RESEARCH ARTICLE

Polyketide synthase pathways identified from a metagenomic library are derived from soil *Acidobacteria*

Larissa C. Parsley[1], Jonathan Linneman[2], Ann M. Goode[1], Kristen Becklund[3], Isabelle George[4], Robert M. Goodman[5], Nicole B. Lopanik[2] & Mark R. Liles[1]

[1]Department of Biological Sciences, Auburn University, Auburn, AL, USA; [2]Department of Biology, Georgia State University, Atlanta, GA, USA; [3]Department of Plant Pathology, University of Wisconsin, Madison, WI, USA; [4]Unité de Génie Biologique, Université Catholique de Louvain, Louvain-la-Neuve, Belgium; and [5]School of Environmental and Biological Sciences, Rutgers, The State University of New Jersey, New Brunswick, NJ, USA

Correspondence: Mark R. Liles, Department of Biological Sciences, Auburn University, 101 Life Sciences Building, 120 W. Samford Ave., Auburn, AL 36849, USA. Tel.: +334 844 1656; fax: +334 844 1645; e-mail: lilesma@auburn.edu

Received 10 September 2010; revised 22 April 2011; accepted 26 April 2011.

DOI:10.1111/j.1574-6941.2011.01122.x

Editor: Christoph Tebbe

Keywords
polyketide synthase; soil; macroarray; ketosynthase; metagenomics; *Acidobacteria*.

Abstract

Polyketides are structurally diverse secondary metabolites, many of which have antibiotic or anticancer activity. Type I modular polyketide synthase (PKS) genes are typically large and encode repeating enzymatic domains that elongate and modify the nascent polyketide chain. A fosmid metagenomic library constructed from an agricultural soil was arrayed and the macroarray was screened for the presence of conserved ketosynthase [β-ketoacyl synthase (KS)] domains, enzymatic domains present in PKSs. Thirty-four clones containing KS domains were identified by Southern hybridization. Many of the KS domains contained within metagenomic clones shared significant similarity to PKS or nonribosomal peptide synthesis genes from members of the *Cyanobacteria* or the *Proteobacteria* phyla. However, analysis of complete clone insert sequences indicated that the BLAST analysis for KS domains did not reflect the true phylogenetic origin of many of these metagenomic clones that had a %G+C content and significant sequence similarity to genes from members of the phylum *Acidobacteria*. This conclusion of an *Acidobacteria* origin for several clones was further supported by evidence that cultured soil *Acidobacteria* from different subdivisions have genetic loci closely related to PKS domains contained within metagenomic clones, suggesting that *Acidobacteria* may be a source of novel polyketides. This study also demonstrates the utility of combining data from culture-dependent and -independent investigations in expanding our collective knowledge of microbial genomic diversity.

Introduction

The isolation of secondary metabolites produced by soil microorganisms has historically been an effective means of identifying novel compounds with antimicrobial activity. Soil environments contain an impressive 10^{10} microbial cells g^{-1}, while only a relatively small fraction of the extant microorganisms has been cultivated in the laboratory (Hugenholtz *et al.*, 1998; Curtis & Sloan, 2005). While this small percentage of cultured soil microorganisms has already illustrated the phylogenetic diversity of these communities and has provided many important antibiotics to date, a potentially richer resource for novel antibiotic discovery resides within as yet uncultured bacteria (Rondon *et al.*, 2000; Gillespie *et al.*, 2002; Handelsman, 2004). The application of a metagenomic approach to isolate large, contiguous regions of DNA from uncultured microorganisms is a viable strategy to access intact (or nearly intact) genetic pathways that are involved in natural product synthesis. Previous studies have demonstrated the utility of screening environmental metagenomic libraries to identify bioactive molecules (Courtois *et al.*, 2003), proteolytic systems (Beja *et al.*, 2000), and polyketide synthases (PKS) (Moffitt & Neilan, 2003; Ginolhac *et al.*, 2004; Schirmer *et al.*, 2005; Wawrik *et al.*, 2005).

Polyketides are a diverse group of bioactive secondary metabolites produced by bacteria, fungi, and plants and include several therapeutically important drugs such as erythromycin, tetracyclines, immunosuppressants FK506 and rapamycin, and antitumor compounds doxorubicin

FEMS MICROBIOLOGY ECOLOGY

and mithramycin (Staunton & Weissman, 2001). Polyketide biosynthetic pathways are generally classified into three major groups: Type I, Type II, or Type III. Type II PKSs are composed of individual proteins that perform one enzymatic activity iteratively to catalyze the formation of aromatic polyketides such as the antibiotic actinorhodin (Hertweck *et al.*, 2007). Type III PKSs participate in the assembly of small aromatic compounds and are found in plants and bacteria (Moore & Hopke, 2001). Type I PKSs are large, modular, multifunctional proteins composed of enzymatic domains that typically perform a single reaction in polyketide chain assembly (Fischbach & Walsh, 2006). The biosynthesis of Type I polyketides proceeds in an assembly-line fashion: the starter unit is loaded onto the loading module and transferred to the first extender module. After extension and modification by the first module, the nascent polyketide is passed on to the next extender module until it reaches the thioesterase domain, which releases the chain in either a cyclized or a linear form. In most cases, modules have three enzymatic domains that are necessary to catalyze one cycle of chain elongation: β-ketoacyl synthase (KS), acyltransferase, and acyl carrier protein (ACP). The action of optional β-keto reductive domains (ketoreductase, dehydratase, and enoyl reductase) contributes to the diversity of polyketide natural products. Tailoring enzymes such as glycosyl transferases, methyltransferases, and monooxygenases can further modify the metabolites after they are released from the PKS (Rix *et al.*, 2002). The modularity and colinearity of enzyme and product of Type I PKSs make them attractive for genetic engineering and combinatorial biosynthesis to produce novel metabolites (Kittendorf & Sherman, 2006).

While the synthesis of polyketides is often associated with bacteria within the phylum *Actinobacteria* (Staunton & Weissman, 2001; Stinear *et al.*, 2004), PKS genes and metabolites have also been identified in a diverse group of bacteria, including species of *Pseudomonas* (El-Sayed *et al.*, 2003), *Stigmatella* (Beyer *et al.*, 1999), *Sorangium* (Schupp *et al.*, 1995; Beyer *et al.*, 1999; Julien *et al.*, 2000), *Myxococcus* (Simunovic *et al.*, 2006), *Bacillus* (Chen *et al.*, 2006), and *Burkholderia* (Partida-Martinez & Hertweck, 2005). The widespread distribution of PKS pathways is of interest because additional reservoirs of polyketides in soil microorganisms likely exist, but have remained uncharacterized due to cultivation biases. In this study, we used degenerate primers and probes to identify potentially novel PKS pathways from a soil metagenomic library via macroarray hybridization. The isolation and manipulation of PKS pathways using a metagenomic approach provides access to the uncharacterized PKS pathways of as yet uncultured soil microorganisms (Schirmer *et al.*, 2005; Brady *et al.*, 2009) and allows for the targeted detection of such pathways from specific bacterial taxa that are not typically associated with this type of biosynthetic pathway, such as members of the phylum *Acidobacteria*.

