

# Long-Chain *N*-Acyl Amino Acid Synthases Are Linked to the Putative PEP-CTERM/Exosortase Protein-Sorting System in Gram-Negative Bacteria<sup>∇†</sup>

Jeffrey W. Craig, Marisa A. Cherry, and Sean F. Brady\*

Laboratory of Genetically Encoded Small Molecules, Howard Hughes Medical Institute, The Rockefeller University, 1230 York Avenue, New York, New York 10065

Received 31 May 2011/Accepted 5 August 2011

**Clones that encode the biosynthesis of long-chain *N*-acyl amino acids are frequently recovered from activity-based screens of soil metagenomic libraries. Members of a diverse set of enzymes referred to as *N*-acyl amino acid synthases are responsible for the production of all metagenome-derived *N*-acyl amino acids characterized to date. Based on the frequency at which *N*-acyl amino acid synthase genes have been identified from metagenomic samples, related genes are expected to be common throughout the global bacterial metagenome. Homologs of metagenome-derived *N*-acyl amino acid synthase genes are scarce, however, within the sequenced genomes of cultured bacterial species. Toward the goal of understanding the role(s) played by *N*-acyl amino acids in environmental bacteria, we looked for conserved genetic features that are positionally linked to metagenome-derived *N*-acyl amino acid synthase genes. This analysis revealed that *N*-acyl amino acid synthase genes are frequently found adjacent to genes predicted to encode PEP-CTERM motif-containing proteins and, in some cases, other conserved elements of the PEP-CTERM/exosortase system. Although relatively little is known about the PEP-CTERM/exosortase system, its core components are believed to represent the putative Gram-negative equivalent of the LPXTG/sortase protein-sorting system of Gram-positive bacteria. During the course of this investigation, we were able to provide evidence that an uncharacterized family of hypothetical acyltransferases, which had previously been linked to the PEP-CTERM/exosortase system by bioinformatics, is a new family of *N*-acyl amino acid synthases that is widely distributed among the PEP-CTERM/exosortase system-containing *Proteobacteria*.**

Diverse microbial communities are now routinely explored and characterized through metagenomic analysis. Generally, DNA extracted from complex environmental samples (environmental DNA [eDNA]) is cloned into standard *Escherichia coli*-based vectors, and the resulting eDNA libraries are subjected to functional and/or sequence-based analyses. Activity-dependent screens of eDNA libraries that target antibacterial activity have been used to identify eDNA clones that contain genes encoding small-molecule biosynthetic machineries (3). To date, the most common small molecules reported from antibiosis screens of *E. coli*-based soil eDNA libraries are long-chain *N*-acyl amino acids (8). Extracts of antibacterially active eDNA clones have yielded several distinct varieties of *N*-acyl amino acids, including *N*-acyl derivatives of tyrosine, phenylalanine, tryptophan, and arginine (Fig. 1, compounds 1 to 4, respectively) (6, 7, 12). eDNA clones that produce *N*-acyl amino acids have been identified in metagenomic studies spanning geographically and compositionally distinct soils, suggesting that *N*-acyl amino acid-producing bacteria may be common to the majority of soil ecosystems. Despite the apparent ubiquity of *N*-acyl amino acids, their biological purpose and ultimate fate remain unclear.

The biosynthesis of all known metagenome-derived *N*-acyl amino acids is carried out by a diverse group of enzymes referred to as *N*-acyl amino acid synthases (NASs), which utilize acyl-(acyl-carrier proteins) (acyl-ACPs) and amino acids as substrates (4). Genes with homology to known eDNA-derived *N*-acyl amino acid synthase genes are scarce within the sequenced genomes of cultured bacteria, and in only one instance has a predicted NAS gene from a cultured organism been functionally verified (4). Consequently, appreciation of the microorganisms that produce *N*-acyl amino acids and their significance as members of microbial communities is limited. Most of what is currently known about bacterial NASs has been provided by structural, kinetic, and mechanistic studies of the eDNA-derived *N*-acyltyrosine synthase (NasY) FeeM (37). In particular, it has been noted that FeeM shares several important characteristics with *N*-acylhomoserine lactone synthases, including structural resemblance to the GCN5-related *N*-acyl transferases and the use of acyl-ACP rather than acyl-coenzyme A (CoA) as an acyl donor (4, 37).

The *N*-acylhomoserine lactone products of *N*-acylhomoserine lactone synthases are used as quorum-sensing signaling molecules by Gram-negative bacteria (32). The chemical similarities between *N*-acyltyrosines and *N*-acylhomoserine lactones are readily apparent, and it has been proposed that *N*-acyltyrosines may play a similar role in bacterial signaling (Fig. 1, compounds 1 and 5) (4). It has also been suggested that some eDNA-derived NASs might represent an expansion of the lipid biosynthetic machinery responsible for the synthesis of bacterial-type ornithine-containing lipids, which are struc-

\* Corresponding author. Mailing address: Laboratory of Genetically Encoded Small Molecules, The Rockefeller University, 1230 York Avenue, New York, NY 10065. Phone: (212) 327-8280. Fax: (212) 327-8281. E-mail: sbrady@rockefeller.edu.

† Supplemental material for this article may be found at <http://jbb.asm.org/>.

∇ Published ahead of print on 12 August 2011.

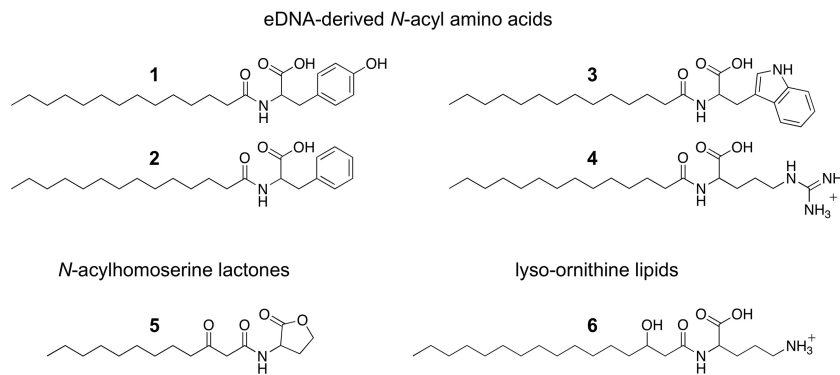


FIG. 1. Long-chain *N*-acyl amino acids isolated from antibacterially active eDNA clones: *N*-tetradecanoyl tyrosine (1), *N*-tetradecanoyl phenylalanine (2), *N*-tetradecanoyl tryptophan (3), and *N*-tetradecanoyl arginine (4). An *N*-acylhomoserine lactone (5) is produced by LasI from *P. aeruginosa* PAO1 [*N*-(3-oxododecanoyl)homoserine lactone]. A lyso-ornithine lipid (6) is produced by OlsB from *S. meliloti* strain 1021 [*N*-(3-hydroxyhexadecanoyl)ornithine].

tural lipids similar to phosphatidic acid (17). The first step of ornithine-lipid biosynthesis requires NAS-like activity and is performed by the enzyme OlsB, which catalyzes the formation of lyso-ornithine lipid from ornithine and 3-hydroxyacyl-ACP (Fig. 1, compound 6) (16).

