JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

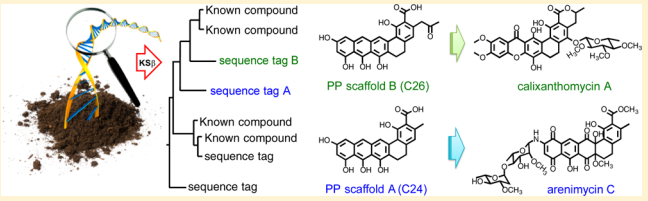
Mining Soil Metagenomes to Better Understand the Evolution of Natural Product Structural Diversity: Pentangular Polyphenols as a Case Study

Hahk-Soo Kang and Sean F. Brady*

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

Supporting Information

ABSTRACT: Sequence-guided mining of metagenomic libraries provides a means of recovering specific natural product gene clusters of interest from the environment. In this study, we use ketosynthase gene (KS) PCR amplicon sequences (sequence tags) to explore the structural and biosynthetic diversities of pentangular polyphenols (PP). In phylogenetic analyses, eDNA-derived sequence tags often fall between closely related clades that are associated with gene clusters known to encode distinct chemotypes. We show that these common "intermediate" sequence tags are useful for guiding the discovery of not only novel bioactive metabolites but also collections of closely related gene clusters that can provide new insights into the evolution of natural product structural diversity. Gene clusters corresponding to two eDNA-derived KS_β sequence tags that reside between well-defined KS_β clades associated with the biosynthesis of (C24)-pradimicin and (C26)-xantholipin type metabolites were recovered from archived soil eDNA libraries. Heterologous expression of these gene clusters in *Streptomyces albus* led to the isolation of three new PPs (compounds **1**−**3**). Calixanthomycin A (**1**) shows potent antiproliferative activity against HCT-116 cells, whereas arenimycins C (**2**) and D (**3**) display potent antibacterial activity. By comparing genotypes and chemotypes across all known PP gene clusters, we define four PP subfamilies, and also observe that the horizontal transfer of PP tailoring genes has likely been restricted to gene clusters that encode closely related chemical structures, suggesting that only a fraction of the "natural product-like" chemical space that can theoretically be encoded by these secondary metabolite tailoring genes has likely been sampled naturally.



1. INTRODUCTION

The analysis of DNA sequence is increasingly being used to guide the discovery of natural product biosynthetic gene clusters capable of encoding for novel metabolites.[1] When studying cultured bacteria, it is possible to fully sequence a bacterial genome and therefore carry out detailed analyses of complete biosynthetic gene clusters to identify those likely to encode for novel metabolites. Similar full gene cluster analyses are not usually possible when exploring complex environmental microbiomes because the immense size of most metagenomes precludes full sequencing and assembly of complete gene clusters. To circumvent this limitation, we and others have explored the use of the detailed phylogenetic analysis of individual, highly conserved natural product biosynthetic genes as markers for guiding the discovery of functionally novel natural product biosynthetic gene clusters from complex metagenomes.[1b,2] In this approach, individual biosynthetic genes are PCR-amplified from environmental DNA (eDNA) and sequenced in bulk. The resulting amplicon sequences (i.e., natural product sequence tags) are then aligned with reference sequences from functionally characterized gene clusters.[2g] In previous studies, we have shown that sequence tags that are only distantly related to any known sequence (e.g., sequence tag A in Figure 1) can guide the identification of gene clusters that encode structurally novel bioactive natural products.[2b,c] We have also shown that sequence tags that are very closely related (e.g., sequence tag B in Figure 1) to known genes can guide the discovery of gene clusters that encode congeners of known molecules.[1b,2d]

While the utility of sequence tags that reside at the phylogenetic boundaries is now clear, many sequence tags fall at a more ambiguous intermediate distance from known sequences (e.g., sequence tag C in Figure 1). We hypothesized that such "intermediate" sequence tags, especially those that fall between clades containing sequences known to encode for structurally distinct subclasses of natural products, might be useful for guiding the discovery of gene clusters that both encode novel bioactive metabolites and provide insights into the evolution of natural products structural diversity. Here, we explore these concepts through the comparison of the genotypes (i.e., gene content) and chemotypes (i.e., molecules encoded by) of eDNA-derived aromatic polyketide (PK) pentangular polyphenols (PPs).

Aromatic PKs are a large group of secondary metabolites generated by iterative (i.e., type II) PKS biosynthetic systems.[3] The initial step in aromatic PK biosynthesis involves the action of