Materials and methods

Soil collection and DNA isolation

Soil samples were collected from soil cores (10–50 cm depth) from an agricultural soil in the Central Sands area of Wisconsin at the University of Wisconsin-Madison's Hancock Agricultural Research Station (HARS) in Hancock, WI. The soil was a Plainfield loamy sand taken from the plow zone at HARS. Bacterial cells were collected after soil homogenization by differential centrifugation and then embedded and lysed within an agarose plug. High-molecular-weight metagenomic DNA was recovered and purified as described previously (Liles *et al.*, 2008, 2009).

Fosmid library construction

The metagenomic library was constructed in the pCC1FOS vector with the CopyControl^{TM} fosmid library system according to the manufacturer's instructions (Epicentre, Madison, WI). Briefly, purified metagenomic DNA was randomly sheared and size selected from a pulsed-field agarose gel before ligation into the pCC1FOS vector. Ligated DNA was then packaged (MaxPlax Lambda Packaging Extract) and titered onto the EPI100^{TM}-T1 cloning strain (Epicentre) and plated onto Luria–Bertani containing 12.5 μg mL^{-1} chloramphenicol for the selection of recombinant clones. To induce the production of higher fosmid copy numbers, arabinose (0.01% v/v) was added to the media with the targeted clones. The recombinant *Escherichia coli* clones were robotically picked in duplicate onto nylon membranes at the Clemson University Genomics Institute, resulting in duplicate clone macroarrays. Macroarrays were generated with and without arabinose induction of fosmid copy number and contained 9216 fosmid clones with an average insert size of 42 kb (*c.* 100 *E. coli* genome equivalents).

Screening of fosmid macroarrays

Pooled fosmid DNA served as a template in a PCR to generate a PKS probe for downstream library screening. The β-ketosynthase (KS) domain was targeted with the degenerate primer set 5LL and 4UU (Table 1) (D. Sherman, University of Michigan, pers. commun.). The amplification was conducted with a touchdown PCR, with a 2-min denaturation at 95 °C, followed by 14 cycles of a touchdown PCR (95 °C for 30 s, annealing beginning at 63 °C and ramping down 1 °C per cycle to 50 °C for 30 s, and 72 °C for 1 min) and then 30 cycles of amplification (95 °C for 30 s, 50 °C for 30 s, and 72 °C for 1 min), and then 5 min at 72 °C. Seven other Type I and Type II PKS-specific or NRPS-

Table 1. Oligonucleotide primers used in this study

1		
	A3	GCSTACSYSATSTACACSTCSGG
	A7R	SASGTCVCCSGTSCGGTAS
2		
	K1	TSAAGTCSAACATCGGBCA
	M6R	CGCAGGTTSCSGTACCAGTA
3		
	KSLF	CCSCAGSAGCGCSTSYTSCTSGA
	KSLR	GTSCCSGTSCCGTGSGYSTCSA
4		
	KSDPQQF	MGNGARGCNNWNSMNATGGAYCCNCARCANMG
	KSHGTGR	GGRTCNCCNARNSWNGTNCCNGTNCCRTG
5		
	5LL	GGRTCNCCIARYTGIGTICCIGTICCRTGIGC
	4UU	MGIGARGCIYTICARATGGAYCCICARCARMG
6		
	F	TSGCSTGCTTGGAYGCSATC
	R	TGGAANCCGCCGAABCCGCT
7		
	540F	GGITGCACSTCIGGIMTSGAC
	1100R	CCGATSGCICCSAGIGAGTG
8		
	kSALPHA	TTCGGCGGXTTCCAGTCXGCCATG
	ACP	CCXATGCTCAGCXACCGCGACGACCT
9		
	mtaD-F	GTGGTGCTGGAGCAGTCG
	mtaD-R	TCACCGCTCAGCGATGTC

Primer sets 1 through 8 were used in the generation of ketosynthase-specific probes (Seow *et al*., 1997; Metsa-Ketela *et al*., 2002; Piel, 2002; Ginolhac *et al*., 2004; Ayuso-Sacido & Genilloud, 2005; Wawrik *et al*., 2005). Primer set 5, specific to Type I KS domains, was the only primer set that yielded a PCR product using a pooled metagenomic library DNA template while not giving a similar product with *Escherichia coli* genomic DNA template. Primer set 9 was specific to the *mtaD* homologous sequence present in fosmid clone A12 and was used to identify *mtaD* gene sequences from cultured *Acidobacteria*.

specific primer sets were also evaluated for probe synthesis (Table 1), but were not used because they either yielded PCR products from control reactions containing *E. coli* genomic DNA as a template or did not yield the expected PCR product based on sequences of representative amplicons. The Type I KS domain amplicon resulting from PCR with the primers 5LL and 4UU was labeled with digoxigenin using the DIG PCR Synthesis system (Roche, Indianapolis, IN) and used as a probe in macroarray screening by colony blot hybridization at 42 °C with medium hybridization washing stringency as per the manufacturer's instructions. The hybridization was performed using the digoxigenin detection system (Roche) with detection mediated by the CSPD chemiluminescent substrate. The detection of positive clones was accomplished after multiple exposure times for each blot. Clones that exhibited chemiluminescence in duplicate ($n = 34$) were considered KS positive and subjected to further analyses.

Table 2. List of *Acidobacteria* strains, their phylogenetic classification, and the presence or absence of genetic loci associated with type I PKS pathways

Acidobacteria strain	*Acidobacteria* subdivision	KS domain	*mtaD* domain
IGE012	1	+	−
IGE013	1	+	−
IGE015	3	−	+
IGE017	4	−	+
IGE001	6	−	+
IGE002	6	+	+

The *Acidobacteria* cultured isolates (George *et al*., manuscript in preparation) were screened by PCR using either KS domain (Table 1, primer set 5) or *mtaD*-specific (Table 1, primer set 9) primer sets. In every case of a negative PCR result, a control PCR was performed and the genomic DNA template did provide a 16S rRNA gene amplicon using the 'universal' Bacteria primers 27F and 1492R (Lane *et al*., 1985). The phylogeny of the *Acidobacteria* KS or *mtaD* genetic loci are presented in Figs 2 and 4, respectively.

Verification and sequencing of KS domains from positive clones and *Acidobacteria*-cultured isolates

To confirm that the 34 hybridization-positive clones were unique and to estimate insert size, restriction fragment length polymorphism (RFLP) profiles were generated for each clone using the restriction enzyme BamHI. The resulting digest reactions were electrophoresed on a 0.8% (w/v) 1 × TAE pulsed-field agarose gel at 5 V cm^{-1} with a 1–15-s switch time for 16 h.

Twenty-one of the KS-positive clones produced a KS amplicon following PCR with the 5LL/4UU primer set to verify the presence of a KS domain. To obtain sequence data from these clones, the KS primers were modified with T3 (5LL) and T7 (4UU) primer recognition sequences at the 5′-end of each respective primer and used in a secondary PCR to generate KS amplicons for sequencing. The KS genes were then sequenced using primers T3 and T7, which resulted in higher quality sequence data than using the degenerate 5LL or 4UU primers. All 34 KS-positive clones were end sequenced using the vector primers EpiFOSF and EpiFOSR.

