

Expanding Small-Molecule Functional Metagenomics through Parallel Screening of Broad-Host-Range Cosmid Environmental DNA Libraries in Diverse *Proteobacteria*^{∇†}

Jeffrey W. Craig, Fang-Yuan Chang, Jeffrey H. Kim, Steven C. Obiajulu, and Sean F. Brady*

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

Received 8 September 2009/Accepted 1 January 2010

The small-molecule biosynthetic diversity encoded within the genomes of uncultured bacteria is an attractive target for the discovery of natural products using functional metagenomics. Phenotypes commonly associated with the production of small molecules, such as antibiosis, altered pigmentation, or altered colony morphology, are easily identified from screens of arrayed metagenomic library clones. However, functional metagenomic screening methods are limited by their intrinsic dependence on a heterologous expression host. Toward the goal of increasing the small-molecule biosynthetic diversity found in functional metagenomic studies, we report the phenotypic screening of broad-host-range environmental DNA libraries in six different proteobacteria: *Agrobacterium tumefaciens*, *Burkholderia graminis*, *Caulobacter vibrioides*, *Escherichia coli*, *Pseudomonas putida*, and *Ralstonia metallidurans*. Clone-specific small molecules found in culture broth extracts from pigmented and antibacterially active clones, as well as the genetic elements responsible for the biosynthesis of these metabolites, are described. The host strains used in this investigation provided access to unique sets of clones showing minimal overlap, thus demonstrating the potential advantage conferred on functional metagenomics through the use of multiple diverse host species.

Uncultured bacteria are predicted to be a significant reservoir of novel small-molecule biosynthetic machinery (19, 34). One means by which to access the biosynthetic potential contained within the genomes of uncultured bacteria is functional metagenomics (19). This approach involves cloning DNA directly from naturally occurring microbial populations (environmental DNA [eDNA]) and screening the resulting clone libraries for phenotypes traditionally associated with the production of secondary metabolites. A major limitation of functional metagenomics is its reliance on a foreign host to facilitate the expression of eDNA-derived genes and gene clusters (17). Codon bias, missing substrates, and the inability to recognize foreign regulatory elements, including promoters and ribosomal binding sites, are just some of the obstacles that are likely to limit the success of expression-dependent studies with any single host organism. Circumventing these obstacles through an expansion of the collection of hosts available for functional metagenomic studies should increase the efficacy of this approach.

Soil ecosystems are rich in bacterial diversity, and the majority of soil-dwelling bacteria remain recalcitrant to standard microbial culture methods (33, 39). Large-scale metagenomic sequencing studies indicate that soil microbiomes are often dominated by five bacterial phyla: *Alphaproteobacteria*, *Betaproteobacteria*, *Gammaproteobacteria*, *Acidobacteria*, and *Acti-*

nobacteria (14). Bacteria from these common phyla are therefore appealing hosts for use in functional metagenomic studies of soil-derived eDNA libraries.

In this study, six unique bacterial hosts [*Agrobacterium tumefaciens* (*Alphaproteobacteria*), *Burkholderia graminis* (*Betaproteobacteria*), *Caulobacter vibrioides* (*Alphaproteobacteria*), *Escherichia coli* (*Gammaproteobacteria*), *Pseudomonas putida* (*Gammaproteobacteria*), and *Ralstonia metallidurans* (*Betaproteobacteria*)] from the superphylum *Proteobacteria* were explored as hosts for functional metagenomic screening (Table 1). Soil eDNA libraries constructed in an IncP1- α group broad-host-range cosmid vector were independently introduced into each of the six host proteobacteria, and the resulting eDNA libraries were screened for colonies exhibiting antibacterial activity, altered pigmentation, and altered colony morphology. Each of these phenotypes is easily detectable and frequently associated with small-molecule production (5). Individual clones possessing one or more of these natural-product-associated phenotypes were recovered from high-throughput primary screens and subsequently analyzed for the production of clone-specific metabolites. Clone-specific small molecules responsible for the observed phenotypes and the genetic elements that encode their biosyntheses are described here.

MATERIALS AND METHODS

eDNA isolation and library construction. Two soil samples collected in Pennsylvania (one from a deciduous forest and the other from the mud at the bottom of a small creek bed) and one collected in Oregon (from a cold desert covered with sand and clay) were used as sources of eDNA. Soils were collected manually using clean gardening spades and stored at 4°C for less than 2 weeks before processing. Large debris (rocks, sticks, and roots) was removed prior to the addition of soil lysis buffer by passing each soil sample sequentially through 3.35- and 1.00-mm sieves. Following the general detergent-based lysis strategy (sodium dodecyl sulfate and cetyltrimethylammonium bromide) of Zhou et al., a one-to-

* 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://aem.asm.org/>.

[∇] Published ahead of print on 15 January 2010.

TABLE 1. Strains, clones, and plasmids used in this study

Strain, clone, or plasmid	Description	Source or reference(s)
Strains		
<i>E. coli</i> EC100	<i>Gammaproteobacteria</i> host	Epicentre
<i>E. coli</i> EC300	Copy control <i>E. coli</i> strain	Epicentre
<i>E. coli</i> S17.1	Strain used for conjugal mating	36
<i>P. putida</i> KT2440	<i>Gammaproteobacteria</i> host	3
<i>R. metallidurans</i> CH34	<i>Betaproteobacteria</i> host	29
<i>B. graminis</i> C4D1M	<i>Betaproteobacteria</i> host	40
<i>C. vibrioides</i> CB15	<i>Alphaproteobacteria</i> host	32
<i>A. tumefaciens</i> LBA4404	<i>Alphaproteobacteria</i> host	22
eDNA clones		
EC5	Antibacterially active, hosted by <i>E. coli</i> , produces <i>N</i> -acyl amino acids	This paper
EC6	Antibacterially active (nonreproducible), hosted by <i>E. coli</i>	This paper
RM19	Heme (protoporphyrin IX) producer, hosted by <i>R. metallidurans</i>	This paper
RM35	Antibacterially active, hosted by <i>R. metallidurans</i> , contains a cryptic NRPS gene cluster	This paper
RM44	Antibacterially active, hosted by <i>R. metallidurans</i> , produces compounds 1 and 2	This paper
RM57	Antibacterially active, hosted by <i>R. metallidurans</i> , produces type III polyketides	11
RM3/AT1	Carotenoid producer, hosted by either <i>A. tumefaciens</i> (AT1) or <i>R. metallidurans</i> (RM3)	This paper, 11
AT3	Wrinkled colony phenotype, hosted by <i>A. tumefaciens</i>	This paper
AT5	Wrinkled colony phenotype, hosted by <i>A. tumefaciens</i>	This paper
Plasmids		
pTAC-MAT-Tag-2	<i>tac</i> promoter-based expression vector	Sigma-Aldrich
pTAC-NasA	pTAC-MAT-Tag-2 with <i>nasA</i> , S2K, and I3L	This paper
pJWC1	IncP1- α group broad-host-range cosmid	11

one mixture of soil and soil lysis buffer was heated briefly at 70°C, after which any remaining particulate matter was removed by centrifugation (44). Crude eDNA was precipitated from the resulting supernatant using isopropanol and collected by centrifugation. All crude eDNA samples were then washed with 70% ethanol and resuspended in TE buffer (10 mM Tris-HCl, 1 mM Na EDTA, pH 8.0). The remaining soil particulate matter and humic substances were removed by large-scale gel purification on a 1% agarose gel (16 h at 20 V). Purified high-molecular-weight eDNA was recovered from the gel by electroelution (2 h at 100 V) and concentrated by isopropanol precipitation. eDNA was then blunt ended (End-It; Epicentre), ligated into ScaI-digested and calf intestinal alkaline phosphatase (CIP)-treated pJWC1 (11), packaged into lambda phage *in vitro* (MaxPlax Packaging Extracts; Epicentre), and transfected into *E. coli* EC100 cells. Each 100,000- to 500,000-member eDNA library was prepared as a collection of unique 20,000- to 50,000-member sublibraries. The individual sublibraries were miniprep, and the resulting DNA was used for either direct transformation of heterologous expression hosts or transformation of *E. coli* conjugal mating strain S17.1.