The clustering of functionally related genes is a common feature of bacterial chromosomes (1, 42). We therefore hypothesized that an accounting of the genes surrounding known NAS genes might reveal genetic traits shared by NAS gene-containing bacteria and that this information might provide insight into the biological role(s) performed by *N*-acyl amino acids. To this end, we chose to analyze the genetic context of eDNA-derived NAS genes found in previous antibiosis screens. In total, 14 eDNA cosmids containing characterized NAS genes were fully sequenced and annotated to identify genetic elements conserved across this family of clones (Table 1). From this analysis, we have uncovered a genetic link between NAS genes and genes corresponding to conserved elements of the PEP-CTERM/exosortase system, a putative protein-sorting system associated with exopolysaccharide (EPS) production in Gram-negative bacteria. We also found that members of a previously uncharacterized group of proteins, the putative PEP-CTERM/exosortase system-associated acyl-

transferases (ExoATs; InterPro [integrative protein signature database] accession number IPR022484), constitute a novel family of *N*-acyl amino acid synthases that is widely distributed among the PEP-CTERM/exosortase system-containing *Proteobacteria*. Rigorous characterization of newly identified *N*-acyl amino acid synthase-containing bacterial species should allow for the functional relationship between *N*-acyl amino acid synthases and the PEP-CTERM/exosortase system to be fully revealed.

#### MATERIALS AND METHODS

**Sequencing and analysis of eDNA clones.** Fourteen cosmid eDNA clones containing previously characterized NAS genes were fully sequenced using a combination of 454 pyrosequencing (GS FLX+ system, GS FLX Titanium XLR70; Roche Diagnostics Corp.) and standard Sanger-type sequencing (Table 1). Sequences from each respective eDNA insert were analyzed for putative open reading frames (ORFs) using GLIMMER, version 3.02 (14), and SoftBerry FGENESB (bacterial operon and gene prediction program [SoftBerry, Inc., Mount Kisco, NY]). Predicted ORFs from each respective eDNA insert were annotated using InterProScan, which provides output data for multiple protein classification applications, including HMMProfam and HMMTigr (19, 22, 28).

**Construction of heterologous expression constructs.** Cosmids pCSL11, pCSL132, and pCSLC3 were used as PCR templates for the amplification of NAS and ExoAT genes. The following primer pairs were used for PCR amplification of individual NAS and ExoAT targets: for the ExoAT132 gene amplicon, ExoAT132-NdeI-F and ExoAT132-PstI-R; for the ExoATC3 gene amplicon, ExoATC3-NcoI-F and ExoATC3-SbfI-R; for the ExoAT11 gene amplicon, ExoAT11-NdeI-F and ExoAT11-PstI-R; for the NasY2 gene amplicon, NasY2-NdeI-F and NasY2-PstI-R; and for the NasR gene amplicon, NasR-NdeI-F and NasR-PstI-R (Table 2 gives the primer sequences). PCRs were carried out under the following conditions using a Biometra TGradient thermocycler: 100- $\mu$ l reaction mixtures consisted of 1.25 $\times$  Thermopol buffer (NEB), 500  $\mu$ M deoxynucleoside triphosphate (dNTP) mix, primers at 0.4  $\mu$ M each, 100 ng of template DNA, and 0.5  $\mu$ l per reaction of both Phusion Hot Start DNA Polymerase and Dynazyme II DNA Polymerase (NEB-Finnzymes); the PCR program consisted of 1 cycle at 95°C for 120 s, followed by 31 cycles of 30 s at 97°C, 30 s at 64°C, and 50 s at 72°C, followed by 120 s at 72°C. All PCR amplicons were subsequently digested with NdeI and PstI (except for the ExoATC3 gene amplicon, which required NcoI and SbfI) in preparation for cloning into the custom *E. coli* expression vector p4R-P<sub>lac</sub> (Epoch Biolabs, Inc., Sugarland, TX), which was prepared by reciprocal enzymatic digestion and calf intestinal phosphatase (CIP) treatment (NEB). Insert and vector samples were ligated (Fast-Link DNA Ligation kit; Epicentre), transformed into Transformax electrocompetent *E. coli* EC100 (Epicentre), and selected on Luria-Bertani agar (LB-agar) containing 200  $\mu$ g ml<sup>-1</sup> ampicillin.

**Bacterial cultures, organic extraction, and HPLC-MS analysis.** Fifty-milliliter cultures of *E. coli* transformed with NAS or ExoAT gene expression constructs were grown in LB medium containing 200  $\mu$ g ml<sup>-1</sup> ampicillin for 72 h at 30°C.

TABLE 1. eDNA cosmid clones used in this study

Clone	NAS	Amine substrate used by NAS	PEP-CTERM motif (no. of ORFs)	ExoAT and PrsKRT	Reference
pCSL12	NasY1	Tyrosine	Yes (2) <sup>a</sup>		6
pCSL132	NasY2	Tyrosine	Yes (3)	Yes	4
pCSL144	NasY3	Tyrosine			4
pCSLC2	FecM	Tyrosine			5
pCSLC3	NasY5	Tyrosine	Yes (2)	Yes	4
pCSLD10	NasY6	Tyrosine	Yes (1) <sup>a</sup>		4
pCSLF42	NasY7	Tyrosine			4
pCSLF43	NasY8	Tyrosine	Yes (1) <sup>a</sup>		4
pCSLG7	NasY9	Tyrosine	Yes (1) <sup>a</sup>		4
pCSLG10	NasY10	Tyrosine	Yes (1) <sup>a</sup>		4
pCSL1	NasW	Tryptophan			7
pCSL11	NasR	Arginine	Yes (2) <sup>a</sup>	Yes	7
pCSL142	NasP1	Phenylalanine			12
pECS	NasP2	Phenylalanine			13

<sup>a</sup> In these six clones, a gene encoding a putative PEP-CTERM motif protein is located directly adjacent to the previously characterized NAS gene.

TABLE 2. PCR primers used in this study

Primer	Sequence (5'-3') <sup>a</sup>
ExoAT132-NdeI-F.....	GATCTACATATGAAAAGCCCGCATCGACTTTGCGTC
ExoAT132-PstI-R.....	GATCTTCTGCAGTTTTTATTTAGTCACACCGGTTTGCAGCAGCTGTCC
ExoATC3-NcoI-F.....	GATCCCATGGAAAGCCCGCATCGACTTTGCGTCGCCAAGAG
ExoATC3-SbfI-R.....	CTAGCCTGCAGGGCTCAAAGTTCCCTCGCTTTTATTTAGTCACG
ExoAT11-NdeI-F.....	GATCTACATATGATTGTGCCCCGATAAGCCGCCACAG
ExoAT11-PstI-R.....	GATCTACTGCAGTCACGAGCCGCGAAACCACGTCAGCAG
NasY2-NdeI-F.....	GATCTTATATGTCTCTACCTGCTTACCACCTGAATCC
NasY2-PstI-R.....	GATCTACTGCAGCCAATCAAGACAAAGCAATAGCAAACCTTGTTCC
NasR-NdeI-F.....	GATCATCATATGCAGCCAGAGATCTTCGCGCTTCGTTATG
NasR-PstI-R.....	GATCTTCTGCAGTCTCAGTCGCTCACATCTCGCCGCGGAAC

<sup>a</sup> Relevant restriction sites are in boldface.