KS domains were also PCR amplified from three *Acidobacteria* genomes out of the six *Acidobacteria* isolates that were tested (I. F. George, M. Hartmann, M. R. Liles & S. N. Agathos, unpublished data) (Table 2). Each *Acidobacteria* isolate was grown on 1/100th strength nutrient agar for 3 weeks and bacterial cells were recovered from the plate using a sterile swab. Genomic DNA was extracted from *Acidobacteria* cultures using a genomic DNA isolation kit (Promega). The 5LL/4UU primer set was used in a PCR with *Acidobacteria* genomic DNA as a template, as described above. Because the yield of genomic DNA was low for each *Acidobacteria* isolate, a control PCR targeting the 16S rRNA

gene was performed for each DNA sample using the 'universal bacteria' primers 27F (5′-AGAGTTTGATC MTGGCTCAG-3′) and 1492R (5′-GYTACCTTGTTAC GACTT-3′) (Lane *et al.*, 1985), and only isolates that yielded a 16S rRNA gene amplicon were included in this study. KS domain PCR products were cloned into the pCR2.1 vector (Invitrogen, Carlsbad, CA) and sequenced using vector primers. A consensus DNA sequence was generated for each respective KS amplicon derived from an *Acidobacteria* culture. The KS domain sequences were submitted to GenBank under the accession numbers JF342575–JF342588.

KS domain phylogenetic analysis

KS gene sequences from 19 PCR-positive fosmid clones and three *Acidobacteria* isolates were aligned with KS domain sequences from other bacterial sources using CLUSTALX software. A minimum evolution phylogenetic tree was generated and bootstrap analysis (5000 iterations) was conducted using MEGA4 (Tamura *et al.*, 2007). Note that the KS domains from two of the 21 PCR-positive fosmid clones were excluded due to poor sequence quality.

Southern hybridization of putative KS-positive clones

To investigate the 13 clones that were KS-positive after colony blot hybridization, but failed to yield a KS PCR product, a Southern hybridization was performed using the same digoxigenin-labeled heterogeneous KS probe as that used in previous experiments. Fosmid DNA from each of the 13 clones was restriction digested with BamHI, electrophoresed through a 1% (w/v) 1 × TAE agarose gel, transferred to a nylon membrane using the Whatman TurboBlotter system (Kent, UK), and hybridized with the KS probe. DNA fragments hybridized to the KS gene probe were visualized using the NBT/BCIP colorimetric detection method (Roche).

Insert sequencing of fosmid DNA clones

Ten clones were selected for complete insert sequencing by 454 pyrosequencing based on their predicted phylogenetic and functional domains. Fosmid DNA was extracted from clones A2, A3, A11, A12, B1, B8, B10, B11, C1, and C5 using the Qiagen Large-Construct DNA Isolation kit (Qiagen, Valencia, CA), with an additional plasmid-safe exonuclease digestion step for 4 h at 37 °C to remove contaminating genomic DNA. Purified DNA was sent to the Lucigen Corporation (Middleton, WI) for bar-coded shotgun subclone library construction and then sequenced at the EnGenCore Center at the University of South Carolina (Columbia, SC) using a Genome Sequencer FLX system (Roche, Nutley, NJ) according to the manufacturer's instructions. One half of a pyrosequencing plate was devoted to the fosmid clones, providing at least 47-fold coverage for each fosmid DNA insert. Primer walking and Sanger sequencing was used to extend DNA sequences from fosmid clone contigs. DNA sequencing reactions were conducted using a fosmid DNA template with specific oligonucleotide primers at the Lucigen Corporation or the Georgia State University DNA/Protein Core facility (Atlanta, GA). The insert sequences for clones A2, A3, A11, A12, and C5 were submitted to GenBank under the accession numbers of JF342589–JF342593, respectively.

Bioinformatic analysis of fosmid-derived ORFs

Fosmid DNA sequences generated by 454 sequencing were assembled into contiguous fragments (contigs) using the CLC Genomics Workbench (Cambridge, MA) assembly algorithm, and contigs were analyzed for the presence of putative PKS-related genes. ORFs were identified using a GENEMARK heuristic approach for gene prediction in prokaryotes (http://opal.biology.gatech.edu/GeneMark/). Additionally, GLIMMER 3.02 and NCBI's ORF FINDER (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) were utilized to corroborate the predicted ORFs obtained from GENMARK analysis. The prediction of gene function was accomplished by comparing each ORF sequence against the GenBank nr/nt database using the BLASTP and BLASTN algorithms. Any predicted ORFs lacking any significant homology to other gene products in GenBank (E value > 0.001) were further investigated for secondary structures (profiles, patterns, blocks, motifs, and protein families) using a web server (http://motif.genome.jp/). Additional assembly and ORF identification for clone A3 were performed using LASERGENE software (DNASTAR, Madison, WI). Putative gene functions were identified using BLASTX and BLASTP algorithms in the GenBank nr/nt database, with further motif analysis generated by the ExPASy Proteomics Server (http://ca.expasy.org).

Analysis of clone A12

Preliminary sequence data from clone A12 indicated that this clone may have significant homology to the genome of *Solibacter usitatus* from the phylum *Acidobacteria*. Additionally, a sequence with significant similarity to an MtaD-like protein involved in polyketide (myxothiazol) biosynthesis was detected on this clone. Because the fosmid clone A12 insert sequence was not complete, a Southern blot using an *mtaD* probe was conducted to verify its presence on the clone. Purified DNA from fosmid A12 was restriction digested with the enzymes BamHI, ClaI, EcoRI, EcoRV, Hind III, NotI, PstI, SalI, SmaI, SphI, or XbaI and electrophoresed through a 0.8% (w/v) 1 × TAE agarose gel. The digested DNA was transferred to a nylon membrane and

hybridized with the *mtaD* homolog probe as described above for the KS domain probe.

Detection of *mtaD* homologs from cultured *Acidobacteria*

Pure bacterial cultures representing subdivisions 1, 3, 4, and 6 of *Acidobacteria* (Table 2) were screened for *mtaD* homologous sequences by PCR using *mtaD*-specific primers developed from the fosmid A12 sequence (Table 1). The following touchdown thermal cycling program was used: initial denaturing at 94 °C for 2 min, followed by 14 cycles of denaturing at 94 °C for 30 s, stepwise annealing from 63 to 50 °C for 30 s, and extension at 72 °C for 2 min, and an additional 30 cycles of the above with a 50 °C annealing temperature. Amplicons of the predicted size of 482 bp were subcloned into the pCR 2.1 vector (Invitrogen) according to the manufacturer's instructions and inserts were amplified via colony PCR using the vector primers M13F and M13R. The resulting PCR amplicons were purified using the Promega Wizard PCR Clean-up System (Madison, WI) and submitted to the Auburn University Genetic Analysis Laboratory for sequencing. The sequences were trimmed for quality and compared with their nearest significant neighbors in GenBank by multiple alignment (CLUSTALX) and maximum parsimony analysis (MEGA4).