Preparation of electrocompetent cells. Electrocompetent cells were prepared using established protocols with only minor deviations. Briefly, all strains were grown at 30°C in their preferred media (*R. metallidurans* CH34, SOB medium [38]; *C. vibrioides* CB15, *Caulobacter* medium [CM; ATCC medium 36] [18]; *B. graminis* C4D1M, Luria-Bertani medium [LB] [13]; *A. tumefaciens* LBA4404, yeast extract-mannitol medium [YM] [27]) to optical densities at 600 nm (OD_{600}) between 0.5 and 1.0, chilled on ice, harvested by centrifugation, washed three times with ice-cold 10% glycerol, resuspended in 10% glycerol, and flash frozen.

Transformation of eDNA cosmid libraries by electroporation and conjugation. Individual 80- μ l aliquots of electrocompetent cells were thawed on ice and mixed with 10 μ l of cosmid DNA (250 ng ml⁻¹), transferred to 1.0-mm electroporation cuvettes, pulsed at 1.8 kV (2.20 kV for *A. tumefaciens*) for 6 ms using a Bio-Rad MicroPulser, and then immediately mixed with 1 ml of recovery medium (SOC for *B. graminis* and *R. metallidurans*, YM for *A. tumefaciens*, and CM for *C. vibrioides*). The resulting cell suspensions were incubated with shaking at 30°C for 2 to 3 h and subsequently plated at titers of 1,500 to 2,500 colonies per 150-mm-diameter plate onto the following selection plates: *B. graminis* and *R. metallidurans*, LB-tetracycline (20 μ g ml⁻¹); *A. tumefaciens*, LB-tetracycline (5 μ g ml⁻¹); *C. vibrioides*, CM-tetracycline (4 μ g ml⁻¹). Biparental conjugation was used to transfer eDNA libraries from *E. coli* S17.1 to *P. putida* as previously described (28). Tetracycline (50 μ g ml⁻¹) and irgasan (20 μ g ml⁻¹) were used to select for exconjugants and counterselect against the *E. coli* donor, respectively.

After 1 to 2 days, exconjugants were scraped from the selection plates, resuspended in 10% glycerol, and flash frozen in liquid nitrogen. Titrations of dilutions of frozen *P. putida* stocks containing the eDNA libraries were then determined, and the bacteria were plated at 500 colonies per plate onto 150-mm-diameter plates of LB-tetracycline (50 μ g ml⁻¹).

Phenotypic screening of eDNA libraries. eDNA libraries transferred to individual host *Proteobacteria* were plated to achieve between 1.5- and 3-fold coverage of each library. After recovery of pigmented colonies and colonies displaying altered morphologies, the screening plates were overlaid with a thin layer (15 ml plate⁻¹) of LB top agar (LB with 6 g liter⁻¹ agar) containing a 1:200 dilution of *B. subtilis* 1E9 (grown in LB-tetracycline to an OD_{600} of 1.0). Overlaid plates were allowed to solidify at room temperature before being transferred to 30°C. After 24 to 48 h at 30°C, the overlaid plates were screened visually for colonies that produced a zone of growth inhibition in the resulting *B. subtilis* lawn. Colonies displaying this phenotype were picked with sterile toothpicks, streaked onto selective media, and retested to confirm the original phenotype. Cosmid DNA from each hit, regardless of the host origin, was miniprep using standard Qiagen protocols and then electroporated into CopyControl EC300 *E. coli* (Epicentre) to facilitate genetic analysis. Each cosmid was sequenced by 454 pyrosequencing, and the sequences were analyzed for putative open reading frames (ORFs) using GLIMMER3.02 (12) and SoftBerry FGENESB: Bacterial Operon and Gene Prediction Program (<http://linux1.softberry.com/berry.phtml?topic=fgenesb&group=programs&subgroup=gfindb>). Digital images were captured with a Nikon Coolpix S550 digital camera and processed minimally (image cropping, brightness, and contrast adjustment) with Microsoft Office Picture Manager.

Organic extraction, compound isolation, and structural analysis. One-liter liquid cultures (LB containing tetracycline at 20 μ g ml⁻¹ and kanamycin at 12 μ g ml⁻¹) of *R. metallidurans* CH34 clone RM44 were grown at 30°C with shaking (200 rpm) for 3 to 4 days. Liquid cultures were then extracted with ethyl acetate, and the resulting extracts were dried under vacuum. Crude extracts were resuspended in a minimal volume of high-performance liquid chromatography (HPLC) grade methanol and then subjected to preparative HPLC under the following conditions (Waters XBridge C₁₈ column [10 by 150 mm], 7 ml/min): 5 min at 50:50 H₂O-methanol with 0.1% formic acid, followed by a linear gradient of 50:50 H₂O-methanol with 0.1% formic acid to 100% methanol with 0.1% formic acid over 15 min, followed by 100% methanol with 0.1% formic acid for 4 min. Pooled fractions containing compounds 1 and 2 (min 11.5 to 13.5) were resuspended in 50:50 methanol-acetonitrile and subjected to a second round of preparative HPLC as follows: 2 min at 25:75 acetonitrile-H₂O with 0.1% formic

acid, followed by a linear gradient of 25:75 acetonitrile-H₂O with 0.1% formic acid to 100% acetonitrile over 20 min, followed by 100% acetonitrile for 2 min. Fractions containing pure compound 1 (min 8.5 to 10.0) were pooled and dried under vacuum, resulting in a final yield of ~2.5 mg liter⁻¹. Fractions containing pure compound 2 (min 8.5 to 10.0) were pooled and dried under vacuum, resulting in a final yield of ~0.25 mg liter⁻¹.

Compound 1, 3-hydroxy myristoylputrescine [IUPAC: *N*-(4-amino-2-hydroxybutyl)tetradecanamide]. White powder; ¹H NMR (600 MHz, CD₃OD) 3.79 (C-3, 1H, m), 3.29 (C-4, 1H, dd, 13, 5.5), 3.18 (C-4, 1H, dd, 13, 7.4), 3.07 (C-1, 2H, m), 2.23 (C-6, 2H, t, 7.6) 1.82 (C-2, 1H, m), 1.68-1.62 (C-2, C-7, 3H, m), 1.4-1.2 (m), 0.92 (C-18, 3H, t, 7.0); ¹³C NMR (150 MHz, CD₃OD) 176.9 (C-5), 69.9 (C-3), 46.3 (C-4), 38.5 (C-1), 37.0 (C-6), 33.1, 32.7 (C-2), 30.8 (m), 30.6, 30.5 (m), 30.4, 27.0 (C-7), 23.7, 14.4 (C-18); high-resolution electrospray ionization mass spectrometry (HRESIMS) *m/z* 315.2998 [M]⁺ (calculated for C₁₈H₃₉N₂O₂, 315.3006).

Compound 2, myristoylputrescine. White powder; ¹H NMR (600 MHz, CD₃OD) 3.20 (C-4, 2H, t, 7), 2.94 (C-1, 2H, t, 7.5), 2.18 (C-6, 2H, t, 8), 1.65 (C-2, 2H, m), 1.64-1.56 (C-3, C-7, 4H, m), 1.2-1.4 (m), 0.90 (C-18, 3H, t, 7); ¹³C NMR (150 MHz, CD₃OD) 176.5 (C-5), 40.3 (C-1), 39.4 (C-4), 37.2 (C-6), 33.1, 30.8 (m), 30.6, 30.5 (m), 30.4, 27.5 (C-3), 27.1 (C-7), 26.0 (C-2), 23.7, 14.4 (C-18); HRESIMS *m/z* 299.3039 [M]⁺ (calculated for C₁₈H₃₉N₂O, 299.3057).