Received: October 15, 2014
Published: December 18, 2014

   dx.doi.org/10.1021/ja510606j | J. Am. Chem. Soc. 2014, 136, 18111−18119

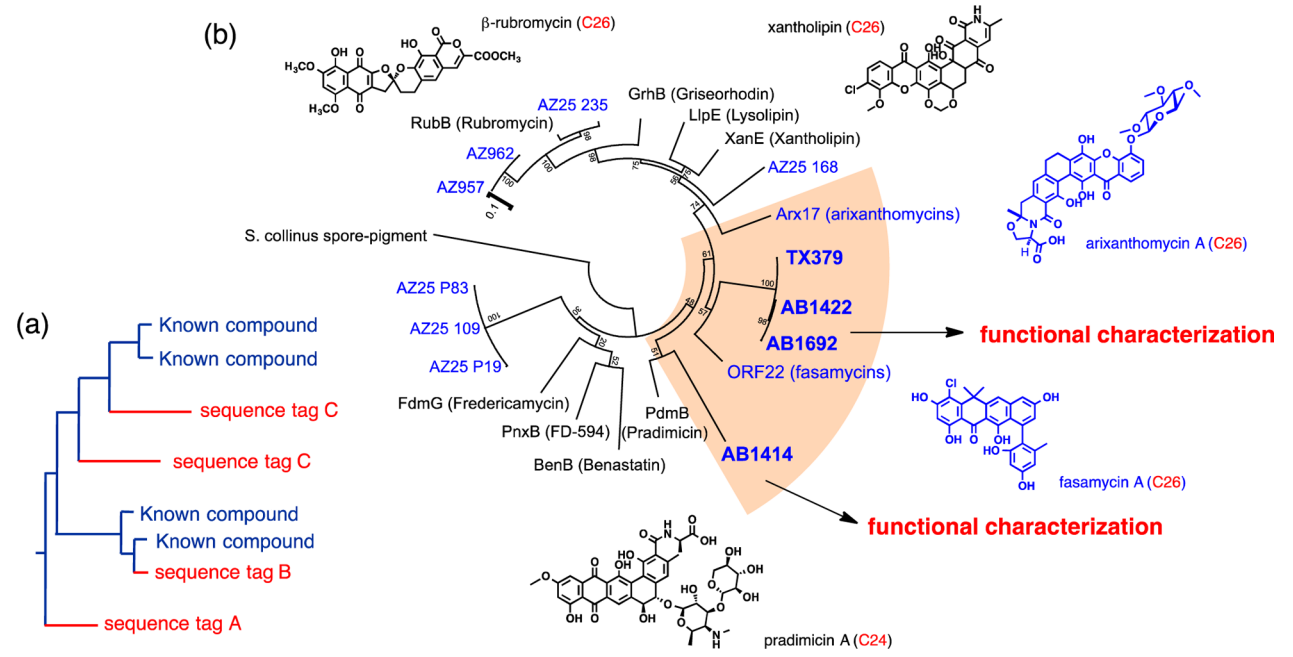


Figure 1. Mining of soil eDNA for pentangular polyphenols. (a) Phylogeny-guided strategy for the discovery of new natural products. (b) The maximum likelihood[5] tree of partial KS_β gene sequences is shown. AZ, AB and TX denote amplicon sequences derived from Arizona, California and Texas desert soil eDNA libraries, respectively. A number of eDNA-derived PP KS_β sequence tags fall between clades associated with PPs derived from polyketide precursors of different lengths (C24 and C26). The specific group of intermediate eDNA sequence tags used in this study are highlighted in orange.

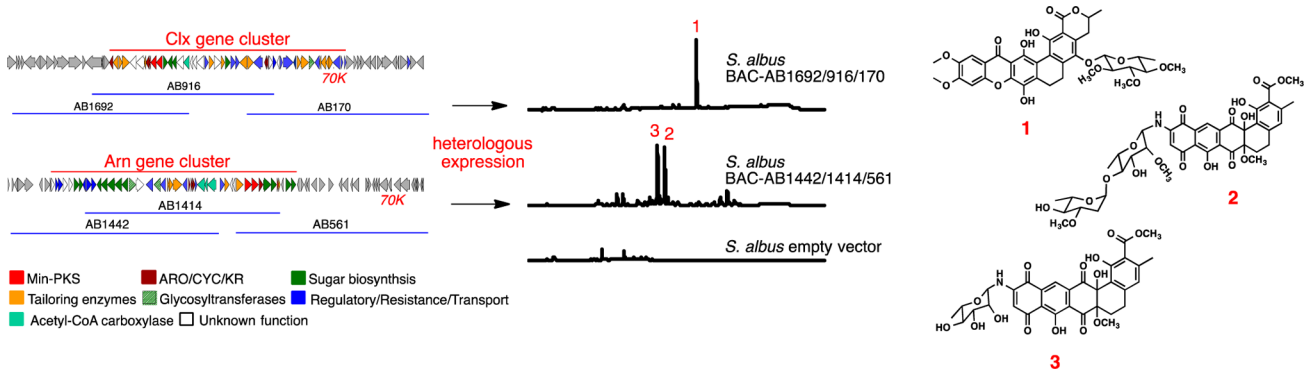


Figure 2. Heterologous expression and the structures of calixanthomycin A (**1**), and arenimycins C (**2**) and D (**3**). Two PP gene clusters were recovered from California desert soil eDNA library on three overlapping eDNA cosmid clones (AB1692/916/170 and AB1442/1414/561). They were each reconstructed into BAC clones using TAR. The heterologous expression of these two PP gene clusters by *S. albus* led to the production of three clone-specific metabolites, compound **1** from the strain *S. albus* BAC-AB1692/916/170, and compounds **2** and **3** from the strain *S. albus* BAC-AB1442/1414/561. Detailed HPLC trace analysis can be found in the Supporting Information (Figure S2).

a minimal polyketide synthase (min-PKS). The min-PKS is composed of the three proteins KS_α, KS_β and ACP, which together catalyze the iterative condensation of malonyl-CoA into polyketides ranging from 16 to 30 carbons in length. PPs are constructed from the largest nonreduced PK precursors known (24−30 carbons in length) and are, therefore, potentially the most structurally diverse subclass of metabolites in this family.[4]