Results

Generation of a KS-specific probe

Multiple primer sets specific to PKS or NRPS pathways were tested for their use to prepare a probe for macroarray hybridization (Table 1). In each experiment with different primer sets, a pooled fosmid library DNA template was used as a template. Every primer set that generated a product from the library template (but not the *E. coli* DNA template) was cloned and a representative number of clones ($n = 8$) were sequenced to ensure that the amplicon was the intended gene target. The only primer set that produced the desired PCR amplicon using a metagenomic library template and not with host genomic DNA template was primer set 4UU/5LL that is specific to the KS domain, the most conserved region of Type I PKS pathways (Beja *et al.*, 2000; Hertweck *et al.*, 2007). The KS domain amplicon resulting from a pooled metagenomic library DNA template was selected for digoxigenin labeling and subsequent macroarray hybridization.

Identification of KS domain-containing clones from the fosmid library

A total of 34 clones from the arabinose-induced macroarray exhibited chemiluminescence after hybridization with the digoxigenin-labeled Type I KS probe (in duplicate), indicating the presence of a KS domain and therefore a possible PKS pathway (Fig. 1a). Only 19 KS-containing clones were identified from the macroarray prepared without arabinose induction (data not shown), which were also detected in the arabinose-induced macroarray in addition to 15 other clones. The position of the duplicate clones relative to one another enabled the determination of the specific fosmid clone identity and RFLP profiles of the 34 clones confirmed that they contained unique inserts (Fig. 1b). Of the 34 fosmid clones that were identified by hybridization, only 21 yielded a PCR product using primers 4UU and 5LL, suggesting that the remaining 13 clones may contain KS domains that are phylogenetically divergent from known genes or may be misidentified clones or clones with mismatches with the 4UU/5LL primer set.

Phylogenetic analysis of KS-positive clones

DNA sequences derived from KS amplicons obtained from the fosmid clones were compared with the GenBank nr/nt database and were found to have significant similarity to known KS genes (similarities ranging from 51 to 73%). As the highest percent similarity in amino acid sequence homology observed was 73%, it is likely that these KS domains and linked pathways are phylogenetically distant from the previously characterized KS domains present in GenBank databases. Minimum evolution analysis of library-derived KS domains and known KS domains indicates that the KS domains identified from the soil metagenomic

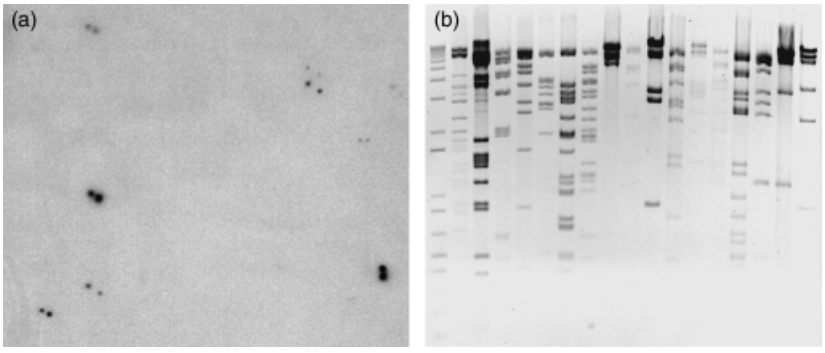


Fig. 1. (a) Clones from a soil metagenomic library exhibit chemiluminescence after hybridization with a digoxigenin-labeled KS domain amplicon probe prepared from a pooled metagenomic library DNA template. Clones are spotted in duplicate, so that replicate hybridization signals provide an indication of positive clone hybridization. (b) BamHI restriction digests of KS domain-containing fosmid clones were used to estimate the size for each metagenomic clone insert.

library represent diverse lineages, but are still related to known KS domains (Fig. 2). The KS domains encoded by these fosmid clones showed similarity to those found in pathways responsible for a variety of polyketide compounds, such as the myxobacterial compounds epothilone and stigmatellin (Beyer *et al.*, 1999), the cyanobacterial jamaicamide (Edwards *et al.*, 2004), and the actinobacterial polyketide antibiotics erythromycin (Donadio *et al.*, 1991) and pikromycin (Xue *et al.*, 1998). Although many of these library-derived KS domains have similarity to known KS domains, the resultant chemical moieties produced by these pathways in their native host may be structurally distinct from known polyketides because the other modules present in these PKS pathways may not be similar in organization to known pathways.

Southern hybridization of PCR-negative fosmid clones

Of the 34 fosmid clones identified by the original library macroarray hybridization, 13 clones did not yield a KS domain PCR product despite multiple attempts at PCR amplification using a variety of reaction conditions. To determine whether each of these fosmid clone DNAs would hybridize with KS domains, a Southern hybridization was performed using a digoxigenin-labeled KS domain probe generated by PCR amplification of the KS domains from a pool of the 21 KS domain PCR-positive fosmid DNAs. Eight of the 13 PCR-negative clones yielded KS-positive Southern blot results, thus suggesting the presence of a putative KS domain on these clones that had a mismatch(es) with the 4UU/5LL primer set. The five fosmid clones that were originally considered KS-positive by macroarray hybridization, but have not yielded a KS PCR product or shown a positive result in subsequent hybridizations may represent false-positive macroarray hybridizations, misidentification of the fosmid clone from the metagenomic library stored at $-80\,^{\circ}C$, or may require lower stringency hybridization conditions for the detection of divergent KS domains.

454 sequencing of fosmid clones

Ten fosmid clones ($n = 7$ PCR positive; $n = 3$ PCR negative) were selected for whole-insert sequencing by 454 pyrosequencing based on their predicted phylogeny and likelihood of containing PKS pathways. Preliminary phylogenetic analysis of these clones indicated that the predicted bacterial lineages of the clone insert sequences include members of the phyla *Acidobacteria*, *Cyanobacteria*, and multiple subdivisions of the *Proteobacteria*. To date, four of the 10 clone inserts have been completely assembled and contain many ORFs with similarity to gene products from various bacterial lineages associated with polyketide synthesis, such as acyl carrier proteins and β-ketoacyl synthases, as well as cell transporters and translocation systems and enzymes involved in polyketide modification [fosmid clone A3 (PCR positive) depicted in Fig. 3; fosmid clones A2 (PCR negative), A11 (PCR negative), and C5 (PCR positive) shown in Supporting Information, Fig. S1]. Clone A3 revealed the presence of predicted genes with significant similarity to genes involved in both PKS and NRPS synthesis that suggests a hybrid pathway; in addition, on one end of the clone A3 insert are genes with significant similarity to genes from the *Acidobacteria* isolate *S. usitatus* and a high %G+C content consistent with an *Acidobacteria* origin (Fig. 3). Preliminary annotation of the remaining six clones has also revealed the presence of similar ORFs as well as many sequences with no significant homologs in GenBank or that have currently unassigned functions (data not shown).

Bioinformatic analysis of clone A12

Sequence data from clone A12 generated two points of interest. First, this clone appears to contain a gene with 52% identity and 70% similarity to a gene that is homologous to *mtaD*, known to encode an enzyme from the myxobacterium *Stigmatella aurantiaca* that is involved in chain extension during the synthesis of the polyketide myxothiazol (Silakowski *et al.*, 1999). Second, much of the DNA carried on clone A12 shares significant similarity to the genome of *S. usitatus*, a representative of the phylum *Acidobacteria* (see Fig. S2).