Cloning and expression of the EC5 *N*-acyl amino acid synthase (NAS) gene in *E. coli*. The 657-bp EC5 NAS gene was PCR amplified (30 cycles of 97°C for 30 s, 65°C for 30 s, and 72°C for 15 s) from EC5 using primers EC5-NAS-HindIII-F (5'-ACCATGAAGCTTCTGATTCGCCCGTGTATCGGGACAG-3') and EC5-NAS-BglIII-R (5'-GGAGATCTTTCATCGGAGCGCCTCCGATTCATGATC-3'). This primer set introduces the following amino acid changes into NAS: S2K and I3L. The PCR amplicon was double digested with HindIII and BglIII and then ligated into the correspondingly double-digested/CIP-treated pTAC-MAT-Tag-2 expression vector (E5405; Sigma-Aldrich). LB cultures (0.1 mM isopropyl-β-D-thiogalactopyranoside [IPTG], 100 μg ml⁻¹ ampicillin) of *E. coli* EC100 containing the NAS expression construct were incubated at 30°C with shaking for 3 days and then extracted with an equal volume of ethyl acetate. The dried extracts were analyzed by normal-phase thin-layer chromatography (TLC) and reversed-phase analytical HPLC for the presence of clone-specific metabolites. Analytical reversed-phase HPLC conditions were as follows: Waters XBridge C₁₈ column (4.6 by 150 mm), 1.5 ml/min, 3 min at 80:20 H₂O-methanol with 0.1% formic acid, followed by a linear gradient of 80:20 H₂O-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.

Nucleotide sequence accession numbers. The GenBank accession numbers corresponding to individual cosmid eDNA clones are as follows: EC5, GQ869383; EC6, GQ869384; RM3/AT1, FJ151553 (11); RM57, FJ151552 (11); RM44, GQ869386; RM35, GQ869384; AT3, GQ869381; and AT5, GQ869382.

RESULTS

Broad-host-range library construction and screening. We previously reported the construction of an IncP1-α group broad-host-range cosmid cloning vector, pJWC1, designed to facilitate functional metagenomic screening efforts (11). In this study, pJWC1 was used to create cosmid-based eDNA libraries from soil samples collected in Reading, PA (RPA, deciduous forest topsoil, 170,000 clones), Wellsboro, PA (JPA, creek bed mud/sediment, 450,000 clones), and Smith Rock, OR (SROR, sand/clay-covered desert soil, 130,000 clones). The soils chosen for this study are representative of three different surface terrains, all of which were free from obvious pollution, contamination, or other signs of human interference. Each library was initially constructed in *E. coli* and then transferred to *A. tumefaciens*, *B. graminis*, *C. vibrioides*, and *R. metallidurans* by electroporation and to *P. putida* by biparental conjugation. In each case, the retransformation of randomly chosen library clones back into *E. coli* confirmed that eDNA cosmids were stably maintained throughout the transfer and subsequent screening process despite the presence of functional recombination machineries in all hosts other than *E. coli* (data not shown).

eDNA libraries maintained in each host proteobacterium

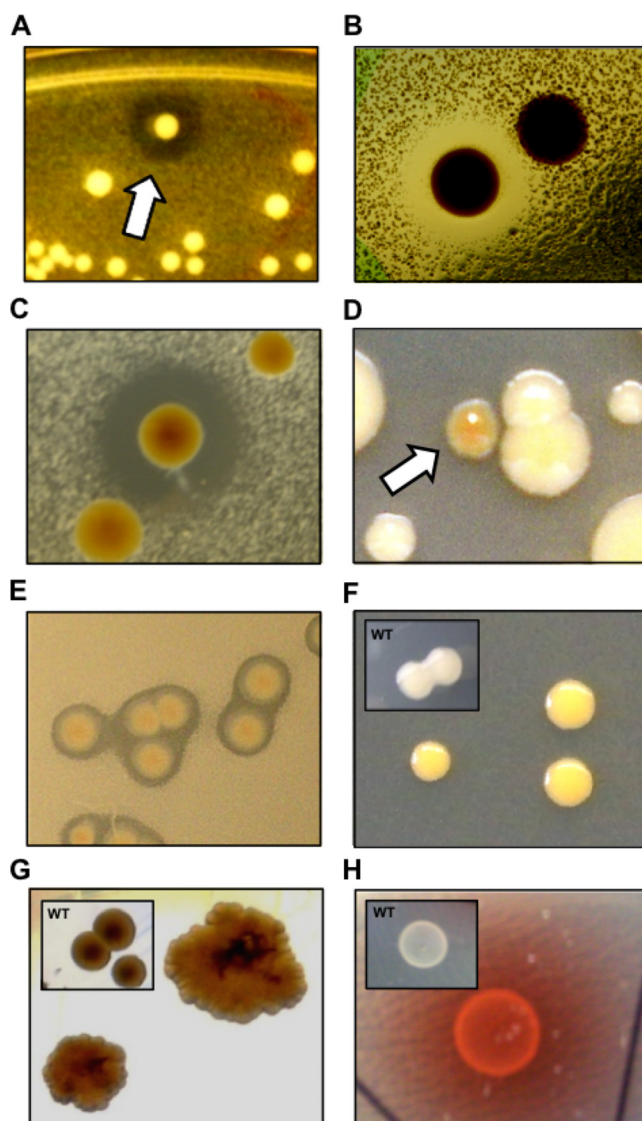


FIG. 1. eDNA library clones displaying phenotypes of interest. Representative colonies of interest as they appeared on the original assay plates: A, antibacterially active colonies found in an *E. coli*-based library; B and C, antibacterially active colonies found in *R. metallidurans*-based libraries; D, pigmented colonies found in a *P. putida*-based library. The arrows in panels A and D indicate specific clones of interest. Representative clones of interest after being recovered from primary assay plates: E, antibacterially active colonies recovered from an *E. coli*-based library; F, pigmented colonies recovered from an *A. tumefaciens*-based library; G, wrinkled colonies recovered from an *A. tumefaciens*-based library; H, pigmented colony recovered from an *R. metallidurans*-based library. The insets in panels F to H show wild-type (WT) colonies.

were distributed onto 150-mm-diameter plates at titers of 500 to 2,500 colonies per plate. These arrayed eDNA libraries were allowed to mature for 4 to 6 days and then screened for colonies that exhibited antibacterial activity, altered pigmentation, or altered colony morphology (Fig. 1; Table 2). Antibacterial screens were carried out using the two-layer overlay method with tetracycline-resistant *B. subtilis* (BGSCID 1E9) as the assay strain. This approach provides a very simple method of

TABLE 2. eDNA clones identified in phenotypic screens of soil eDNA libraries

Host strain	Subphylum	No. of colonies displaying phenotype				
		Pigmentation			Antibiosis	Morphology
		Heme	Melanin	Other		
<i>E. coli</i> EC100	<i>Gammaproteobacteria</i>				2	
<i>P. putida</i> KT2440	<i>Gammaproteobacteria</i>	>10 ^a				
<i>R. metallidurans</i> CH34	<i>Betaproteobacteria</i>	>10 ^a	5	3	4	
<i>B. graminis</i> C4D1M	<i>Betaproteobacteria</i>	6			2	
<i>A. tumefaciens</i> LBA4404	<i>Alphaproteobacteria</i>			1		2
<i>C. vibrioides</i> CB15	<i>Alphaproteobacteria</i>					

^a Only a subset of brown-colored colonies was recovered from assay plates.

identifying diffusible small-molecule antibiotics. Alternative screening approaches would be required to identify compounds that remain sequestered within the host cell. Prior to the antibiosis overlay assay, each arrayed library was first examined for clones exhibiting altered pigmentation or altered colony morphology.