The highly conserved KS_α and KS_β genes from the min-PKS have proved particularly useful as sequence tags for guiding the discovery of new aromatic polyketides from metagenomes.[2b−d,f] While eDNA KS_α and KS_β sequencing studies suggested the presence of large numbers of unique PP clusters in the environment, the biosynthetic landscape of PPs from culture-based studies remained quite limited.[6] In this study PP gene clusters associated with eDNA-derived KS_α and KS_β sequence tags that fall between well-defined clades containing genes from the biosynthesis of pradimicin and xantholipin type PP metabolites were recovered from archived soil eDNA libraries. The molecules encoded by these PP gene clusters were accessed through heterologous expression in *Streptomyces albus* and found to encode new bioactive pentangular polyphenols with potent antiproliferative and antibacterial activities. By comparing the gene cluster genotypes and chemotypes across this expanded pentangular polyphenol biosynthetic landscape, we gain specific insights into pentangular polyphenol biosynthesis and general insights into potential limitations of the evolution of natural product structural diversity.

2. RESULTS AND DISCUSSION

KS_α and KS_β Sequence-Guided Mining of Soil eDNA Libraries. To identify novel gene clusters encoding PP PKs, archived soil eDNA libraries were screened with degenerate primers targeting either the KS_α or KS_β genes of the min-PKS.

dx.doi.org/10.1021/ja510606j | J. Am. Chem. Soc. 2014, 136, 18111−18119

KS_α and KS_β phylogenetic trees were constructed with the resulting eDNA amplicon sequences and sequences from previously characterized type II PKS gene clusters deposited in GenBank. In this phylogenetic analysis we identified two sequence tags of interest (KS_β tag AB1692 and KS_α tag AB1414) from our Anza Borrego Desert (AB) eDNA library. Both sequence tags fall between well-defined clades containing KS genes from gene clusters that encode structurally distinct PP subfamilies, pradimicins and xantholipin (Figure 1b). Sequence tags AB1692 and AB1414 were used to guide the recovery of eDNA cosmid clones containing PP biosynthetic gene clusters associated with these KS sequences. The recovered cosmids were PGM (Personal Genome Machine)-sequenced and this sequence data was subsequently used to guide the recovery of additional cosmid clones overlapping each end of the two min-PKS containing cosmids. Sequencing and bioinformatics annotation (Tables S3 and S4, Supporting Information) of each set of three overlapping cosmid clones revealed a biosynthetic gene cluster flanked by collections of genes predicted to encode primary metabolic enzymes, suggesting that both gene clusters were recovered in their entirety on three overlapping cosmids.

Heterologous Expression of the eDNA-derived PP Gene Clusters. To access the metabolites encoded by the AB1692 and AB1414 PP gene clusters, each set of three overlapping eDNA cosmid clones (AB1692, AB916 and AB170; AB1442, AB1414 and AB561) was assembled into bacterial artificial chromosomes (BAC-AB1692/916/170 and BAC-AB1442/1414/561) using pathway specific pTARa (*E. coli*:yeast:*Streptomyces*) shuttle capture vectors and transformation-associated recombination (TAR) in yeast (Figure 2).[7] These approximately 90 kb BAC constructs were PGM-sequenced to ensure correct assembly and transferred into *Streptomyces albus* by intergenic conjugation and integrated into the *S. albus* genome using the ΦC31 integrase. For heterologous expression purposes, the resulting strains, *S. albus* BAC-AB1692/916/170 and *S. albus* BAC-AB1442/1414/561, were grown at 30 °C in R5A liquid media for 9 days. LC−MS analysis of EtOAc extracts derived from these cultures showed the presence of one and two major clone-specific peaks, respectively (Figure 2). The three clone-specific metabolites were purified from the extracts of 1 L cultures using, in each case, two rounds of C_{18} reversed-phase HPLC. This gave calixanthomycin A (**1**) (12 mg/L) from *S. albus* BAC-AB1692/916/170, and arenimycins C (**2**) (4.6 mg/L) and D (**3**) (2.5 mg/L) from *S. albus* BAC-AB1442/1414/561.

Structures of Calixanthomycin A (1), Arenimycin C (2) and Arenimycin D (3). The structures of compounds **1**−**3** were determined by spectroscopic methods, including HRESIMS, and 1D and 2D NMR (Figure 3, see Supporting Information Discussions S1 and S2 for structure elucidation details).

The structure of calixanthomycin A (**1**) features a xanthone-containing PP core structure. The polycyclic xanthone core is tailored by oxidation, O-methylation and glycosylation. Rare structural features include the F ring lactone, which is more commonly seen as a lactam in xanthone-containing PPs, and the ortho-dimethoxy functionality on the A ring. Calixanthomycin A (**1**) is most closely related to IB-00208 from an unsequenced marine *Actinomadura sp.* (Figure S3).[8] They differ by the position of methoxy groups around the A ring, the presence or absence of the double bond in the D ring and the oxidation state of the C ring.