However, the MtaD-like domain did not assemble with other contigs from A12 based on available fosmid sequences; thus, Southern hybridization was performed to verify its presence on this clone. Multiple bands were observed for clone A12 using restriction enzymes that were not predicted to cut within the probe sequence, indicating that this may be a repeating domain within the clone (data not shown), and explains its problematic assembly with other clone insert sequences. The *mtaD* homolog gene sequence was used to generate primer sets specific to this domain to query cultured *Acidobacteria* genomes.

Identification of *mtaD* homologous sequences and KS domains from cultured *Acidobacteria*

Using primer set #9 targeting the *mtaD* homolog from metagenomic clone A12 or primer set #5 targeting KS domains (Table 1), the cultured *Acidobacteria* were screened for the presence of *mtaD* or KS domains. Four *mtaD* homologs were identified from *Acidobacteria* strains IGE015 (subdivision 3), IGE017 (subdivision 4), IGE002 (subdivision 6), and IGE001 (subdivision 6), showing at least 99% identity between the *mtaD* homolog sequence from metagenomic clone A12 and each of the *Acidobacteria* *mtaD* amplicon sequences. Minimum evolution trees of MtaD homologous sequences and their nearest neighbors

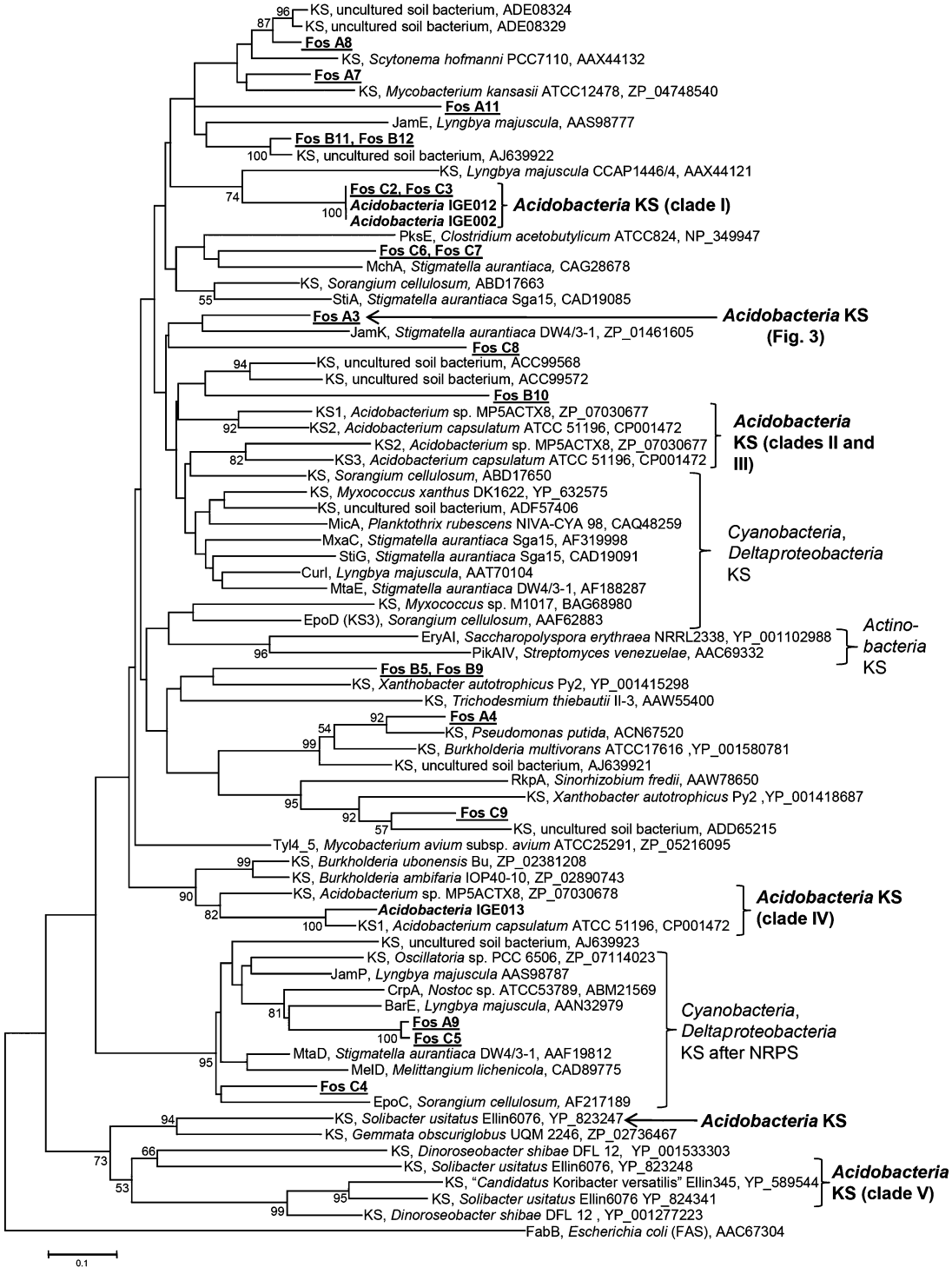
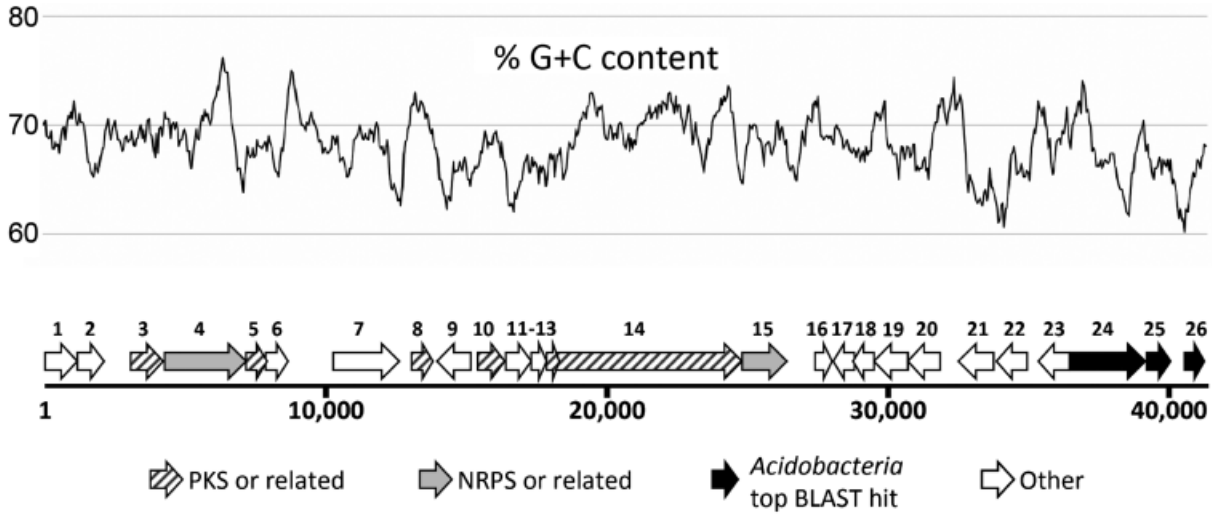


Fig. 2. Minimum evolution tree based on the amino acid sequences of KS gene sequences derived from metagenomic clones, *Acidobacteria* cultures (in bold), and the GenBank nr/nt database. KS domain sequences recovered from the soil metagenomic library are underlined and in bold. KS amino acid sequences from fosmids C2 and C3 were 98.6% identical, B5 and B9 were 99% identical, C6 and C7 were 100% identical, and B11 and B12 were 100% identical. KS sequences from fosmids C4, C5, and A9 group with a clade of KSs that follow NRPS enzymatic domains for hybrid PKS/NRPS biosysnthesis in *Cyanobacteria* and*Deltaproteobacteria*. Bootstrap analysis (performed 5000 times) with percentages $> 50\%$ are indicated at the nodes. Scale bar represents 0.1 amino acid substitutions per site.