Gammaproteobacteria. Standard cloning strains of the gammaproteobacterium *E. coli* have been the hosts of choice for most functional metagenomic screening efforts. Therefore, *E. coli* was included in this study to establish a base for comparing alternative bacterial hosts. No *E. coli*-hosted eDNA clones were found to display altered pigmentation or altered colony morphology during screens of any of the three libraries. Two antibacterially active eDNA clones, EC5 from JPA and EC6 from RPA, were, however, found during overlay assays performed on the three libraries. Upon retransformation, clone EC5 continued to confer antibacterial activity on the *E. coli* host, while EC6 did not reproduce the originally observed bioactivity (Fig. 1E and A, respectively). Bioinformatic analysis of the fully sequenced eDNA insert from EC6 did not reveal any putative ORFs with homology to known biosynthetic enzymes, and therefore this clone was not investigated further (GenBank accession no. GQ869384).

HPLC-mass spectrometry (HPLC-MS) analysis of ethyl acetate extracts derived from LB-grown cultures of antibacterially active clone EC5 suggested the presence of a clone-specific collection of *N*-acyl aromatic amino acids. Clones that produce *N*-acyl amino acids are frequently encountered in antibacterial screens of *E. coli*-hosted eDNA libraries (6). Individual *N*-acyl amino acids produced by EC5 were purified from ethyl acetate extracts by a combination of normal-phase flash chromatography and reversed-phase HPLC. HPLC-MS analysis coupled with ¹H NMR studies of the two most abundant chemical species isolated from these ethyl acetate extracts indicated that the major clone-specific compounds produced by EC5 are spectroscopically and chromatographically identical to the 12-carbon *N*-acyl derivative of phenylalanine (observed *m/z* = 348.2) and to the 14-carbon *N*-acyl derivative of tryptophan (observed *m/z* = 415.3) (see Fig. S1 in the supplemental material) (7, 9).

The fully sequenced EC5 eDNA insert (GenBank accession no. GQ 869383) was found to contain a putative ORF showing low-level (<20%) sequence identity to known NASs. To test its role as a predicted NAS, this ORF was PCR amplified and then cloned into *E. coli* expression plasmid pTAC-MAT-Tag-2 (E5405; Sigma-Aldrich). HPLC-MS analysis of the ethyl ace-

tate extract derived from an *E. coli* culture transformed with this expression construct confirmed that it contained a mixture of long-chain *N*-acyl aromatic amino acids indistinguishable from that found in the extract derived from the original EC5 clone, confirming the NAS designation given to this gene.

The NAS (*nasA*) from EC5 is predicted to reside in a two-gene operon (*nasAB*). The second ORF in this operon (*nasB*) contains a set of domains identical to those present in the E1 α , E1 β , and E2 subunits of 2-oxo acid dehydrogenase complexes, as well as a C-terminal β -keto-acyl synthase III domain (KAS III), which is predicted to be involved in the synthesis of acyl carrier protein (ACP)-linked fatty acids (15, 30). In a BLAST search, the only other putative enzyme found to contain the same predicted multidomain architecture is encoded within the genome of *Agrobacterium radiobacter* K84 (GenBank accession no. ACM27363) (37). The putative *A. radiobacter* KAS III gene is also part of a predicted two-gene operon showing the same organization as the EC5 *nasAB* operon. In the operon from *A. radiobacter*, a putative fatty acid hydroxylase/sterol desaturase domain-containing enzyme (GenBank accession no. ACM27362) sits in place of the NAS gene found in the eDNA-derived operon.

The predicted domain functions of eDNA-derived multidomain KAS III (NasB) suggest that it might link the catabolism of 2-oxo acids to the production of ACP-linked fatty acids, which are known substrates for NASs. The major end products of 2-oxo acid dehydrogenase complexes are acyl coenzymes A (acyl-CoAs) derived from the decarboxylation of a variety of primary metabolic substrates. Acyl-CoAs generated by the early domains in NasB could therefore function as substrates for the synthesis of ACP-linked fatty acids by the terminal type III β -keto-ACP synthase domain (Fig. 2). The successful production of *N*-acyl amino acids in our *E. coli*-based heterologous expression system with the NAS alone indicates that if this multidomain KAS III is a source of ACP-linked fatty acids, it is not the exclusive source of ACP-linked fatty acids used by the EC5 NAS.

The soil-dwelling saprophyte *P. putida* was selected as a second representative host from the *Gammaproteobacteria* due to its frequent use as an expression host for natural product biosynthetic pathways (43). The large colonies that are produced by *P. putida* are too easily dispersed by top agar overlays to permit screening of *P. putida*-based libraries for antibacterial activity using the two-layer overlay method. Surmounting this screening problem by generating mutant *P. putida* strains that are more compatible with high-throughput overlays could

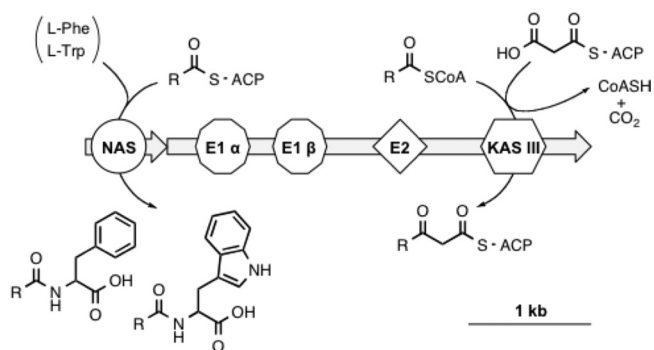


FIG. 2. The EC5 putative NAS operon (*nasAB*). The multidomain ORF located immediately downstream of the EC5 NAS is predicted to produce ACP-linked fatty acids that could serve as one source of acyl-ACPs used in the biosynthesis of *N*-acyl aromatic amino acids.

potentially transform *P. putida* into a productive host for antibiosis detection. While we were not able to screen for antibacterially active *P. putida*-based clones, 11 pigmented clones with colors ranging from deep pink to brown were identified in visual screens of the three eDNA libraries hosted by *P. putida* (Fig. 1D). As seen with the brown-pigmented *R. metallidurans* clones described later in more detail, HPLC-MS analysis of crude organic extracts derived from liquid cultures of these colored clones indicated that all of the pigmented clones overproduced endogenous porphyrin ring-containing metabolites, and thus none were pursued further.

Betaproteobacteria. In previous work, we identified carotenoid- and polyketide-producing clones from phenotypic screens of the JPA and SROR libraries hosted by the betaproteobacterium *R. metallidurans* and showed that these same clones could not be identified using an *E. coli* host (GenBank accession no. FJ151553 and FJ151552) (11). During the present investigation, screening of the RPA library in *R. metallidurans* led to the identification of four reddish brown clones, six dark brown clones, and two clones displaying antibacterial activity (RM35 and RM44).

Reddish brown melanin-producing clones and dark brown heme-producing clones have been previously reported from functional metagenomic screens of *E. coli*-hosted libraries (7, 23–25). The reddish brown and dark brown clones identified in screens of *R. metallidurans*-hosted libraries were therefore easily dereplicated as heme and melanin producers, respectively. Mass and UV absorption spectra observed in HPLC-MS analysis of acidified ethyl acetate extracts from the six dark brown clones showed that each overproduced the heme-related pigment protoporphyrin IX (observed $m/z = 563.3$) (21). One representative member of this family of clones, RM19, was transposon mutagenized and reintroduced into *R. metallidurans* for phenotypic screening. As reported from studies of brown clones found in *E. coli*-hosted libraries, all of the transposon insertions that knocked out color production were found in a predicted glutamyl-tRNA reductase (HemA), the enzyme that catalyzes the first committed step in heme biosynthesis. Acidified ethyl acetate extracts derived from cultures of the reddish brown clones were all found to contain the same non-colored clone-specific metabolite (observed $m/z = 169.1$). This compound was purified from the ethyl acetate extract of a

single member of this family of clones by normal-phase flash chromatography. The ^1H NMR spectrum of the purified compound was found to be identical to that of homogentisic acid (commercial standard, Sigma-Aldrich 53560). The accumulation of this common microbial metabolite, produced by the enzyme hydroxyphenylpyruvate dioxygenase, is known to result in reddish brown pyomelanin pigments (Fig. 1H) (31).