The structures of arenimycins C (**2**) and D (**3**) produced by *S. albus* BAC-AB1442/1414/561 feature a highly oxidized benzo-

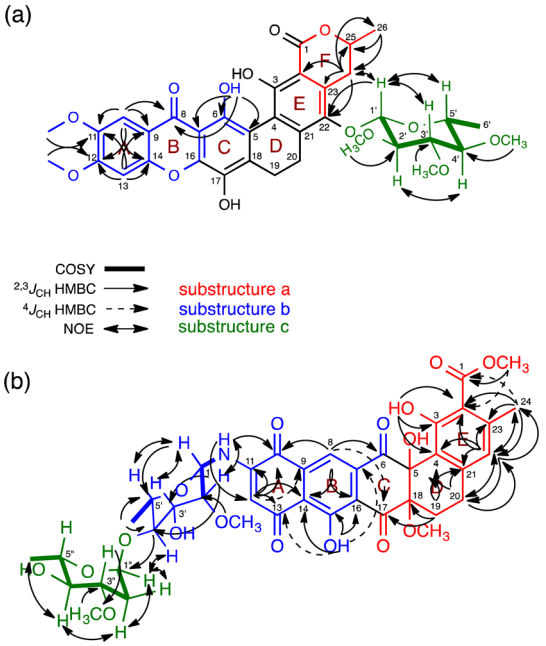


Figure 3. Key 2D NMR correlations used for determining the structures of calixanthomycin A (**1**) and arenimycin C (**2**).

[*a*]naphthacene quinone core (rings A through E). This core structure is appended with a disaccharide (*O*Me-L-rhamnose-OMe-L-Olivose) in **2** and a monosaccharide (L-rhamnose) in **3** via a rare *N*-glycosidic linkage. These structures share an aglycone with the SF2446s from a soil actinomycete *Streptomyces* sp. SF2446, and the recently reported arenimycins A and B from a marine actinomycete *Salinispora* sp. (Figure S3).[9]

Biological Activities of Calixanthomycin A (1), Arenimycin C (2) and Arenimycin D (3). Compounds **1**−**3** were evaluated for antibacterial and antiproliferative activities (Table I). Calixanthomycin A (**1**) is extremely toxic to HCT-116 cancer

Table I. Antibacterial and Antiproliferative Activities of Compounds 1−3

		1	2	3
antibacterial[a]	MRSA USA300	50	0.098	0.19
	B. subtilis RM125	3.1	0.0015	0.39
	VRE EF16	>50	12.5	>50
	E. coli DRC39	>50	>50	>50
antiproliferative[b]	HCT116	0.00043	0.17	2.8

[a]MIC in μg/mL. [b]IC_{50} in μM

cells (IC_{50} 0.43 nM); however, it shows no antibacterial activity against MRSA (Methicillin-resistant *S. aureus*) and VRE (Vancomycin-resistant *Enterococcus faecium*) at the highest concentrations tested (50 μg/mL), and weak activity against *B. subillis* (MIC = 3.1 μg/mL). This is quite interesting as most reported xanthone-containing PPs display a broad spectrum of activity with both potent human cell cytotoxicity and antibacterial activities.[10] Arenimycin C (**2**) shows potent Gram-positive antibacterial activity (MIC = 98 ng/mL against MRSA; 1.5 ng/mL against *B. subtilis*) and moderate cytotoxicity against HCT-116 cells (IC_{50}= 0.17 μM).

Proposed Biosynthetic Schemes for Calixanthomycin A and Arenimycins C and D. The biosynthesis of compounds **1**−**3** can be rationalized based on the predicted functions of the