Individual 30-ml aliquots were then extracted 1:1 with ethyl acetate, dried under vacuum, and resuspended in methanol. Each sample was then subjected to analytical high-performance liquid chromatography–mass spectrometry (HPLC-MS) under the following conditions (Waters XBridge C<sub>18</sub> 5- $\mu$ m column [4.6 by 150 mm]; flow rate, 1.5 ml min<sup>-1</sup>): 3 min at 50:50 H<sub>2</sub>O-methanol with 0.1% formic acid, followed by a linear gradient from 50:50 H<sub>2</sub>O-methanol with 0.1% formic acid to 100% methanol with 0.1% formic acid over 12 min, followed by 100% methanol with 0.1% formic acid for 5 min. The same gradient conditions were used for preparatory HPLC-MS (Waters XBridge C<sub>18</sub> 5- $\mu$ m column [10 by 150 mm] with a flow rate of 7.0 ml min<sup>-1</sup>; Waters 2767 Sample Manager). All HPLC-MS data were recorded on a Waters Micromass ZQ instrument (Waters 515 HPLC Pump/Waters 2525 Binary Gradient Module) with a Waters 2996 Photodiode Array Detector and processed using MassLynx Mass Spectrometry software (version 4.1). Nuclear magnetic resonance (NMR) data were recorded on a 600-MHz spectrometer (Bruker) and processed using MestReNova software (version 6.1.0-6267).

**Nucleotide sequence accession numbers.** Sequence data have been submitted to the GenBank database under accession numbers JF429409 (CSL12), JF429410 (CSL132), JF429412 (CSL144), JF429413 (CSLC2), JF429414 (CSLC3), JF429415 (CSLD10), JF429416 (CSLF42), JF429417 (CSLF43), JF429404 (CSLG7), JF429405 (CSLG10), JF429407 (CSL1), JF429408 (CSL11), JF429411 (CSL142), and JF429406 (EC5).

## RESULTS

**Many eDNA-derived NAS genes are located near genes encoding putative PEP-CTERM motif-containing proteins.** *N*-Acyltyrosines are the most frequently encountered subset of *N*-acyl amino acids identified from activity-based screens of soil eDNA libraries. The corresponding NAS enzymes that produce *N*-acyltyrosines, referred to as NasYs, can be divided into three groups based on sequence homology: group 1, NasY1-NasY2, FeeM, NasY5, and NasY7; group 2, NasY6 and NasY8 to NasY10; and others, NasY3 (4). Two NasYs (NasY6 and NasY10, both from group 2) show low-level similarity (~15% identity) to individual *N*-acylhomoserine lactone synthases (4); however, the majority of NAS enzymes do not show similarity to known enzymes, and none is predicted to contain the Pfam “autoinducer synthetase” domain PF00765 that is commonly found in *N*-acylhomoserine lactone synthases. There are two known *N*-acylphenylalanine-producing NASs, NasP1 and NasP2 (NasA), that are nonhomologous to one another and lack similarity to the NasYs (12, 13). There are also singular examples of *N*-acyltryptophan-producing (NasW) and *N*-acylarginine-producing (NasR) NASs, neither of which resembles other previously described NAS enzymes (7). The 25- to 40-kb eDNA inserts on which these 14 NAS genes are found were each sequenced, annotated, and compared with one another to identify conserved genetic features.

Eight NAS gene-containing eDNA clones were found to contain at least one gene encoding a hypothetical protein bearing a PEP-CTERM motif (VPEP, Pfam PF07589; InterPro IPR011449) (Table 1 and Fig. 2A). The PEP-CTERM motif is a short C-terminal homology domain believed to be analogous to the LPXTG motif used for protein sorting in Gram-positive bacteria (18). The N-terminal regions of PEP-CTERM motif proteins are generally predicted to contain signal peptides but are otherwise variable. It has been proposed that the putative transpeptidase EpsH (exopolysaccharide locus protein H, InterPro IPR013426), or exosortase, is responsible for the proteolytic/transpeptolytic processing of PEP-CTERM motif proteins, most likely for targeting these proteins across the outer membrane (Fig. 2B) (18). In total, 13 genes encoding PEP-CTERM motif proteins were identified from within these eight eDNA clones. The N-terminal regions of these 13 hypothetical proteins are predicted to contain signal peptides but are not otherwise conserved. Most, however, are acidic proteins of similar sizes (~200 to 300 residues) that contain an abundance of residues used for O-linked and N-linked glycosylation (Ser/Thr and Asn, respectively). Within six of these eight clones, at least one gene encoding a putative PEP-CTERM motif protein was found directly adjacent to a previously characterized *NAS* gene.

Bacterial species that encode putative PEP-CTERM motif proteins are found predominantly in sediments, soils, and biofilms; they possess an outer membrane and invariably contain genes for exopolysaccharide (EPS) production (18). In the *Proteobacteria*, an intergenic region containing a putative *cis*-regulatory site with a sigma-54 ( $\sigma^{54}$ ) binding motif precedes the majority of genes encoding PEP-CTERM motif proteins. Haft et al. found that the phylogenetic profile of this presumed regulatory region was identical to that of three conserved protein families: (i) a TPR-repeat protein (PrsT, InterPro IPR014266), (ii) a transmembrane histidine kinase (PrsK, InterPro IPR014265), and (iii) a sigma-54-interacting response regulator (PrsR, InterPro IPR014264), the latter two of which form a two-component regulatory system. Phylogenetic profile comparisons such as this are used to infer functional relationships between genes and other genetic features through a pattern of cooccurrence across multiple species (25). The underlying methodology is premised on the assumption that elements with interrelated functions will be preferentially retained with one another through evolution and lateral transfer. Based on this phylogenetic association, the PrsK/PrsR two-component system was hypothe-