ORF	Length (bp)	Top BLASTp hit	Nearest neighbor	E-value	Similarity (%)
1	1164	Cysteine desulfurase family protein	*Heliobacterium modesticaldum*	1e-88	66
2	966	Peptidase U32	*Pirellula staleyi*	5e-16	48
3	1170	Butyryl-CoA dehydrogenase	*Bacillus* sp. SG-1	6e-53	54
4	2877	AMP-dependent synthetase and ligase *	*Nostoc punctiforme* *	8e-133	61
5	744	ABC transporter related (macrolide-specific)	*Clostridium phytofermentans*	2e-53	68
6	798	Hypothetical protein	*Oceanobacter* sp. RED65	2e-25	52
7	2370	Hypothetical protein	*Halorhodospira halophila*	1e-40	45
8	768	4'-phosphopantetheinyl transferase	*Cyanothece* sp. PCC 7424	7e-23	48
9	1203	Rieske domain protein	*Stigmatella aurantiaca*	2e-76	57
10	957	Acyltransferase domain protein	*Stigmatella aurantiaca*	2e-34	48
11	924	Putative dioxygenase	*Stigmatella aurantiaca*	4e-109	80
12	546	Hypothetical protein	*Stigmatella aurantiaca*	1e-42	62
13	468	Oxidoreductase, short chain dehydrogenase/reductase family	*Stigmatella aurantiaca*	3e-13	56
14	6720	Tylactone synthase modules 4 and 5	*Stigmatella aurantiaca*	0.0	65
15	1590	AMP-dependent synthetase and ligase	*Stigmatella aurantiaca*	3e-168	70
16	603	Collagen triple helix repeat-containing protein	*Syntrophobotulus glycolicus*	4e-16	48
17	702	Putative branched-chain amino acid transport protein	*Moritella* sp. PE36	2e-86	82
18	768	Urea ABC transporter	*Tolumonas auensis*	2e-98	81
19	1182	Possible ABC transporter permease	*Cellvibrio japonicus*	6e-149	82
20	1170	Putative urea short-chain amide or branched-chain amino acid uptake ABC transporter permease	*Photobacterium profundum*	3e-135	77
21	1266	Putative UreA short-chain amide	*Methylobacterium* sp. 4-46	0.0	88
22	1089	Phosphate-selective porin O and P	*Methylovorus* sp. SIP3-4	3e-43	54
23	1098	Hypothetical protein	uncultured marine microorganism	4e-57	56
24	2613	PpiC-type peptidyl-prolyl cis-trans isomerase	*Solibacter usitatus*	0.0	68
25	864	Electron transport protein	*Solibacter usitatus*	2e-76	68
26	678	Hypothetical protein	*Solibacter usitatus*	9e-30	58

Fig. 3. Predicted ORFs from metagenomic clone A3 after complete insert sequencing. ORFs were categorized based on their predicted functions after comparison with the GenBank nr/nt database. (The third BLASTP hit is shown for ORF 4 in order to correlate the predicted function with motif analysis.)

in the GenBank nr/nt database show significant bootstrap support for the affiliation of the fosmid A12 MtaD with homologs in *Acidobacteria* subdivisions 4 and 6, as well as support for the monophyly of all *Acidobacteria* MtaD sequences (Fig. 4). Likewise, *Acidobacteria* strains IGE002 (subdivision 6), IGE012 (subdivision 1), and IGE013

(subdivision 1) had KS domains with significant similarity to KS domain sequences derived from metagenomic clones, and a minimum evolution tree reveals that several of the *Acidobacteria* and metagenomic KS domains together form coherent clades (Fig. 2).

Discussion

We constructed and screened a metagenomic library from soil to detect the presence of Type I PKS pathways within large-insert clones. Probing of the fosmid library, which is estimated to contain over 400 Mbp of cloned metagenomic DNA, revealed the presence of 34 clones that were positive for a KS domain. With the metagenomic library representing approximately 100 *E. coli* genome equivalents and assuming an equal relative abundance of genomes within the library, then at this hit frequency, approximately 40% of the genomes represented within the library contained a KS domain. Compared with previous estimates of KS domain-positive cultured bacteria of 15–25% in different soil types (Khatun *et al.*, 2002), this is a higher frequency, suggesting at the very least that a culture-independent survey of PKS pathways is capable of identifying a significant yield of PKS-containing clones at a frequency within the range of expectation. The observed PKS-containing clone hit frequency of 0.4% for this library is similar to the 0.5–0.7% frequency observed for marine sponge metagenomic libraries (Schirmer *et al.*, 2005) and the 0.2% frequency from a soil metagenomic library (Ginolhac *et al.*, 2004). The other acknowledged constraint in screening a fosmid library is the limited insert size possible for any given clone. Other soil metagenomic libraries have been constructed in bacterial artificial chromosome vectors that exceed the average insert size of this fosmid library, but this study does report the detection of diverse KS-positive clones within a macroarrayed metagenomic library and allow an assessment of the phylogenetic origin and PKS pathway architecture among these clones.

Upon first comparing the KS domains present on metagenomic clones with the GenBank database, it was apparent that the majority of the KS domain-positive clones encoded KS domains most similar to those from *Cyanobacteria* and *Myxobacteria* taxa (among other phyla) and were phylogenetically and (presumably) functionally diverse, including some expected to be involved in lipopolysaccharide and polyketide synthesis, as well as hybrid polyketide/nonribosomal peptide chain elongation. Additional bioinformatic analysis of complete insert sequences from the KS-positive clones indicated many predicted ORFs with significant homology to sequences encoding KS domains, ACPs, and enzymes involved in the postsynthesis modification of various polyketides (Fig. 3 and Figs S1 and S2).