dx.doi.org/10.1021/ja510606j | J. Am. Chem. Soc. 2014, 136, 18111−18119

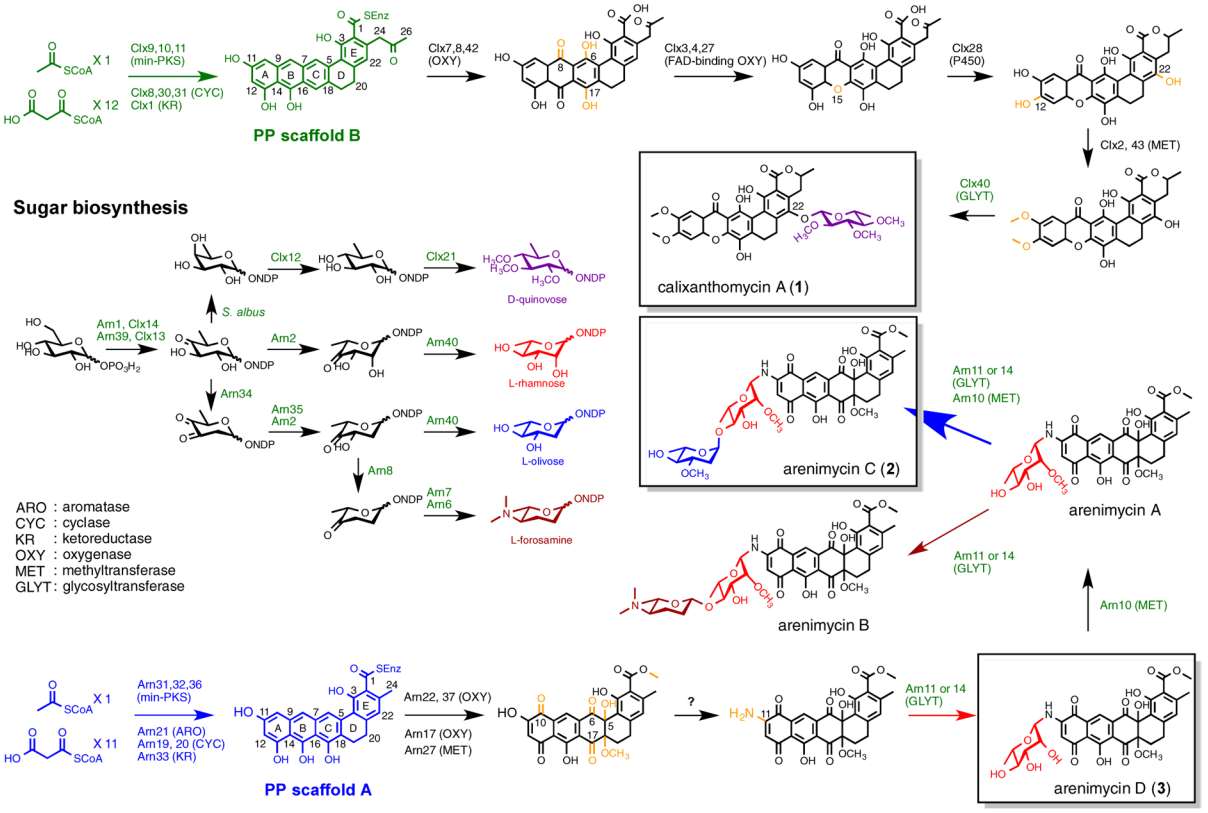


Figure 4. Proposed biosynthesis of calixanthomycin A (**1**), and arenimycins C (**2**) and D (**3**) (in boxes) based on the predicted gene function and Pfam domain analysis. The presence of arenimycin B in the culture extract was confirmed by LC−MS. Detailed gene annotation tables for the Clx and Arn gene clusters appear in Tables S3 and S4.

biosynthetic genes found on each producing BAC construct (Figure 4).

Calixanthomycin A. BAC-AB1692/916/170 is predicted to harbor the 50 kb Clx (calixanthomycin) gene cluster (GenBank KM881706) containing 47 genes involved in the biosynthesis, regulation and resistance of calixanthomycin A (**1**). In our biosynthetic proposal, the predicted min-PKS (Clx9−11), cyclases (Clx8, 30 and 31) and a ketoreductase (Clx1) are responsible for generating a benzo[*a*]naphthacene quinone intermediate using an acetate starter unit and 12 malonyl CoA extension steps. The quinone could then be transformed to a xanthone via an oxidative rearrangement catalyzed by the predicted Bayer-Villiger oxidase (BVO) Clx27 (51% identity to pnxO4 from the FD-594 gene cluster).[6b] The final tailoring steps in our proposed calixanthomycin A (**1**) biosynthesis scheme involve formation of the ortho-dimethoxy functionality in the A ring and glycosylation in the E ring. Formation of the rare ortho-dimethoxy functionality would require reduction at C-13 and hydroxylation at C-12 followed by two O-methylations. The mechanism of C-13 reduction is not clear at this point, but the hydroxylation at C-12 is predicted to be catalyzed by the cytochrome-P450 hydroxylase Clx28 due to its sequence similarity to PnxO5 (61% sequence identity) from the FD-594 gene cluster.[6b] Methylation of the ortho-hydroxyl groups could then occur by the action of the predicted O-methyltransferases Clx2 and Clx43. In our proposed biosynthesis, the resulting hexacyclic xanthone aglycone is appended with a D-quinovose sugar moiety by the predicted glycosyltransferase Clx40.[2f]

The Arenimycins. BAC-AB1442/1414/561 harbors the 40 kb Arn (arenimycin) gene cluster (GenBank KJ440489) containing genes predicted to be involved in the biosynthesis, regulation and resistance of arenimycins C (**2**) and D (**3**). In our proposed biosynthetic scheme the benzo[*a*]naphthacene core is synthesized via a min-PKS (Arn31, 32 and 36), an aromatase (Arn21), two cyclases (Arn19 and 20) and a ketoreductase (Arn33) (Figure 4). On the basis of their high sequence identity to pdmH from the pradimicin gene cluster,[6a] Arn 22 and 37 are predicted to oxidize C-6 and C-10 to form the two 1,4-benzoquinone moieties (A and C rings) seen in **2** and **3**. Arn 17, which shows 54 and 43% sequence identity to GrhO8 from the griseorhodin gene cluster and XanO5 from the xantholipin gene cluster, respectively, is likely to be involved in hydroxylations at the two angular positions (C-5 and C-18).[6g,h] Finally, this aglycone is predicted to be glycosylated by two glycosyltransferases Arn11 and Arn14, completing the biosynthesis of **2** and **3**. The mechanism of *N*-glycosidic bond formation is not clear at this point.

