pignatello, F. K. Gleason, and J. M. Wood.** 1982. Isolation of chlorine-containing antibiotic from the freshwater cyanobacterium *Scytonema hofmanni*. Science **215:**400–402.
15. **Meeks, J. C., J. Elhai, T. Thiel, M. Potts, F. Larimer, J. Lamerdin, P. Predki, and R. Atlas.** 2001. An overview of the genome of *Nostoc punctiforme*, a multicellular, symbiotic cyanobacterium. Photosyn. Res. **70:**86–106.
16. **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.
17. **Moffitt, M. C., and B. A. Neilan.** 2001. On the presence of peptide synthetase and polyketide synthase genes in the cyanobacterial genus Nodularia. FEMS Microbiol. Lett. **196:**207–214.
18. **Neilan, B. A., E. Dittmann, L. Rouhiainen, R. A. Bass, V. Schaub, K. Sivonen, and T. Borner.** 1999. Nonribosomal peptide synthesis and toxigenicity of cyanobacteria. J. Bacteriol. **181:**4089–4097.
19. **Neilan, B. A., D. Jacobs, and A. E. Goodman.** 1995. Genetic diversity and phylogeny of toxic cyanobacteria determined by DNA polymorphisms within the phycocyanin locus. Appl. Environ. Microbiol. **61:**3875–3883.
20. **Orcutt, K. M., U. Rasmussen, E. A. Webb, J. B. Waterbury, K. Gundersen, and B. Bergman.** 2002. Characterization of *Trichodesmium* spp. by genetic techniques. Appl. Environ. Microbiol. **68:**2236–2245.
21. **Rantala, A., D. P. Fewer, M. Hisbergues, L. Rouhiainen, J. Vaitomaa, T. Borner, and K. Sivonen.** 2004. Phylogenetic evidence for the early evolution of microcystin synthesis. Proc. Natl. Acad. Sci. USA **101:**568–573.
22. **Rippka, R., and I. Iteman.** 7 January 2004, posting date. Pasteur Culture Collection of cyanobacteria. Pasteur Culture Collection, Paris, France. [Online.] http://www.pasteur.fr/recherche/banques/PCC.
23. **Schwarzer, D., and M. A. Marahiel.** 2001. Multimodular biocatalysts for natural product assembly. Naturwissenschaften **88:**93–101.

24. **Schwyn, B., and J. B. Neilands.** 1987. Universal chemical assay for the detection and determination of siderophores. Anal. Biochem. **160:**47–56.
25. **Staley, J. T., M. P. Bryant, N. Pfennig, and J. G. Holt (ed.).** 1989. Bergey's manual of systematic bacteriology, vol. 3. Williams and Wilkins, Baltimore, Md.
26. **Tang, L., S. Shah, L. Chung, J. Carney, L. Katz, C. Khosla, and B. Julien.** 2000. Cloning and heterologous expression of the epothilone gene cluster. Science **287:**640–642.
27. **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.
28. **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.
29. **Turner, S.** 1997. Molecular systematics of oxygenic photosynthetic bacteria. Plant Syst. Evol. Suppl. **11:**13–52.
30. **Waterbury, J. B., and R. Rippka.** 1989. Subsection I. Order Chroococcales Wettstein 1924, emend. Rippka et al., 1979, p. 1728–1746. *In* J. T. Staley, M. P. Bryant, N. Pfennig, and J. G. Holt (ed.), Bergey's manual of systematic bacteriology, vol. 3. Williams and Wilkins, Baltimore, Md.
31. **Waterbury, J. B., S. W. Watson, F. W. Valois, and D. G. Franks.** 1986. Biological and ecological characterization of the marine unicellular cyanobacterium *Synechococcus*, p. 71–120. *In* T. Platt and W. K. W. Li (ed.), Photosynthetic picoplankton, vol. 214. Department of Fisheries and Oceans, Ottawa, Canada.
32. **Webb, E. A., J. W. Moffett, and J. B. Waterbury.** 2001. Iron stress in open-ocean cyanobacteria (*Synechococcus*, *Trichodesmium*, and *Crocosphaera* spp.): identification of the IdiA protein. Appl. Environ. Microbiol. **67:**5444–5452.
33. **Wilhelm, S. W., and C. G. Trick.** 1994. Iron-limited growth of cyanobacteria: multiple siderophore production is a common response. Limnol Oceanogr. **39:**1979–1984.
34. **Zar, J.** 1999. Biostatistical analysis., vol. 4. Prentice Hall, Upper Saddle River, New Jersey.
35. **Zehr, J. P., J. B. Waterbury, P. J. Turner, J. P. Montoya, E. Omoregie, G. F. Steward, A. Hansen, and D. M. Karl.** 2001. Unicellular cyanobacteria fix N_2 in the subtropical North Pacific Ocean. Nature **412:**635–638.