However, the presumed phylogenetic affiliation of KS domains identified from some metagenomic clones with KS domains from members of the phyla *Cyanobacteria* and *Proteobacteria* was poorly supported by weak sequence similarity (maximum 73% similarity) and by weak bootstrap support from the phylogenetic analysis. These data called into question the origin of these metagenomic clones and highlighted the gaps in existing sequence databases. The analysis of complete insert sequences (i.e. from clones A3 and A12) indicated that some of the fosmid clones may

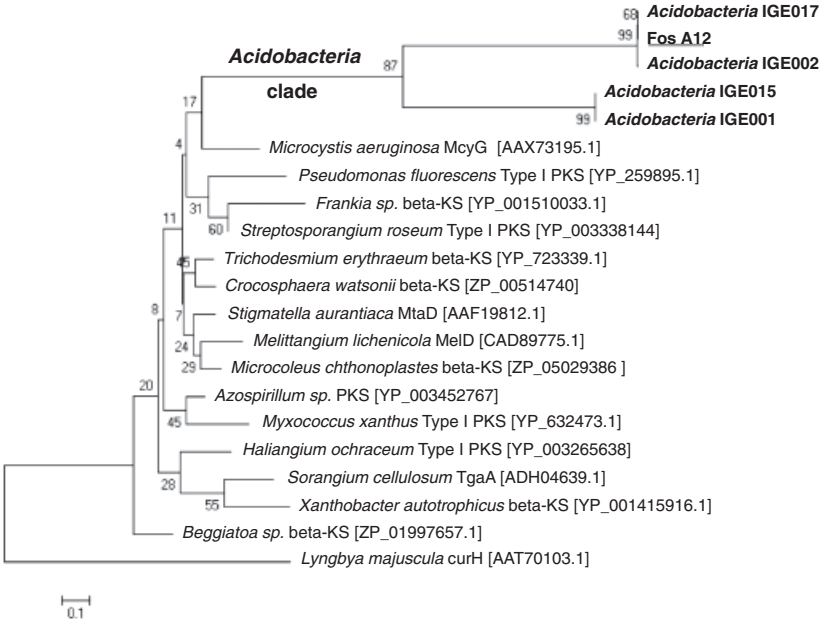


Fig. 4. Minimum evolution tree based on amino acid sequences of MtaD homologous gene sequences derived from metagenomic clone A12 (bold and underlined), *Acidobacteria* cultures (bold), and the GenBank nr/nt database. Numbers at tree nodes represent bootstrap values ($n = 1000$ repetitions; only bootstrap values over 50 are shown), and the tree was rooted with CurH from *Lyngbya majuscula*. Accession numbers for each sequence are in brackets. Scale bar represents 0.1 amino acid substitutions per site. [Correction added 1 June 2011 after online publication: Figure 4 replaced with new version to show current designation of strains]

derive from soil *Acidobacteria*. Fortunately, the availability of cultured *Acidobacteria* isolates from several subdivisions of this phylum provided the ability to query their genomes for PKS domains. The inclusion of PKS domain sequences from these recently cultured *Acidobacteria* indicates that many of these metagenomic clones at first identified as originating from *Cyanobacteria* or *Myxobacteria* taxa are rather derived from this novel phylum that is ubiquitous and abundant in soils. These data demonstrate that the current sequence databases do not adequately reflect the extant diversity of soil bacterial PKS pathways. Some of the observed disparity between the predicted phylogenetic affiliations of the KS domains compared with their flanking sequences (i.e. clone A3 in Fig. 3) may result from the paucity of *Acidobacteria* PKS pathways and genomes in GenBank.

A significant percentage (nearly 40%) of the 34 fosmid clones identified by library macroarray hybridization never yielded a KS domain-specific PCR product, leading to the initial conclusion that these may be false-positive identifications or that the incorrect fosmid clones had been selected from the 384-well formatted library. Subsequent Southern hybridization analysis indicated that eight of these 13 clones contained a KS hybridizing domain(s). It is likely that base pair mismatches between the KS domain PCR primers and fosmid clone DNA prevented the successful PCR amplification of KS genes, and yet they shared sufficient similarity for successful hybridization. This is strong support for the utility of metagenomic macroarray hybridization for the inclusion of metagenomic clones representing diverse phyla such as the *Acidobacteria*, especially when the available 'universal' primer sets designed on the basis of gene sequences derived from cultured taxa may not be inclusive of metagenomic diversity.

Sequence analysis of the PCR-negative clones revealed the presence of PKS-related ORFs that are predicted to originate from diverse bacterial lineages, such as the *Acidobacteria*, *Cyanobacteria*, *Planctomycetes*, and *Proteobacteria* phyla. Many of the predicted gene products appear to be involved in polyketide biosynthesis, postsynthesis modification, or the transport of such compounds outside of the cell. Because of the phylogenetically diverse nature of these sequences, it is unlikely that all of the pathways would have been functionally active in *E. coli*, thus underscoring the advantage of a sequence-based screening approach for the initial identification, to be followed by research to express PKS domains or pathways within other heterologous host(s) predicted to be closely related to the microorganism from which the metagenomic clone was derived.

Analysis of clone A12 revealed the presence of a gene whose top BLAST hit MtaD originates from the myxobacterium *S. aurantiaca* and is involved in the hybrid PKS/NRPS biosynthetic pathway of myxothiazol. Subsequent investigation revealed that clone A12 contained several ORFs with homology to the genome of *S. usitatus* from the *Acidobacteria* phylum. Using the available *Acidobacteria* culture collection, we explored the presence of MtaD-like domains from recently cultured *Acidobacteria*, including two isolates from the previously uncultivated subdivision 6. MtaD-like sequences were identified from four *Acidobacteria* isolates. There was significant affiliation observed between metagenome- and *Acidobacteria*-derived sequences for both MtaD and KS domains. Because these cultured isolates represent multiple *Acidobacteria* subdivisions and were found within soils collected in Belgium, these data indicate a potentially widespread distribution and diversity of PKS pathways among *Acidobacteria* taxa. Further characterization of these *Acidobacteria* strains and their polyketide products will contribute to our understanding of natural product discovery from previously uncultured bacterial phyla. To our knowledge, no polyketide product has been identified from an *Acidobacteria* sp., although the results of one genome sequencing study indicate that members of this phylum may harbor PKS-related genes (Ward *et al.*, 2009). These findings reflect the scarcity of known *Acidobacteria* sequences and indicate that these bacteria may represent an unexplored reservoir of novel PKS pathways and polyketide products.

Acknowledgements

We thank the members of the Liles laboratory for their help in the execution of these experiments and their critical analysis of this publication. Funding for this work was provided by the National Institutes of Health (grant # 1R21AI083852-01).

References

Ayuso-Sacido A & Genilloud O (2005) New PCR primers for the screening of NRPS and PKS-I systems in actinomycetes: detection and distribution of these biosynthetic gene sequences in major taxonomic groups. *Microb Ecol* **49**: 10–24.

Beja O, Suzuki MT, Koonin EV *et al.* (2000) Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environ Microbiol* **2**: 516–529.

Beyer S, Kunze B, Silakowski B & Mueller R (1999) Metabolic diversity in myxobacteria: identification of the myxalamid and the stigmatellin biosynthetic gene cluster of *Stigmatella aurantiaca* Sg a15 and a combined polyketide–(poly)peptide gene cluster from the epothilone producing strain *Sorangium cellulosum* So ce90. *Biochim Biophys Acta* **1445**: 185–195.

Brady SF, Simmons L, Kim JH & Schmidt EW (2009) Metagenomic approaches to natural products from free-living and symbiotic organisms. *Nat Prod Rep* **26**: 1488–1503.

Chen XH, Vater J, Piel J *et al.* (2006) Structural and functional characterization of three polyketide synthase gene clusters in *Bacillus amyloliquefaciens* FZB 42. *J Bacteriol* **188**: 4024–4036.

Courtois S, Cappellano CM, Ball M *et al.* (2003) Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. *Appl Environ Microb* **69**: 49–55.