The Arn gene cluster harbors all genes predicted to be required for the biosynthesis of the sugar moieties (rhamnose and olivose) found in **2** and **3** (Figure 4). Interestingly, it also contains three additional predicted sugar biosynthesis genes, whose functions could not be assigned to a specific transformation in the biosynthesis of the sugar moieties found on **2** or **3**. These include genes predicted to encode for an *N*,*N*-dimethyltransferase (Arn6), an aminotransferase (Arn7) and a 3,4-dehydratase (Arn8). These observations along with the recently reported structure of arenimycin B,[9b] which contains the dimethylated amino deoxysugar forosamine in place of the *O*Me-L-olivose seen in **2**, led us to re-examine culture broth extracts obtained from BAC-AB1442/1414/561 by LC−MS. The selective ion chromatogram for m/z = 809 revealed the presence of a minor clone specific compound with a mass corresponding to that of

dx.doi.org/10.1021/ja510606j | J. Am. Chem. Soc. 2014, 136, 18111−18119

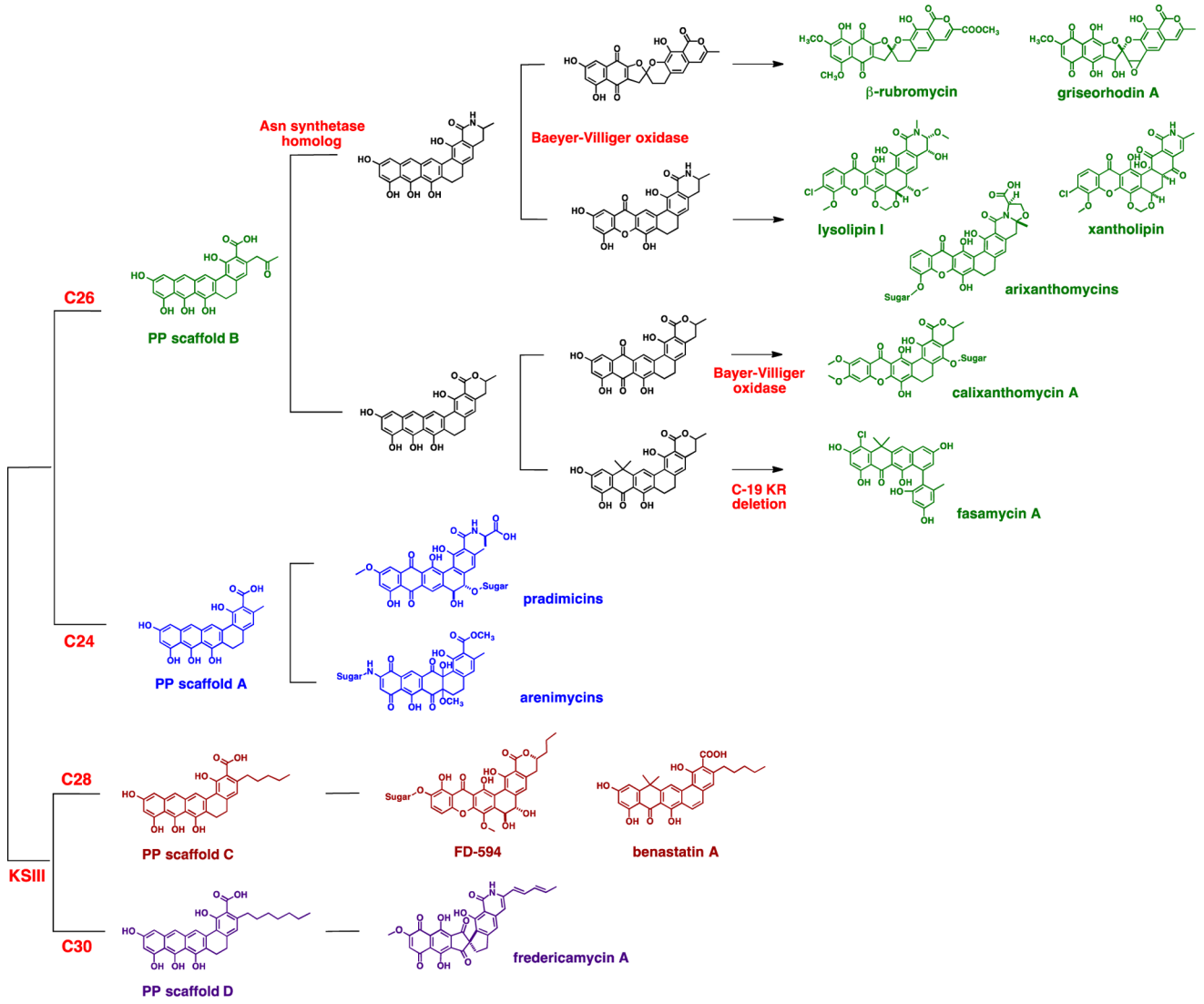


Figure 5. Different PP chemistries are organized according to the phylogenetic analysis of KS_β gene sequences (Figure 1). This analysis reveals four different PP scaffolds with distinct polyketide chain lengths. We have designated these scaffolds PP-A, PP-B, PP-C and PP-D. Each different PP scaffold and the compounds arising form the scaffold are shown in different colors. Structures shown in black are hypothetical intermediates. The structures of sugar moieties were not shown due to the limited space.

arenimycin B (Figure S9). It appears that the eDNA-derived Arn gene cluster has the potential to encode a number of different glycosylated compounds, with arenimycin C (**2**) being the major product. The annotated genome of *Salinispora arenicola* CNB527, the arenimycin B producer, was recently made publicly available (GenBank: NZ_AZXI01000002). A comparison of the arenimycin B and the Arn gene clusters reveals that these two gene clusters are very closely related, showing 90 to 95% sequence identity between most biosynthetic genes.