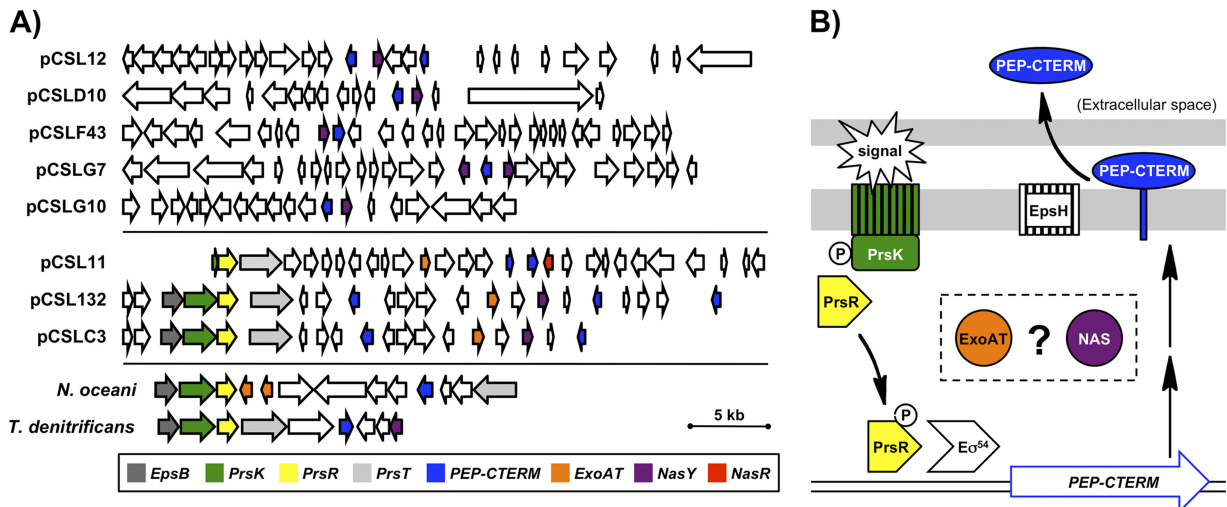


FIG. 2. (A) Open reading frame maps of the eDNA inserts containing previously identified *N*-acyl amino acid synthase genes (either NasY or NasR genes) and genes encoding PEP-CTERM motif proteins. These maps are divided into two sections to separate those that contain whole or partial genes for PrsT, PrsK, and PrsR (middle section) from those that do not (top section). For comparison, the bottom section shows similar regions from the genomes of *Nitrosococcus oceani* ATCC 19707 and *Thiobacillus denitrificans* ATCC 25259. (B) In this overview of the proposed PrsK/PrsR-mediated regulation of PEP-CTERM motif protein expression, an unknown signal initiates phosphorylation between PrsK and PrsR. Phosphorylated PrsR then binds to a proposed response regulator binding motif and activates the alternate  $\sigma^{54}$ -RNA polymerase holoenzyme ( $E\sigma^{54}$ ) in an ATP-dependent process leading to transcription of downstream gene(s) encoding PEP-CTERM motif protein(s). PEP-CTERM motif proteins, anchored to the plasma membrane by their PEP-CTERM motifs, are subsequently cleaved by the transpeptidase EpsH, presumably for targeting across the outer membrane. The contributions of ExoAT and NAS-like enzymes to this process are currently unknown.

sized to control the expression of PEP-CTERM motif proteins in a sigma-54-dependent manner (Fig. 2B) (18). The factor(s) responsible for the activation and/or repression of the PrsK histidine kinase sensor domain is currently unknown, making PrsK/PrsR a “ligand-orphaned” two-component system (11). The N-terminal regions of PrsK homologs are predicted to contain nine transmembrane helices (29). A similar nine-transmembrane helix motif is found in the N-terminal sensor domain of the histidine kinase LuxN from *Vibrio harveyi* (20), the cognate ligand of which is the *N*-acylhomoserine lactone AI-1 (3-hydroxybutanoyl homoserine lactone) (9, 34).

**Other PEP-CTERM/exosortase system components are found within NAS-containing clones.** In addition to EpsH, PrsT, PrsK, and PrsR, several other protein families are also associated with the PEP-CTERM/exosortase system, including a family of sugar transferases (EpsB, InterPro IPR017464) and an uncharacterized family of hypothetical acyltransferases (PEP-CTERM/exosortase system-associated acyltransferases, InterPro IPR022484), here referred to as ExoATs, that are members of the acyl-CoA *N*-acyltransferase superfamily (SSF55729) (Fig. 2B). Members of the ExoAT family are homologous to known *N*-acylhomoserine lactone synthases (LasI, Clusters of Orthologous Groups [COG] of Proteins database COG3916) (35, 36), and many are predicted to contain the Pfam autoinducer synthetase domain PF00765. Of the eight clones encoding PEP-CTERM motif proteins that were identified in this study, three were also found to contain whole or partial genes encoding homologs of PrsT, PrsK, and PrsR as well as a putative ExoAT.

**PEP-CTERM/exosortase system-associated acyltransferases possess *N*-acyl amino acid synthase activity.** Based on the predicted topological similarities between PrsK homologs and the nine-transmembrane histidine kinase LuxN and the sequence

similarities between ExoATs and *N*-acylhomoserine lactone synthases, we hypothesized that ExoATs might produce *N*-acylated small molecules that could function as the cognate ligands required for PrsK/PrsR activation. As an initial effort to explore this hypothesis, we chose to investigate the function of eDNA-derived ExoAT genes using *E. coli*-based heterologous expression experiments. From within the set of eDNA clones containing gene(s) coding for PEP-CTERM motif proteins, two NasY gene-containing eDNA clones (pCSL132 and pCSLC3) and the NasR gene-containing eDNA clone (pCSL11) were each found to contain an uncharacterized gene encoding a hypothetical ExoAT enzyme (Table 1 and Fig. 2A). Each of the eDNA-derived ExoAT genes was cloned into an *E. coli* expression vector downstream of the  $P_{tac}$  promoter, and the resulting ExoAT gene expression constructs were subsequently transformed into *E. coli*. Unexpectedly, analysis of the organic culture broth extracts of all three ExoAT gene expression constructs revealed the presence of long-chain *N*-acyl amino acid mixtures containing compounds identical by mass, retention time, and UV-visible light (Vis) absorption profile to *N*-acyl amino acids produced by previously characterized NAS enzymes (6, 7). In addition, a dominant component of each ExoAT-derived mixture was isolated and analyzed by one-dimensional (1D)  $^1\text{H}$  NMR spectroscopy, confirming that these clone-specific metabolites were indeed *N*-acyl amino acids (6, 7). The two ExoAT genes from the NasY gene-containing clones pCSL132 and pCSLC3 (the ExoAT132 and ExoATC3 genes, respectively) enabled *E. coli* to produce *N*-acyltyrosine mixtures while the ExoAT gene from the NasR gene-containing clone pCSL11 (harboring the ExoAT11 gene) afforded the production of *N*-acylarginines (Fig. 3).