Two clone-specific metabolites were isolated from ethyl acetate extracts derived from LB cultures of antibacterially active clone RM44 (Fig. 1C). The chemical structure of each compound was determined by one- and two-dimensional NMR experiments and HRESIMS. HRESIMS indicated that the molecular formulas of compounds 1 and 2 are $\text{C}_{18}\text{H}_{38}\text{N}_2\text{O}_2$ and $\text{C}_{18}\text{H}_{38}\text{N}_2\text{O}$, respectively. Compound 1, which is only predicted to differ from compound 2 by the presence of an additional oxygen atom, is present in the crude extract in 10-fold greater abundance than compound 2.

^1H and ^1H - ^1H correlation spectroscopy (COSY) NMR spectra of compound 1 suggested the presence of a trisubstituted four-carbon spin system and a long-chain fatty acid substructure (see Fig. S2 in the supplemental material). Based on carbon chemical shift data, the trisubstituted four-carbon spin system is predicted to be functionalized at both C-1 (H1 3.07 ppm) and C-4 (H4A 3.29 ppm, H4B 3.18 ppm) with nitrogens and at C-3 (H3 3.79 ppm) with an oxygen atom. Long-range ^1H - ^{13}C heteronuclear multiple bond coherence (HMBC) correlations from the C-4 protons of the 4-amino-2-hydroxybutamine substructure to the carbonyl carbon of the fatty acid substructure established the structure of compound 1 as an *N*-acylated derivative of 4-amino-2-hydroxybutamine. Based on the molecular formula for compound 1, the acyl side chain must be a fully saturated 14-carbon fatty acid. The final structure of compound 1 is therefore *N*-(4-amino-2-hydroxybutyl)tetradecanamide (Fig. 3A).

The 1D ^1H and ^1H - ^1H COSY spectra of compound 2 indicated the presence of four methylene groups organized as a linear four-carbon spin system, as well as a long-chain fatty acid substructure (see Fig. S3 in the supplemental material). Chemical shift analysis suggested that both ends of the four-carbon spin system are functionalized with nitrogen atoms (H-1, 2.94 ppm; H-4, 3.20 ppm). As seen in the ^1H - ^{13}C HMBC spectrum of compound 1, there is a correlation from protons on the terminal carbon (C-4) of the four-carbon spin system to the carbonyl carbon of the fatty acid substructure. Based on the molecular formula for compound 2, the acyl side chain must be a fully saturated 14-carbon fatty acid. The final structure of compound 2 is therefore *N*-(4-aminobutyl)tetradecanamide or myristoylputrescine. To the best of our knowledge, compound 1 is a new natural product that has not been previously isolated from cultured bacteria. Compound 2 is also a new metabolite differing in acyl chain length from the known bacterial natural product palmitoylputrescine (8).

Saturating transposon mutagenesis of the RM44 cosmid indicated that a single ORF was responsible for both the observed antibiosis and the production of compounds 1 and 2 (Fig. 3B). This ORF is most similar to the biosynthetic enzyme responsible for the production of palmitoylputrescine (GenBank accession no. AAV33349) and the N terminus of a putative serine-pyruvate aminotransferase from the genome of *Bacillus* sp. strain NRRL B-14911 (GenBank accession no. EAR68203). The entire RM44

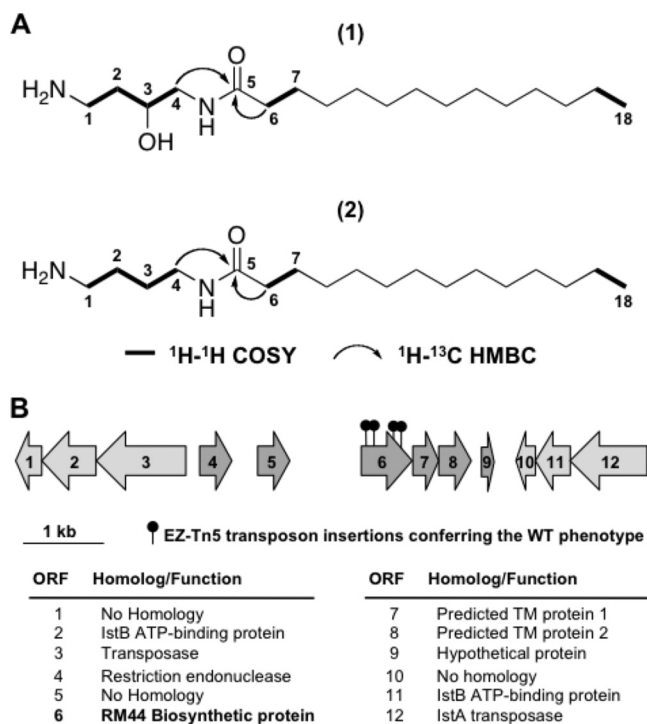


FIG. 3. (A) Key ^1H - ^1H COSY and ^1H - ^{13}C HMBC correlations used to establish the structures of compounds 1 and 2. (B) Transposons that knock out the production of compounds 1 and 2 all fall into a single ORF (ORF6). This biosynthetic ORF is located within a genetic island flanked on both sides by IstA/IstB-like, IS21 family transposable sequences (ORFs 2 and 3 and ORFs 11 and 12).

eDNA insert was subsequently sequenced (GenBank accession no. GQ869386), and the RM44 biosynthetic ORF was found to be located within an 8-kb region flanked on both sides by genes homologous to the tandem IstA/IstB transposase elements contained within IS21 family transposable sequences. In addition to the RM44 biosynthetic ORF, this region contains genes for a putative restriction endonuclease and two short hypothetical proteins predicted to contain multiple transmembrane helices (10). In combination, the latter two ORFs, which sit directly adjacent to the RM44 biosynthetic ORF, are predicted to contain a total of seven transmembrane helices. The predicted topology of these transmembrane helices is similar to that of the transmembrane regions of canonical receptors of the seven-transmembrane diverse intracellular signaling module family (*7TMR-DISM_7TM*, PF07695) (2). Consistent with this observation, each of these putative ORFs shows the greatest sequence similarity to the N-terminal transmembrane regions of putative multisensor signal transduction histidine kinases (GenBank accession no. ACL15452 and ACA59209, respectively).

RM35, the second *R. metallidurans*-based antibacterially active hit we identified in overlay assays, lost activity upon restreaking, and the cosmid isolated from this clone did not confer antibacterial activity on *R. metallidurans* upon retransformation (Fig. 1B). This cosmid was fully sequenced (GenBank accession no. GQ869384), and the eDNA insert was found to contain a 16-kb nonribosomal peptide synthetase (NRPS)-based gene cluster that resembles lipopeptide gene clusters known to produce

antibiotics (see Fig. S4 in the supplemental material). For a description of the RM35 NRPS-based gene cluster and its predicted product, see the supplemental material.

In screens using libraries hosted by a second representative betaproteobacterium, *B. graminis*, we identified two antibacterially active clones surrounded by hazy zones of growth inhibition and six clones displaying light brown to dark brown pigmentation. Neither antibacterially active clone reproduced the antibacterial phenotype upon retransformation. As seen with brown clones found in *P. putida*- and *R. metallidurans*-hosted libraries, all of the brown *B. graminis* clones showed evidence of elevated heme metabolite production by normal-phase TLC, and thus none of these clones were pursued further.