PP Gene Clusters. In total, 12 PP biosynthetic gene clusters have now been sequenced and functionally characterized, including eight clusters from culture-based studies and four from culture-independent studies.[2c,f,6] In addition to the eDNA-derived Clx and Arn clusters described here, we previously reported eDNA gene clusters that encode for fasamycin and arixanthomycin type PPs.[2c,f] KS_α and KS_β sequence tags associated with these four eDNA clusters were selected by us for detailed analyses because they all fall between well characterized KS clades (i.e., intermediate sequence tags). In the following analyses, we explore this closely related collection of biosynthetic gene clusters to gain insights into the evolution of natural product structural diversity.

Four Distinct Scaffolds Are Generated by PP Associated Min-PKSs. The individual proteins that makeup min-PKS (KS_α, KS_β, and ACP) are highly conserved across type II PKS gene clusters. KS_α and KS_β phylogenetic trees show very similar topologies, both of which correlate closely with differences in the core polyketide structure (e.g., chain length and cyclization pattern) encoded by the gene cluster from which the min-PKS arises.[11] KS_β and, to a lesser extent, KS_α genes have thus proved to be useful phylogenetic markers for predicting differences in polyketide core structures encoded by type II PKS gene clusters.[12]

While PPs have so far been considered a single class of type II polyketides, based on both KS phylogeny and natural product structure, it appears that they arise from four distinct min-PKS lineages. These lineages differ in starter unit selectivity (acetate or hexanoate) and the number of malonyl CoA chain extension steps they carry out (11 or 12). PP polyketide precursors range from 24 to 30 carbons in length and generate four unique PP

dx.doi.org/10.1021/ja510606j | J. Am. Chem. Soc. 2014, 136, 18111−18119

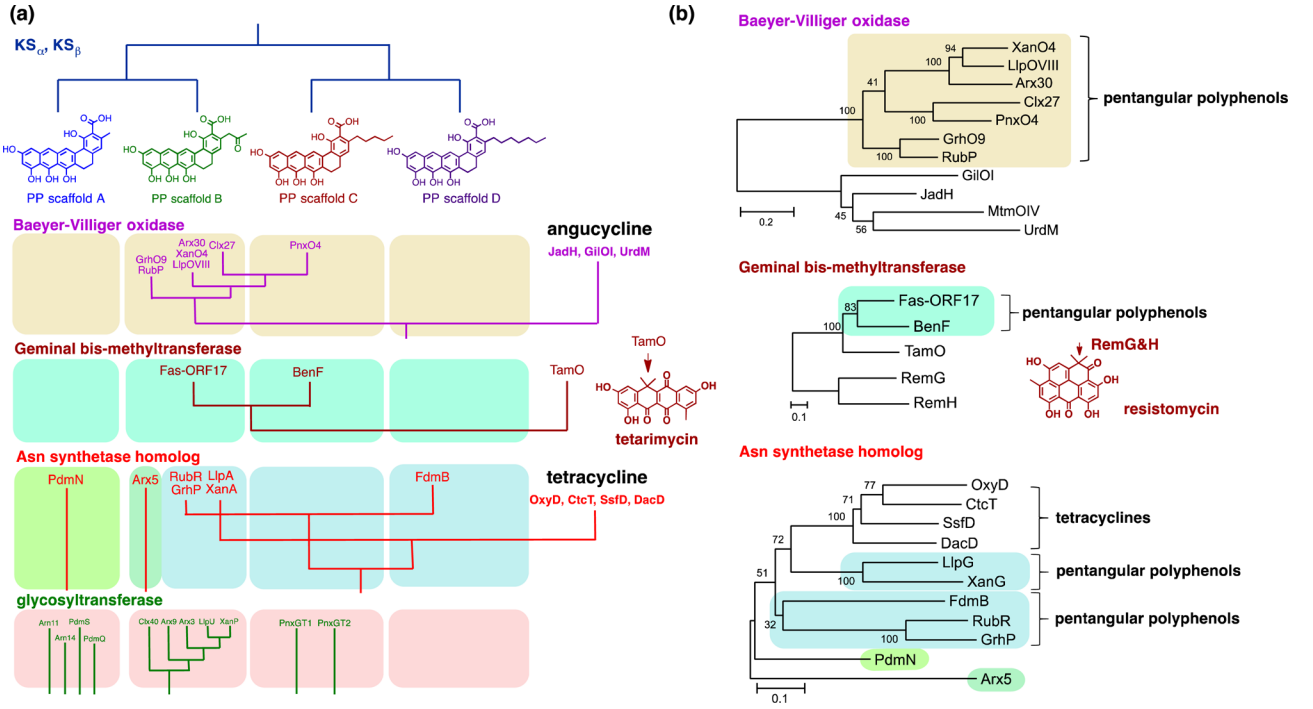


Figure 6. Overview of PP biosynthesis gene phylogenies. (a) Phylogenetic relationships between KS$_\alpha$, KS$_\beta$, Baeyer–Villiger oxidase, Asn synthase homologue, geminal bis-methyltransferase and glycosyltransferase genes from PP biosynthetic gene clusters are reconstructed from the maximum likelihood phylogenetic trees to highlight the relationships between genes from clusters of different PP lineages. (b) The parent maximum likelihood phylogenetic trees for BVO genes, GBM genes and Asn synthase homologue genes are shown.

scaffolds that we have designated PP-A (C24–acetate and 11 extensions), PP-B (C26–acetate and 12 extensions), PP-C (C28–hexanoate and 11 extensions) and PP-D (C30–hexanoate and 12 extensions) (Figure 5). The 30-carbon D scaffold is the largest polyketide chain observed in aromatic polyketide biosynthesis. With more than 30 predicted B scaffold-based natural products reported in the literature, this 26-carbon scaffold, which arises from an acetate starter unit and 12 malonyl CoA extension steps is responsible for most of the structural diversity seen in known PPs (Figure 5). The only reported gene clusters predicted to encode the A, C, and D scaffolds are the pradimicins[13] (PP-A), arenimycins[14] (PP-A), benastatins[15] (PP-C), FD-594s[16] (PP-C) and fredericamycins[17] (PP-D).