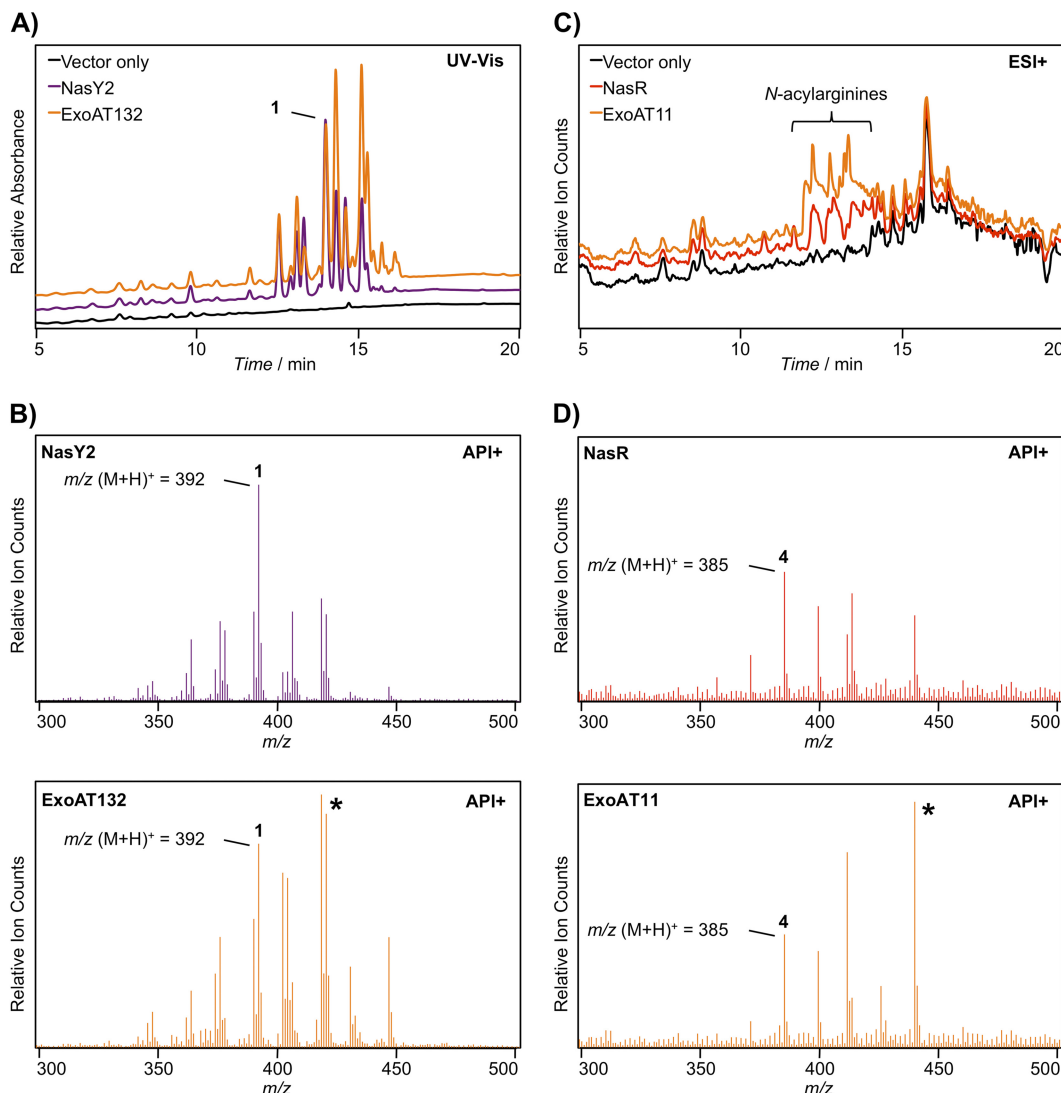


FIG. 3. (A) UV-Vis chromatograms from the HPLC-MS analysis of organic extracts of *E. coli* heterologous expression constructs containing the NasY2 or ExoAT132 gene or empty vector. Prominent signals from *N*-acyltyrosines occur between minutes 12 and 17. (B) Atmospheric pressure ionization-positive (API+) mode ionization data corresponding to minutes 11 to 18 of the HPLC-MS analysis of the NasY2 (top panel) and ExoAT132 (bottom panel) genes. The peak with  $m/z$  (M+H)<sup>+</sup> of 392, corresponding to 1 (*N*-tetradecanoyl tyrosine), is highlighted in both panels. The relatively larger peaks with  $m/z$  (M+H)<sup>+</sup> of 418 and 420 that are marked by a single asterisk in the ExoAT132 gene panel correspond to the saturated and monounsaturated 16-carbon *N*-acyl derivatives of tyrosine. (C) Relative ion counts (electrospray ionization-positive [ESI+] mode) over the course of the HPLC-MS analysis of organic extracts of *E. coli* heterologous expression constructs containing the NasR or ExoAT11 gene or empty vector. (D) API-positive mode ionization data corresponding to minutes 10 to 18 of the HPLC-MS analysis of the NasR (top panel) and ExoAT11 (bottom panel) genes. The peak with  $m/z$  (M+H)<sup>+</sup> of 385, corresponding to 4 (*N*-tetradecanoyl arginine), is highlighted in both panels. The largest peak in the ExoAT11 gene panel with  $m/z$  (M+H)<sup>+</sup> of 439, marked by a single asterisk, corresponds to the monounsaturated 18-carbon *N*-acyl derivative of arginine.

To investigate the functional relationship between NAS and ExoAT genes located on the same eDNA clone, the NasY2 gene from pCSL132 and NasR gene from pCSL11 were heterologously expressed in *E. coli*, and extracts of these cultures were compared directly with extracts of cultures expressing the ExoAT132 and ExoAT11 genes, respectively. Organic extracts from cultures expressing the NasY2 and ExoAT132 genes contained complex mixtures of long-chain *N*-acyltyrosines, and although the profiles of these mixtures were similar, they were not identical (Fig. 3A). For instance, while compound 1 [*N*-tetradecanoyl tyrosine; C<sub>14:0</sub>,  $m/z$  (M+H)<sup>+</sup> = 392] was a major

component of both extracts, the ExoAT132 gene extract contained relatively greater amounts of the saturated and monounsaturated 16-carbon *N*-acyl derivatives of tyrosine [C<sub>16:0</sub>,  $m/z$  (M+H)<sup>+</sup> = 420; C<sub>16:1</sub>,  $m/z$  (M+H)<sup>+</sup> = 418] (Fig. 3B). The NasY2 and ExoAT132 genes are located less than 2.5 kb apart from one another on clone pCSL132. These genes show little sequence similarity (14% translated and aligned identity and 34% similarity [Clustal W]), making their apparent functional equivalence all the more striking. Organic extracts from cultures expressing the NasR and ExoAT11 genes contained complex mixtures of long-chain *N*-acylarginines, including com-

pound 4 [*N*-tetradecanoyl arginine;  $C_{14:0}$ ,  $m/z$  (M+H)<sup>+</sup> = 385] (Fig. 3C). The ExoAT11 gene extract, however, contained relatively greater amounts of the 16-carbon and 18-carbon monounsaturated *N*-acyl derivatives of arginine [ $C_{16:1}$ ,  $m/z$  (M+H)<sup>+</sup> = 411;  $C_{18:1}$ ,  $m/z$  (M+H)<sup>+</sup> = 439] (Fig. 3D). The observed preferences of ExoAT132 and ExoAT11 for slightly longer acyl substrates than those utilized by NasY2 and NasR were reproducibly observed over multiple trials. The significance of such subtle differences is unclear at this time.

**ExoAT and NAS genes show distinct phylogenetic distribution patterns.** Members of the ExoAT gene family were first identified through their phylogenetic association with genomes containing the PEP-CTERM/exosortase system. In the case of all three NAS gene-containing eDNA clones from which we identified an ExoAT gene conferring the production of *N*-acyl amino acids, we also identified whole or partial genes corresponding to the conserved PEP-CTERM/exosortase system proteins PrsT, PrsK, and PrsR (Table 1 and Fig. 2A). To more fully understand the relationship between the PrsK/PrsR two-component system and *N*-acyl amino acid synthases, we searched for patterns in the phylogenetic distributions of ExoAT genes and NAS-like genes across the 69 species of *Proteobacteria* whose genomes contain both the PrsK and PrsR genes. Figure 4 illustrates this relationship as a 16S rRNA-based phylogenetic tree to which we have appended symbols indicating both the presence and number of putative ExoAT and NAS-like genes within the genome of each species.