Curtis TP & Sloan WT (2005) Exploring microbial diversity – a vast below. *Science* **309**: 1331–1333.

Donadio S, Staver MJ, McAlpine JB, Swanson SJ & Katz L (1991) Modular organization of genes required for complex polyketide biosynthesis. *Science* **252**: 675–679.

Edwards DJ, Marquez BL, Nogle LM, McPhail K, Goeger DE, Roberts MA & Gerwick WH (2004) Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium *Lyngbya majuscula*. *Chem Biol* **11**: 817–833.

El-Sayed AK, Hothersall J, Cooper SM, Stephens E, Simpson TJ & Thomas CM (2003) Characterization of the mupirocin biosynthesis gene cluster from *Pseudomonas fluorescens* NCIMB 10586. *Chem Biol* **10**: 419–430.

Fischbach MA & Walsh CT (2006) Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. *Chem Rev* **106**: 3468–3496.

Gillespie DE, Brady SF, Bettermann AD, Cianciotto NP, Liles MR, Rondon MR, Clardy J, Goodman RM & Handelsman J (2002) Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. *Appl Environ Microb* **68**: 4301–4306.

Ginolhac A, Jarrin C, Gillet B *et al.* (2004) Phylogenetic analysis of polyketide synthase I domains from soil metagenomic libraries allows selection of promising clones. *Appl Environ Microb* **70**: 5522–5527.

Handelsman J (2004) Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol R* **68**: 669–685.

Hertweck C, Luzhetsky A, Rebets Y & Bechthold A (2007) Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Nat Prod Rep* **24**: 162–190.

Hugenholtz P, Goebel BM & Pace NR (1998) Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J Bacteriol* **180**: 4765–4774.

Julien B, Shah S, Ziermann R, Goldman R, Katz L & Khosla C (2000) Isolation and characterization of the epothilone biosynthetic gene cluster from *Sorangium cellulosum*. *Gene* **249**: 153–160.

Khatun S, Waugh ACW, Redpath MB & Long PF (2002) Distribution of polyketide synthase genes in bacterial populations. *J Antibiot* **55**: 107–108.

Kittendorf JD & Sherman DH (2006) Developing tools for engineering hybrid polyketide synthetic pathways. *Curr Opin Biotech* **17**: 597–605.

Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML & Pace NR (1985) Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *P Natl Acad Sci USA* **82**: 6955–6959.

Liles MR, Williamson LL, Rodbumrer J, Torsvik V, Goodman RM & Handelsman J (2008) Recovery, purification, and cloning of high-molecular-weight DNA from soil microorganisms. *Appl Environ Microb* **74**: 3302–3305.

Liles MR, Williamson LL, Rodbumrer J, Torsvik V, Parsley LC, Goodman RM & Handelsman J (2009) Isolation and Cloning of High-Molecular-Weight Metagenomic DNA from Soil Microorganisms. *Cold Spring Harb Press Prot*, DOI:10.1101/pdb.prot5383.

Metsa-Ketela M, Halo L, Munukka E, Hakala J, Mantsala P & Ylihonko K (2002) Molecular evolution of aromatic polyketides and comparative sequence analysis of polyketide ketosynthase and 16S ribosomal DNA genes from various streptomyces species. *Appl Environ Microb* **68**: 4472–4479.

Moffitt MC & Neilan BA (2003) Evolutionary affiliations within the superfamily of ketosynthases reflect complex pathway associations. *J Mol Evol* **56**: 446–457.

Moore BS & Hopke JN (2001) Discovery of a new bacterial polyketide biosynthetic pathway. *Chem BioChem* **2**: 35–38.

Partida-Martinez LP & Hertweck C (2005) Pathogenic fungus harbours endosymbiotic bacteria for toxin production. *Nature* **437**: 884–888.

Piel J (2002) A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of *Paederus* beetles. *P Natl Acad Sci USA* **99**: 14002–14007.

Rix U, Fischer C, Remsing LL & Rohr J (2002) Modification of post-PKS tailoring steps through combinatorial biosynthesis. *Nat Prod Rep* **19**: 542–580.

Rondon MR, August PR, Bettermann AD *et al.* (2000) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl Environ Microb* **66**: 2541–2547.

Schirmer A, Gadkari R, Reeves CD, Ibrahim F, DeLong EF & Hutchinson CR (2005) Metagenomic analysis reveals diverse polyketide synthase gene clusters in microorganisms associated with the marine sponge *Discodermia dissoluta*. *Appl Environ Microb* **71**: 4840–4849.

Schupp T, Toupet C, Cluzel B, Neff S, Hill S, Beck J & Ligon J (1995) A *Sorangium cellulosum* (myxobacterium) gene cluster for the biosynthesis of the macrolide antibiotic soraphen A: cloning, characterization, and homology to polyketide synthase genes from actinomycetes. *J Bacteriol* **177**: 3673–3679.

Seow KT, Meurer G, Gerlitz M, Wendt-Pienkowski E, Hutchinson CR & Davies J (1997) A study of iterative type II polyketide synthases, using bacterial genes cloned from soil DNA: a means to access and use genes from uncultured microorganisms. *J Bacteriol* **179**: 7360–7368.

Silakowski B, Schairer HU, Ehret H *et al.* (1999) New lessons of combinatorial biosynthesis from myxobacteria: the myxothiazol biosynthetic gene cluster of *Stigmatella aurantiaca* DW4/3-1. *J Biol Chem* **274**: 37391–37399.

Simunovic V, Zapp J, Rachid S, Krug D, Meiser P & Muller R (2006) Myxovirescin A biosynthesis is directed by hybrid polyketide synthases/nonribosomal peptide synthetase, 3-hydroxy-3-methylglutaryl-CoA synthases, and trans-acting acyltransferases. *Chem BioChem* **7**: 1206–1220.

Staunton J & Weissman KJ (2001) Polyketide biosynthesis: a millennium review. *Nat Prod Rep* **18**: 380–416.

Stinear TP, Mve-Obiang A, Small PLC *et al.* (2004) Giant plasmid-encoded polyketide synthases produce the macrolide toxin of *Mycobacterium ulcerans*. *P Natl Acad Sci USA* **101**: 1345–1349.

Tamura K, Dudley J, Nei M & Kumar S (2007) *MEGA4*: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**: 1596–1599.

Ward NL, Challacombe JF, Janssen PH *et al.* (2009) Three genomes from the phylum *Acidobacteria* provide insight into the lifestyles of these microorganisms in soils. *Appl Environ Microb* **75**: 2046–2056.

Wawrik B, Kerkhof L, Zylstra GJ & Kukor JJ (2005) Identification of unique type II polyketide synthase genes in soil. *Appl Environ Microb* **71**: 2232–2238.

Xue YQ, Zhao LS, Liu HW & Sherman DH (1998) A gene cluster for macrolide antibiotic biosynthesis in *Streptomyces venezuelae*: architecture of metabolic diversity. *P Natl Acad Sci USA* **95**: 12111–12116.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Predicted ORFs from Clones A2, A11, and C5 after complete insert sequencing.
Fig. S2. Predicted ORFs from metagenomic clone A12, shown with their corresponding predicted functions and nearest neighbors in the GenBank nr/nt database.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.