Alphaproteobacteria. A nonpathogenic strain of the rhizosphere-dwelling bacterium *A. tumefaciens* was chosen as a representative of the *Alphaproteobacteria* for its ease of transformation and compatibility with IncP group plasmids (35). Although no antibacterially active clones were detected in overlay assays, one yellow-pigmented clone and two clones with altered colony morphology were identified during screens using this host.

The yellow-pigmented clone, AT1, found while screening the JPA library (Fig. 1F) is identical to the cosmid clone, RM3, that we identified in an earlier screen of this library for clones that confer the production of color to *R. metallidurans* (11). This clone contains a six-gene operon that is predicted to code for the production of the carotenoid β -carotene. This is the only case where we identified the same clone in screens using different bacterial hosts. The carotenoid biosynthetic gene cluster found in this clone sits directly adjacent to and colinear with the N-terminal fragment of the *sacB* gene into which eDNA inserts are cloned. Only 59 bp of eDNA was captured upstream of the first ORF in the carotenoid biosynthetic operon, suggesting that the vector-associated *sacB* promoter, and not an eDNA-derived promoter, may be responsible for driving the expression of this gene cluster in both *R. metallidurans* and *A. tumefaciens*.

The two *A. tumefaciens* clones that display altered colony morphology, AT3 and AT5 (Fig. 1G), were identified from the JPA and RPA libraries, respectively. These clones share a similarly wrinkled appearance, and both are viscous upon physical manipulation. While recent studies suggest that small molecules may play a role in regulating colony phenotypes by exerting control over motility, adhesion, and biofilm formation, no clone-specific metabolites were identified in the culture broths or cell pellets of either clone (4). The eDNA inserts from both AT3 and AT5 were fully sequenced (GenBank accession no. GQ869381 and GQ869382), and no putative ORFs with homology to known biosynthetic enzymes were identified.

eDNA library screens using a second representative alphaproteobacterium, *C. vibrioides*, did not result in the identification of clones displaying any of the predefined phenotypes evaluated in this study.

Retransformation of active cosmids into alternative hosts. The three eDNA libraries used in this study were screened at multiple-library coverage to help ensure that all library clones were examined in each host. To address the possibility that some cosmid DNAs found to confer a phenotype of interest were not actually screened in each host species during our

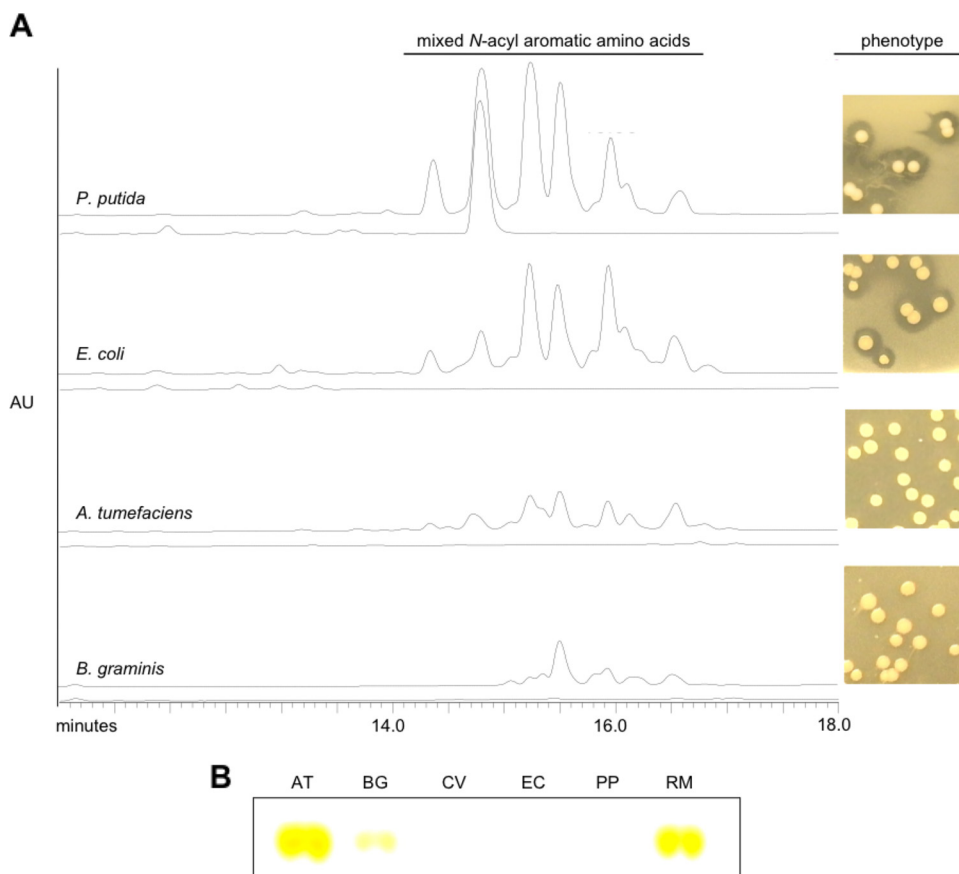


FIG. 4. (A) Analytical HPLC traces of the organic extracts derived from cultures of different proteobacteria transformed with clone EC5. HPLC traces of extracts derived from vector control cultures are shown below each EC5 HPLC trace. The region containing *N*-acyl aromatic amino acids is highlighted. Representative images of the phenotype observed for each host in an antibacterial overlay assay are shown alongside the HPLC traces. AU, absorbance units. (B) Analytical TLC analysis was performed on acetone extracts from the cell pellets of different proteobacteria harboring cosmid RM3/AT1. The region of the TLC plate that is predicted to contain β -carotene is shown. AT, *A. tumefaciens*; BG, *B. graminis*; CV, *C. vibrioides*; EC, *E. coli*; PP, *P. putida*; RM, *R. metallidurans*.

initial high-throughput screens, each cosmid of interest was transformed separately into all six hosts. These transformants were then examined in phenotypic assays to determine the species specificity of the phenotypes associated with each eDNA cosmid. In only one case did an easily observable phenotype appear in a host species that differed from the one identified in the initial high-throughput screens. While this set of eDNA clones did not generally confer easily observable phenotypes to other hosts, there were several examples where the new transformants were found to produce lower amounts of a secondary metabolite(s) than that required for detection in our original phenotypic screens.

Upon retransformation, RM3/AT1, RM44, RM57, AT3, and AT5 all conferred the originally observed phenotypes on the specific hosts they were recovered from, but none conferred these phenotypes to any other host used in this study. In one case, eDNA cosmid RM44 could not be successfully transformed into *B. graminis* despite numerous trials.

Although high-throughput screening of *P. putida*-based libraries was not feasible using the two-layer overlay method, individual cosmid clones hosted by *P. putida* are amenable to this technique. All cosmids that conferred antibacterial activity to other hosts

were therefore subjected to overlay assays in *P. putida*. From these overlay assays, cosmid EC5 was found to confer the production of antibacterial activity to *P. putida* (Fig. 4). HPLC analysis of the ethyl acetate extract of the culture broth of *P. putida* transformed with cosmid EC5 revealed the presence of mixed *N*-acyl aromatic amino acids in amounts comparable to that found when using *E. coli* as the expression host (Fig. 4A). A similar analysis was performed with extracts of the culture broths of the remaining four hosts that did not produce an obvious antibacterial phenotype when transformed with the EC5 cosmid. This analysis revealed trace amounts of mixed *N*-acyl aromatic amino acids in extracts of both the *A. tumefaciens* and *B. graminis* culture broths. These trace levels were apparently too low to be detected by a two-layer overlay screen.