The sequences of all previously described NASs were used as BLASTp search queries to identify additional NAS-like homologs. Our searches returned 18 potential group 1 NasY homologs, all of which belonged to species of PrsK/PrsR-containing *Proteobacteria*. These putative group 1 NasYs were the only NAS homologs to contribute to the relationships depicted in Fig. 4. Two putative group 2 NasYs were also identified, both of which are predicted NasY homologs encoded within the genome of “*Candidatus Solibacter usitatus* Ellin6076,” a member of the phylum *Acidobacteria*. The 45 ExoAT gene homologs represented in Fig. 4 constitute 90% of the ExoAT genes that have been identified within the sequenced genomes of *Proteobacteria* (50 ExoAT gene sequences are from members of the *Proteobacteria*, and one is from the cyanobacterium *Cyanothece* sp. strain PCC 7822). In total, 32 of the 69 PrsK/PrsR-containing species of *Proteobacteria* were found to contain at least one ExoAT or NAS-like gene, and 18 of those 32 species contained at least two such genes, in various combinations.

Both ExoAT and NAS-like genes were seen in the *Betaproteobacteria* and *Deltaproteobacteria* (Fig. 4). Eight of the 12 PrsK/PrsR-containing *Betaproteobacteria* species contained genes corresponding to both types of enzymes, while PrsK/PrsR-containing *Deltaproteobacteria* contained either ExoAT or NAS-like genes but never both. Members of the *Gamma-proteobacteria*, the largest group of PrsK/PrsR-containing *Proteobacteria*, did not contain NAS-like genes but occasionally possessed numerous ExoAT genes. Neither type of enzyme was found in the *Alphaproteobacteria*.

Next, we examined the promoter regions located immediately upstream of eDNA-derived genes encoding PEP-CTERM motif proteins in each of the NasY gene-containing eDNA clones in

order to identify the PEP-CTERM *cis*-regulatory site that is conserved among *Proteobacteria* (18). Five of seven such promoters from clones harboring group 1 NasY genes contained both the putative PrsR response regulator binding motif and the sigma-54 binding motif associated with these regulatory regions. This was not the case for any of four such promoters found on clones harboring group 2 NasY genes. Differences between group 1 and group 2 NasYs are likely reflective of the taxonomic division between the *Proteobacteria* and *Acidobacteria*, the latter of which is represented by few sequenced genomes despite being one of the most abundant, diverse, and widely distributed bacterial phyla throughout the environment (27, 38). Although the PEP-CTERM motif has been found in many divisions of Gram-negative bacteria, the PrsK/PrsR two-component system and the ExoAT-type and group 1 NasY-type *N*-acyl amino acid synthases have been found almost exclusively in *Proteobacteria*.

## DISCUSSION

Genes with interrelated biological functions are often clustered in bacterial genomes, and this feature can aid in the functional classification of uncharacterized genes and enzymes (30, 40). With this in mind, we began our investigation of bacterial *N*-acyl amino acids with the complete genetic characterization of NAS-containing eDNA clones. Our analysis indicates that NAS genes are commonly found adjacent to genes predicted to encode PEP-CTERM motif-containing proteins (8 of 14 clones examined), an association that was particularly strong among both group 1 and group 2 NasY gene-containing eDNA clones. This newly discovered link between bacterial NASs and PEP-CTERM motif proteins then led us to investigate the function of the PEP-CTERM/exosortase system-associated acyltransferase family of enzymes, which we have referred to as ExoATs. Three of the eDNA clones used in this study were found to contain ExoAT genes that conferred the production of long-chain *N*-acyl amino acids to *E. coli*, thus indicating that the ExoAT family of enzymes (InterPro IPR022484) is a new class of bacterial *N*-acyl amino acid synthases found predominantly in the PEP-CTERM/exosortase system-containing *Proteobacteria*. In the case of all three eDNA clones that encoded a previously characterized NAS and a previously uncharacterized ExoAT, both the NAS and the ExoAT were found to synthesize the same type of *N*-acyl amino acid with similar distributions of acyl-chain lengths and saturations.

Although never experimentally validated, the ExoAT family of enzymes has been hypothesized to contribute to the chemical modification of exopolysaccharide and biofilm structural components (HMM summary page for accession number TIGR03694 [http://www.jcvi.org/cgi-bin/tigrfams/HmmReportPage.cgi?acc=TIGR03694]). Several lines of evidence suggest that the *N*-acyl amino acid synthase activity of ExoATs and NASs observed in heterologous expression experiments is relevant to the native biological functions of these enzymes in environmental bacteria. First, studies have reported the isolation of bacterial *N*-acyl amino acids from cultured environmental isolates, yet the enzymatic origins of these metabolites were not described (26, 33, 41). Second, microbial enzymes that either hydrolytically degrade or fur-

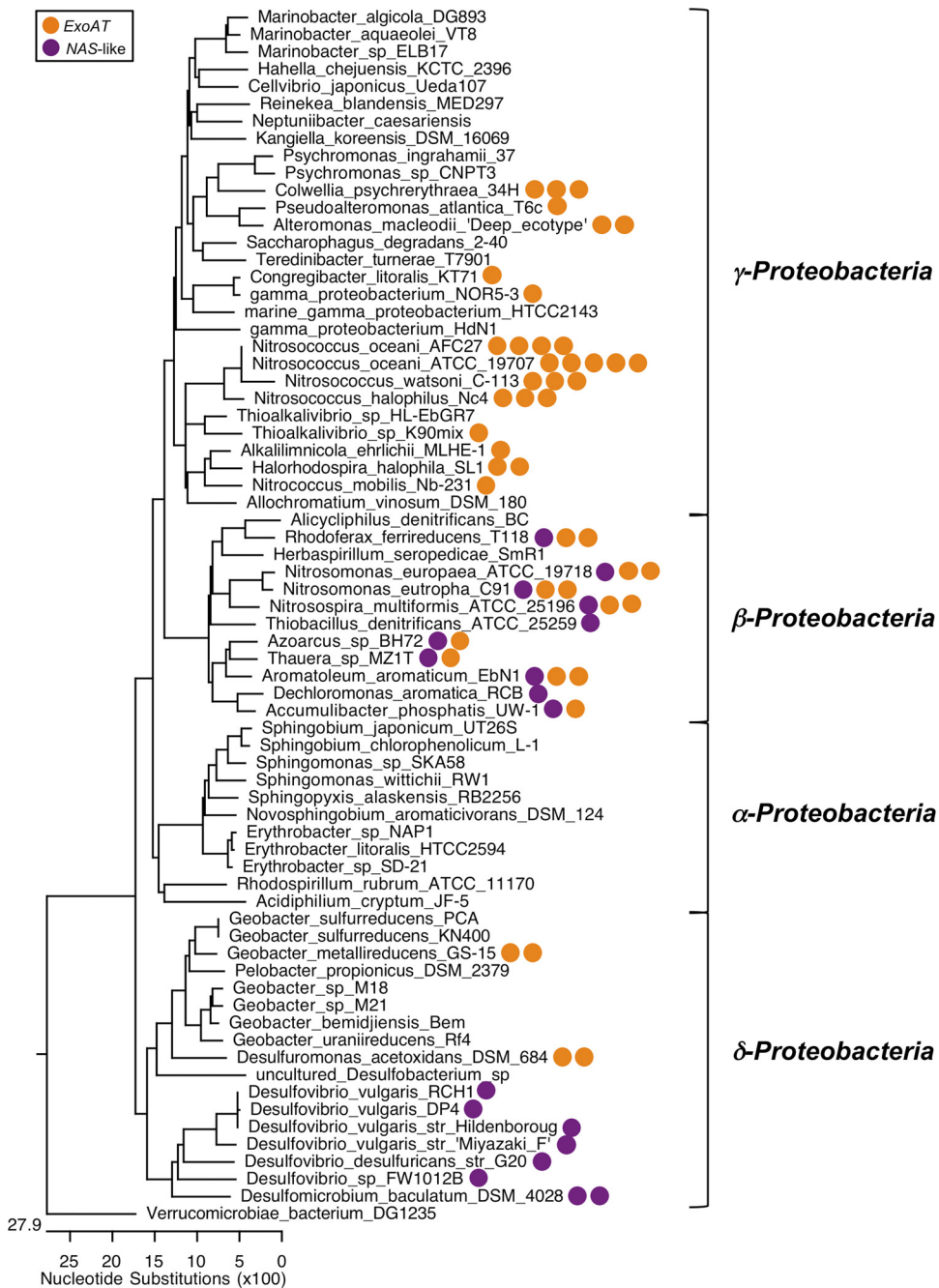


FIG. 4. 16S rRNA phylogram of the 69 species of *Proteobacteria* whose genomes contain PrsK and PrsR genes (out of 71 species total). The phylogram was constructed using 16S sequences from the Silva comprehensive rRNA database SSU Ref 104, trimmed to the region corresponding to positions 133 to 1178 of the *E. coli* K-12 MG1655 16S rRNA gene *rsg*, and aligned using Clustal W. The trimmed 16S rRNA sequence from *Verrucomicrobiae bacterium* DG1235 (a PrsK/PrsR-containing member of the phylum *Verrucomicrobia*) was used as an outgroup. For each ExoAT gene homolog (InterPro IPR022484) found within a genome, an orange dot is appended next to the species name. A purple dot is appended next to the name of each species whose genome contains a putative homolog of the group 1 NasY genes.

ther modify long-chain *N*-acyl amino acids have been recognized for many years (5, 15, 21, 31). Third, *E. coli*-based heterologous expression of related *N*-acylhomoserine lactone synthases and lyso-ornithine lipid synthases (OlsB homologs) results in the accumulation of only the corresponding natively produced metabolites (16, 39). Fourth, the eDNA-derived *N*-acylphenylalanine synthase NasA (NasP2)

produces *N*-acyl amino acids not only when heterologously expressed in *E. coli* but also when expressed in *Pseudomonas putida* KT2440, *Agrobacterium tumefaciens* LBA4404, and *Burkholderia graminis* C4D1M (13). This result suggests that *N*-acyl amino acid biosynthesis is the default function of NasA within representative genetic and metabolic backgrounds spanning the *Alpha*-, *Beta*-, and *Gamma*proteobac-



teria. And, fifth, the binding pocket of the eDNA-derived NAS FeeM is an appropriate size to accommodate its long-chain *N*-acyltyrosine products but appears unlikely to be able to accept large or polymeric amine substrates (37). According to Van Wagoner and Clardy (37), FeeM has  $K_m$  values of  $\sim 10 \mu\text{M}$  and  $\sim 80 \mu\text{M}$  for octanoyl-FeeL (its acyl-ACP substrate) and L-tyrosine, respectively.

Although *N*-acyl amino acid-producing eDNA clones have been identified by virtue of the antibacterial activity bestowed upon them through the production of *N*-acyl amino acids, the relatively high concentrations of *N*-acyl amino acids required to exert this effect have generally evoked caution when the biological purpose of these metabolites is discussed within an environmental context. Chemical signaling, on the other hand, has been one of the prevailing hypotheses regarding the more plausible alternative functions of *N*-acyl amino acids. The association of both the previously characterized NASs and the newly described ExoATs with the ligand-orphaned PrsK/PrsR two-component system is therefore intriguing and begs the question of whether *N*-acyl amino acids could serve as chemical ligands for PrsK homologs. While only about half (46%) of the known PrsK/PrsR-containing genomes contain either an ExoAT or NAS-like gene, additional yet undiscovered families of *N*-acyl amino acid synthases might exist to account for this discrepancy. Alternatively, it is not uncommon for bacterial species to contain genes for small-molecule signaling receptors but not the corresponding biosynthetic genes required for production of cognate small molecules (24). Presumably such bacteria use these receptors for interspecies communication events (10, 23).

To allow for the preliminary evaluation of the interplay between a representative PrsK/PrsR two-component system and a representative ExoAT homolog, we attempted to establish *Pseudoalteromonas atlantica* T6c as a model organism for studying PEP-CTERM/exosortase system function. *P. atlantica* T6c encodes homologs of EpsH, PrsT, PrsK, and PrsR in addition to a single ExoAT homolog and at least 13 PEP-CTERM motif proteins (many of which are preceded by putative PrsR and sigma-54 binding sites) (GenBank accession number CP000388). It is also one of the few PEP-CTERM/exosortase system-containing species with a short doubling time and simple culture conditions. Unfortunately, *N*-acyl amino acids were never detected in the organic extract of *P. atlantica* T6c culture broth, and mutant strains of *P. atlantica* T6c lacking either the *prskR* operon or the ExoAT gene were phenotypically indistinguishable from the wild type (based on growth curve analysis and measurements of exopolysaccharide production) (see supplemental material) (2). Establishment of appropriate culture conditions for inducing *N*-acyl amino acid production in *P. atlantica* T6c or the establishment of alternative PrsK/PrsR-containing *Proteobacteria* as model systems for studying PEP-CTERM/exosortase system function will be required to shed light on the biological role(s) performed by bacterial *N*-acyl amino acids within this diverse group of organisms.

#### ACKNOWLEDGMENTS

This work was supported by NIH grant GM077516 and NIH MSTP grant GM07739 (J.W.C.). S.F.B. is a Howard Hughes Medical Institute Early Career Scientist.