Despite the potential for vector-driven expression of the RM3/AT1 carotenoid biosynthetic gene cluster, no hosts other than *R. metallidurans* and *A. tumefaciens* were found to express the intense yellow-pigmented phenotype associated with this cosmid. Because of the potential role the vector-associated promoter may play in the production of carotenoids by RM3/AT1, the production of β -carotene by each host was assessed

by TLC analysis of acetone-extracted cell pellets. Significant amounts of β -carotene were detected only in the extracts from *R. metallidurans* and *A. tumefaciens* (Fig. 4B). A small amount of β -carotene was found in the extract derived from *B. graminis*, although this amount of pigment was not sufficient to appreciably alter the pigmentation of *B. graminis* colonies.

DISCUSSION

The majority of bacteria from soil and aquatic habitats are not readily cultivated using standard microbiological methods. The development of methods to study these organisms by examining DNA cloned directly from environmental samples has been essential to the culture-independent analysis of microbial ecosystems. To access metagenome-derived enzymatic and biosynthetic abilities without prior knowledge of the genetic elements required to confer these abilities, the emerging field of functional metagenomics relies on the screening of arrayed metagenomic libraries for clones displaying functional attributes and phenotypes. The success or failure of functional metagenomics is a function of the host strain used to process the foreign genetic material within eDNA libraries. As the most commonly used host for metagenomic libraries, *E. coli* is estimated to readily express only 40% of the genes derived from diverse microbial origins, with a strong bias against the expression of genes from certain groups of distantly related organisms (17). An expansion of the host repertoire used in functional metagenomics is therefore likely to increase both the number and the diversity of eDNA clones identified from functional metagenomic studies.

The construction of eDNA libraries that are compatible with an expanded host repertoire could be achieved either through the parallel construction of libraries with unique origins of replication or through the construction of eDNA libraries using broad-host-range cloning vectors (26, 28, 42). The IncP1-*a* group of broad-host-range plasmids boasts perhaps the most extensive list of compatible species and therefore represented an obvious choice as the foundation for a broad-host-range eDNA cloning vector (1). Plasmids possessing the RK2 replicon, found within many IncP1-*a* group derivative cloning vectors, are thought to be stably maintained at low copy numbers in the majority of gram-negative bacterial species and in some instances can be stably transferred to gram-positive bacteria, yeast, and mammalian cells (16, 20, 41). In this study, we were able to take advantage of the diverse host range of RK2-based cloning vectors to create eDNA libraries that were stably maintained in six different bacterial expression hosts.

From this initial broad-host-range functional metagenomic study, it is clear that not all hosts perform equally well when subjected to simple phenotypic assays, and it remains a distinct possibility that some surrogate expression hosts will be better suited to using foreign genetic material than others. High-throughput screens of *E. coli*-, *R. metallidurans*-, and *A. tumefaciens*-based libraries each yielded clones that confer the production of small molecules on heterologous hosts, while the *C. vibrioides*-based libraries did not yield any clones displaying the phenotypes of interest. Some phenotypes were also clearly more readily detected in our screens than others. In particular, heme- and melanin-producing clones, which have been reported numerous times from the screening of *E. coli*-based

libraries, were also found in *P. putida*-, *R. metallidurans*-, and *B. graminis*-based libraries. While we have screened for clones that produce antibiotics or colored metabolites, any easily reproducible phenotypic change has the potential to be exploited for the identification of new molecules or genes by functional metagenomics. The altered morphologies seen in environmental clones AT3 and AT5, hosted by *A. tumefaciens*, exemplify how the heterologous expression of eDNA can be used to confer readily observable phenotypes on a receptive expression host.

This study also serves to highlight an important requirement for functional metagenomic screening. For a small-molecule-derived phenotype to be detected using functional metagenomics, the production of a compound(s) associated with a given phenotype must be at or above a certain threshold level. The ability of several hosts in our study to produce detectable amounts of clone-specific metabolites in sensitive analytical assays yet not display relevant clone-specific phenotypes appears to be the result of subthreshold small-molecule production by specific clone-host pairs. The propensity of some hosts to more commonly reach and surpass these threshold levels is an important mechanism by which individual hosts will likely confer an advantage on functional metagenomic screening efforts.

Our previous work established *R. metallidurans* as the first gram-negative host, other than *E. coli*, to lead to the successful identification of secondary-metabolite-producing clones from eDNA libraries (11). In the present study, *R. metallidurans* again proved itself to be a valuable addition to the functional metagenomic screening toolkit, as screens of the RPA library identified several antibacterially active clones, including an antibacterially active clone that produces the novel natural product *N*-(4-amino-2-hydroxybutyl)tetradecanamide. This clone, RM44, along with clone RM35, which harbors an orphan NRPS-type biosynthetic gene cluster, and the previously identified small-molecule-producing eDNA clones RM3 and RM57, demonstrates the potential *R. metallidurans* has for identifying biosynthetic genes and gene clusters from eDNA. The continued exploration of phylogenetically diverse bacteria as hosts for functional metagenomic screening is likely to identify additional microbes that will be rewarding hosts for these types of studies.

During the course of our high-throughput eDNA library screens, there was only one case where the same small-molecule-producing clone was identified using two different host species. Distantly related host species therefore appear unlikely to identify the exact same subset of eDNA clones from phenotype-dependent screens of soil eDNA libraries, despite nearly identical assay conditions and selection criteria. This study provides the most tangible evidence to date that broad-host-range vectors can be effectively exploited for the creation of soil eDNA libraries suitable for functional metagenomic studies using multiple expression hosts. More importantly, our findings suggest that doing so is likely to expand the functional and genetic diversity of eDNA clones obtained from such screens, as well as increase the number of phenotype-specific lead clones available for further study.

ACKNOWLEDGMENTS

This work was supported by the Howard Hughes Medical Institute, NIH grant GM077516, the Beckman Foundation, the Searle Foundation, and NIH grant MSTP GM07739 (J.W.C.).

REFERENCES

- Adamczyk, M., and G. Jagura-Burdzy. 2003. Spread and survival of promiscuous IncP-1 plasmids. *Acta Biochim. Pol.* **50**:425–453.
- Anantharaman, V., and L. Aravind. 2003. Application of comparative genomics in the identification and analysis of novel families of membrane-associated receptors in bacteria. *BMC Genomics* **4**:34.
- Bagdasarian, M., R. Lurz, B. Ruckert, F. C. Franklin, M. M. Bagdasarian, J. Frey, and K. N. Timmis. 1981. Specific-purpose plasmid cloning vectors. II. Broad host range, high copy number, RSF1010-derived vectors, and a host-vector system for gene cloning in *Pseudomonas*. *Gene* **16**(1–3):237–247.
- Bassler, B. L., and R. Losick. 2006. Bacterially speaking. *Cell* **125**:237–246.
- 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., 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., and J. Clardy. 2004. Palmitoylputrescine, an antibiotic isolated from the heterologous expression of DNA extracted from bromeliad tank water. *J. Nat. Prod.* **67**:1283–1286.
- 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.
- Claros, M. G., and G. von Heijne. 1994. TopPred II: an improved software for membrane protein structure predictions. *Comput. Appl. Biosci.* **10**:685–686.
- Craig, J. W., F. Y. Chang, and S. F. Brady. 2009. Natural products from environmental DNA hosted by *Ralstonia metallidurans*. *ACS Chem. Biol.* **4**:23–28.
- Delcher, A. L., D. Harmon, S. Kasif, O. White, and S. L. Salzberg. 1999. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.* **27**:4636–4641.
- Dower, W. J., J. F. Miller, and C. W. Ragsdale. 1988. High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res.* **16**:6127–6145.
- Fierer, N., M. A. Bradford, and R. B. Jackson. 2007. Toward an ecological classification of soil bacteria. *Ecology* **88**:1354–1364.
- Finn, R. D., J. Tate, J. Misty, P. C. Coghill, S. J. Sammut, H. R. Hotz, G. Ceric, K. Forslund, S. R. Eddy, E. L. Sonnhammer, and A. Bateman. 2008. The Pfam protein families database. *Nucleic Acids Res.* **36**:D281–D288.
- Friedman, A. M., S. R. Long, S. E. Brown, W. J. Buikema, and F. M. Ausubel. 1982. Construction of a broad host range cosmid cloning vector and its use in the genetic analysis of *Rhizobium* mutants. *Gene* **18**:289–296.
- Gabor, E. M., W. B. Alkema, and D. B. Janssen. 2004. Quantifying the accessibility of the metagenome by random expression cloning techniques. *Environ. Microbiol.* **6**:879–886.
- Gilchrist, A., and J. Smit. 1991. Transformation of freshwater and marine caulobacters by electroporation. *J. Bacteriol.* **173**:921–925.
- Handelsman, J., M. R. Rondon, S. F. Brady, J. Clardy, and R. M. Goodman. 1998. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem. Biol.* **5**:R245–R249.
- Heinemann, J. A., and G. F. Sprague, Jr. 1989. Bacterial conjugative plasmids mobilize DNA transfer between bacteria and yeast. *Nature* **340**:205–209.
- Herring, P. 1972. Porphyrin pigmentation in deep-sea medusae. *Nature* **238**:276–277.
- Hoekema, A., P. R. Hirsch, P. J. J. Hooykaas, and R. A. Schilperoort. 1983. A binary plant vector strategy based on separation of *vir*- and T-region of the *Agrobacterium tumefaciens* Ti-plasmid. *Nature* **303**:179–180. (Letter.)
- Huang, Y., X. Lai, X. He, L. Cao, Z. Zeng, J. Zhang, and S. Zhou. 2009. Characterization of a deep-sea sediment metagenomic clone that produces water-soluble melanin in *Escherichia coli*. *Mar. Biotechnol. (NY)* **11**:124–131.
- Kim, J. S., H. K. Lim, M. H. Lee, J. H. Park, E. C. Hwang, B. J. Moon, and S. W. Lee. 2009. Production of porphyrin intermediates in *Escherichia coli* carrying soil metagenomic genes. *FEMS Microbiol. Lett.* **295**:42–49.
- Lee, C. M., Y. S. Yeo, J. H. Lee, S. J. Kim, J. B. Kim, N. S. Han, B. S. Koo, and S. H. Yoon. 2008. Identification of a novel 4-hydroxyphenylpyruvate dioxygenase from the soil metagenome. *Biochem. Biophys. Res. Commun.* **370**:322–326.
- Li, Y., M. Wexler, D. J. Richardson, P. L. Bond, and A. W. Johnston. 2005. Screening a wide host-range, waste-water metagenomic library in tryptophan auxotrophs of *Rhizobium leguminosarum* and of *Escherichia coli* reveals different classes of cloned *trp* genes. *Environ. Microbiol.* **7**:1927–1936.
- Lin, J.-J. 1994. Optimization of the transformation efficiency of *Agrobacterium tumefaciens* cells using electroporation. *Plant Sci.* **101**:11–15.
- Martinez, A., S. J. Kolvek, C. L. Yip, J. Hopke, K. A. Brown, I. A. MacNeil, and M. S. Osburne. 2004. Genetically modified bacterial strains and novel bacterial artificial chromosome shuttle vectors for constructing environmental libraries and detecting heterologous natural products in multiple expression hosts. *Appl. Environ. Microbiol.* **70**:2452–2463.
- Mergeay, M., D. Nies, H. G. Schlegel, J. Gerits, P. Charles, and F. Van Gijsegem. 1985. *Alcaligenes eutrophus* CH34 is a facultative chemolithotroph with plasmid-bound resistance to heavy metals. *J. Bacteriol.* **162**:328–334.
- Perham, R. N. 1991. Domains, motifs, and linkers in 2-oxo acid dehydrogenase multienzyme complexes: a paradigm in the design of a multifunctional protein. *Biochemistry* **30**:8501–8512.
- Plonka, P. M., and M. Grabacka. 2006. Melanin synthesis in microorganisms—biotechnological and medical aspects. *Acta Biochim. Pol.* **53**:429–443.
- Poindexter, J. S. 1964. Biological properties and classification of the *Caulobacter* group. *Bacteriol. Rev.* **28**:231–295.
- Rappé, M. S., and S. J. Giovannoni. 2003. The uncultured microbial majority. *Annu. Rev. Microbiol.* **57**:369–394.
- Rondon, M. R., P. R. August, A. D. Bettermann, S. F. Brady, T. H. Grossman, M. R. Liles, K. A. Loiacono, B. A. Lynch, I. A. MacNeil, C. Minor, C. L. Tiong, M. Gilman, M. S. Osburne, J. Clardy, J. Handelsman, and R. M. Goodman. 2000. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl. Environ. Microbiol.* **66**:2541–2547.
- Schmidhauser, T. J., and D. R. Helinski. 1985. Regions of broad-host-range plasmid RK2 involved in replication and stable maintenance in nine species of gram-negative bacteria. *J. Bacteriol.* **164**:446–455.
- Simon, R., U. Priefer, and A. Pühler. 1983. A broad host range mobilization system for *in vivo* genetic engineering: transposon mutagenesis in gram negative bacteria. *Bio/Technology* **1**:784–791.
- Slater, S. C., B. S. Goldman, B. Goodner, J. C. Setubal, S. K. Farrand, E. W. Nester, T. J. Burr, L. Banta, A. W. Dickerman, I. Paulsen, L. Otten, G. Suen, R. Welch, N. F. Almeida, F. Arnold, O. T. Burton, Z. Du, A. Ewing, E. Godsy, S. Heisel, K. L. Houmiel, J. Jhaveri, J. Lu, N. M. Miller, S. Norton, Q. Chen, W. Phoolcharoen, V. Ohlin, D. Ondrusek, N. Pride, S. L. Stricklin, J. Sun, C. Wheeler, L. Wilson, H. Zhu, and D. W. Wood. 2009. Genome sequences of three *Agrobacterium* biovars help elucidate the evolution of multichromosome genomes in bacteria. *J. Bacteriol.* **191**:2501–2511.
- Taghavi, S., D. van der Lelie, and M. Mergeay. 1994. Electroporation of *Alcaligenes eutrophus* with (mega)plasmids and genomic DNA fragments. *Appl. Environ. Microbiol.* **60**:3585–3591.
- Torsvik, V., J. Goksoyr, and F. L. Daae. 1990. High diversity in DNA of soil bacteria. *Appl. Environ. Microbiol.* **56**:782–787.
- Viallard, V., I. Poirier, B. Cournoyer, J. Haurat, S. Wiebkin, K. Ophel-Keller, and J. Balandreau. 1998. *Burkholderia graminis* sp. nov., a rhizospheric *Burkholderia* species, and reassessment of *Pseudomonas phenazinum*, *Pseudomonas pyrocinia*, and *Pseudomonas glathei* as *Burkholderia*. *Int. J. Syst. Bacteriol.* **48**(Pt. 2):549–563.
- Waters, V. L. 2001. Conjugation between bacterial and mammalian cells. *Nat. Genet.* **29**:375–376.
- Wexler, M., P. L. Bond, D. J. Richardson, and A. W. Johnston. 2005. A wide host-range metagenomic library from a waste water treatment plant yields a novel alcohol/aldehyde dehydrogenase. *Environ. Microbiol.* **7**:1917–1926.
- Zhang, H., Y. Wang, and B. A. Pfeifer. 2008. Bacterial hosts for natural product production. *Mol. Pharm.* **5**:212–225.
- Zhou, J., M. A. Bruns, and J. M. Tiedje. 1996. DNA recovery from soils of diverse composition. *Appl. Environ. Microbiol.* **62**:316–322.