#### REFERENCES

- Audit, B., and C. A. Ouzounis. 2003. From genes to genomes: universal scale-invariant properties of microbial chromosome organisation. *J. Mol. Biol.* **332**:617–633.
- Belas, R., D. Bartlett, and M. Silverman. 1988. Cloning and gene replacement mutagenesis of a *Pseudomonas atlantica* agarase gene. *Appl. Environ. Microbiol.* **54**:30–37.
- Brady, S. F. 2007. Construction of soil environmental DNA cosmid libraries and screening for clones that produce biologically active small molecules. *Nat. Protoc.* **2**:1297–1305.
- Brady, S. F., C. J. Chao, and J. Clardy. 2004. Long-chain *N*-acyltyrosine synthases from environmental DNA. *Appl. Environ. Microbiol.* **70**:6865–6870.
- Brady, S. F., C. J. Chao, and J. Clardy. 2002. New natural product families from an environmental DNA (eDNA) gene cluster. *J. Am. Chem. Soc.* **124**:9968–9969.
- Brady, S. F., and J. Clardy. 2000. Long-chain *N*-acyl amino acid antibiotics isolated from heterologously expressed environmental DNA. *J. Am. Chem. Soc.* **122**:12903–12904.
- Brady, S. F., and J. Clardy. 2005. *N*-acyl derivatives of arginine and tryptophan isolated from environmental DNA expressed in *Escherichia coli*. *Org. Lett.* **7**:3613–3616.
- Brady, S. F., L. Simmons, J. H. Kim, and E. W. Schmidt. 2009. Metagenomic approaches to natural products from free-living and symbiotic organisms. *Nat. Prod. Rep.* **26**:1488–1503.
- Cao, J. G., and E. A. Meighen. 1989. Purification and structural identification of an autoinducer for the luminescence system of *Vibrio harveyi*. *J. Biol. Chem.* **264**:21670–21676.
- Case, R. J., M. Labbate, and S. Kjelleberg. 2008. AHL-driven quorum-sensing circuits: their frequency and function among the *Proteobacteria*. *ISME J.* **2**:345–349.
- Cheung, J., and W. A. Hendrickson. 2010. Sensor domains of two-component regulatory systems. *Curr. Opin. Microbiol.* **13**:116–123.
- Clardy, J., and S. F. Brady. 2007. Cyclic AMP directly activates NasP, an *N*-acyl amino acid antibiotic biosynthetic enzyme cloned from an uncultured beta-proteobacterium. *J. Bacteriol.* **189**:6487–6489.
- Craig, J. W., F.-Y. Chang, J. H. Kim, S. C. Obiajulu, and S. F. Brady. 2010. Expanding small molecule functional metagenomics through parallel screening of broad host-range cosmid environmental DNA libraries in diverse *Proteobacteria*. *Appl. Environ. Microbiol.* **76**:1633–1641.
- Delcher, A. L., K. A. Bratke, E. C. Powers, and S. L. Salzberg. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* **23**:673–679.
- Fukuda, H., S. Iwade, and A. Kimura. 1982. A new enzyme: long acyl aminoacylase from *Pseudomonas diminuta*. *J. Biochem.* **91**:1731–1738.
- Gao, J. L., et al. 2004. Identification of a gene required for the formation of lyso-ornithine lipid, an intermediate in the biosynthesis of ornithine-containing lipids. *Mol. Microbiol.* **53**:1757–1770.
- Geiger, O., N. Gonzalez-Silva, I. M. Lopez-Lara, and C. Sohlenkamp. 2010. Amino acid-containing membrane lipids in bacteria. *Prog. Lipid Res.* **49**:46–60.
- Haft, D. H., I. T. Paulsen, N. Ward, and J. D. Selengut. 2006. Exopolysaccharide-associated protein sorting in environmental organisms: the PEP-CTERM/EpsH system. Application of a novel phylogenetic profiling heuristic. *BMC Biol.* **4**:29.
- Hunter, S., et al. 2009. InterPro: the integrative protein signature database. *Nucleic Acids Res.* **37**:D211–D215.
- Jung, K., T. Odenbach, and M. Timmen. 2007. The quorum-sensing hybrid histidine kinase LuxN of *Vibrio harveyi* contains a periplasmically located N terminus. *J. Bacteriol.* **189**:2945–2948.
- Matsumoto, J., and S. Nagai. 1972. Amidohydrolases for *N*-short and long chain fatty acyl-L-amino acids from mycobacteria. *J. Biochem.* **72**:269–279.
- McDowall, J., and S. Hunter. 2011. InterPro protein classification. *Methods Mol. Biol.* **694**:37–47.
- Michael, B., J. N. Smith, S. Swift, F. Heffron, and B. M. Ahmer. 2001. SdiA of *Salmonella enterica* is a LuxR homolog that detects mixed microbial communities. *J. Bacteriol.* **183**:5733–5742.
- Patankar, A. V., and J. E. Gonzalez. 2009. Orphan LuxR regulators of quorum sensing. *FEMS Microbiol. Rev.* **33**:739–756.
- Pellegrini, M., E. M. Marcotte, M. J. Thompson, D. Eisenberg, and T. O. Yeates. 1999. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc. Natl. Acad. Sci. U. S. A.* **96**:4285–4288.
- Peypoux, F., O. Laprevote, M. Pagaduy, and J. Wallach. 2004. *N*-Acyl derivatives of Asn, new bacterial *N*-acyl D-amino acids with surfactant activity. *Amino Acids* **26**:209–214.
- Quaiser, A., et al. 2003. Acidobacteria form a coherent but highly diverse group within the bacterial domain: evidence from environmental genomics. *Mol. Microbiol.* **50**:563–575.
- Quevillon, E., et al. 2005. InterProScan: protein domains identifier. *Nucleic Acids Res.* **33**:W116–W120.
- Reynolds, S. M., L. Kall, M. E. Riffe, J. A. Bilmes, and W. S. Noble. 2008.



- Transmembrane topology and signal peptide prediction using dynamic Bayesian networks. *PLoS Comput. Biol.* **4**:e1000213.
30. Rogozin, I. B., K. S. Makarova, Y. I. Wolf, and E. V. Koonin. 2004. Computational approaches for the analysis of gene neighbourhoods in prokaryotic genomes. *Brief Bioinform.* **5**:131–149.
  31. Shintani, Y., H. Fukuda, N. Okamoto, K. Murata, and A. Kimura. 1984. Isolation and characterization of N-long chain acyl aminoacylase from *Pseudomonas diminuta*. *J. Biochem.* **96**:637–643.
  32. Smith, D., et al. 2006. Variations on a theme: diverse *N*-acyl homoserine lactone-mediated quorum sensing mechanisms in Gram-negative bacteria. *Sci. Prog.* **89**:167–211.
  33. Spitteller, D., K. Dettner, and W. Bolan. 2000. Gut bacteria may be involved in interactions between plants, herbivores and their predators: microbial biosynthesis of *N*-acylglutamine surfactants as elicitors of plant volatiles. *Biol. Chem.* **381**:755–762.
  34. Swem, L. R., D. L. Swem, N. S. Wingreen, and B. L. Bassler. 2008. Deducing receptor signaling parameters from in vivo analysis: LuxN/AI-1 quorum sensing in *Vibrio harveyi*. *Cell* **134**:461–473.
  35. Tatusov, R. L., et al. 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4**:41.
  36. Tatusov, R. L., E. V. Koonin, and D. J. Lipman. 1997. A genomic perspective on protein families. *Science* **278**:631–637.
  37. Van Wagoner, R. M., and J. Clardy. 2006. FeeM, an *N*-acyl amino acid synthase from an uncultured soil microbe: structure, mechanism, and acyl carrier protein binding. *Structure* **14**:1425–1435.
  38. Ward, N. L., et al. 2009. Three genomes from the phylum *Acidobacteria* provide insight into the lifestyles of these microorganisms in soils. *Appl. Environ. Microbiol.* **75**:2046–2056.
  39. Winson, M. K., et al. 1995. Multiple *N*-acyl-L-homoserine lactone signal molecules regulate production of virulence determinants and secondary metabolites in *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci. U. S. A.* **92**:9427–9431.
  40. Wolf, Y. I., I. B. Rogozin, A. S. Kondrashov, and E. V. Koonin. 2001. Genome alignment, evolution of prokaryotic genome organization, and prediction of gene function using genomic context. *Genome Res.* **11**:356–372.
  41. Yagi, H., G. Corzo, and T. Nakahara. 1997. *N*-acyl amino acid biosynthesis in marine bacterium, *Deleya marina*. *Biochim. Biophys. Acta* **1336**:28–32.
  42. Zheng, Y., B. P. Anton, R. J. Roberts, and S. Kasif. 2005. Phylogenetic detection of conserved gene clusters in microbial genomes. *BMC Bioinformatics* **6**:243.