Tailoring Genes. Although both KS$_\alpha$ and KS$_\beta$ gene phylogenies indicate that the four PP scaffolds (A–C) have evolved independently from a common ancestor, similar tailoring modifications are observed across scaffolds from different lineages. This is especially interesting in light of the fact that some of modifications seen in multiple different PP lineages (e.g., the xanthone core and the gem-dimethyl functionality) are in fact quite rare outside of PP biosynthesis. This suggests that functionally successful horizontal transfer of the tailoring genes responsible for these modifications has not occurred globally throughout bacterial secondary metabolism but instead "locally" within the limited universe of PP biosynthesis. To evaluate evolutionary relationships between the genes responsible for the common modifications found across different PP scaffolds, we carried out detailed phylogenetic analyses of the four common tailoring genes seen in PP gene clusters including Baeyer–Villiger oxidase, geminal bis-methyltransferase, Asn-synthetase homologue and glycosyltransferase (Figure 6).

Tailoring Gene Families Found Exclusively in Secondary Metabolism. BVO Genes. The most notable tailoring modification observed in PP biosynthesis is an oxidative rearrangement of the PP scaffold to generate either a xanthone or a spiroketal. These oxidative rearrangements involve the action of FAD-dependent Baeyer–Villiger oxidase (BVO). BVOs are also found in some angucycline gene clusters (GilOI, JadH and UrdM)[18] and in the mithramycin gene cluster (MtmOIV).[19] To determine the phylogenetic relationship between these genes, a maximum likelihood phylogenetic tree was generated using full-length BVO gene sequences (Figure 6). In this analysis, the PP BVO genes form a monophyletic clade that is distinct from other BVO genes, suggesting that the PP BVO gene was acquired once in PP biosynthesis and diverged to enable the production of distinct xanthone and spiroketal functionalities.

Interestingly, the BVO genes Clx27 and PnxO4 from the calixanthomycin A and FD-594 gene clusters, respectively, show the highest sequence identity even though their KS$_\beta$ genes belong to distinct PP lineages (Figure 6). This suggests that once the BVO gene successfully entered the PP biosynthetic universe it was not only passed down vertically through the PP-B lineage, but also horizontally transferred between PP lineages. Curiously, while it appears that this BVO gene was capable of horizontal transfer between PP lineages with related core structures, similar, more global transfer to numerous other sequenced type II PKS gene cluster families is not observed, as the BVO genes found in other type II PKS gene clusters appear to be of distinct phylogenetic lineages.

GBM Gene. Geminal bis-methylation is another rare tailoring reaction that is seen in PP biosynthesis. The introduction of this functional group is known to occur through the action of a single, SAM-dependent methyltransferase (geminal bis-methyltransfer-

dx.doi.org/10.1021/ja510606j | J. Am. Chem. Soc. 2014, 136, 18111–18119

ase, GBM) via two rounds of C-methylation.[20] Known aromatic polyketides with this functionality include two PPs, the fasamycins and the benastatins,[2c,21] a pentacyclic polyketide resistomycin,[22] and the tetracyclic quinone tetarimycin.[23] In resistomycin biosynthesis, two methyl transferases are predicted to be required to introduce the dimethyl functionality, neither of which is closely related to the PP GBMs. The three tetracyclic structures upon which the PP-like GBMs act (fasamycins, benastatins and tetarimycins) are closely related (Figure 6), once again suggesting that horizontal transfer of these tailoring genes has been limited to interchange between gene clusters that encode closely related core structures.

For both *BVO* and *GBM* tailoring genes, it appears that functionally successful horizontal gene transfer has been limited to gene clusters that encode metabolites with closely related core chemical structures. This would explain their presence in different PP lineages and few other gene clusters outside of PP biosynthesis and suggest that for many tailoring enzymes the challenge of changing substrate specificity (or limited substrate promiscuity) may have restricted their more extensive spread throughout natural product biosynthesis by horizontal gene transfer mechanisms.

Tailoring Gene Families That Are Common Across the Global Metagenome. Transamination Genes. The terminal carboxylate in PP structures undergoes a variety of reactions including O-methylation, esterification/cyclization, alanine addition, serine addition/cyclization and amidation/cyclization. Among these, amidation in general and, more specifically, transamination carried out by Asn synthetase homologues is the most commonly seen transformation. In the phylogenetic analysis of PP Asn transaminase genes, we once again see what appears to be horizontal transfer between different PP lineages. We also see what appears to be horizontal transfer between different polyketide classes as the closest relatives of the LlpA and XanA genes from the lysolipin and xantholipin gene clusters, respectively, are found in tetracycline-type gene clusters.

Glycosyltransferase Genes. One of the most common tailoring steps found in aromatic polyketide biosynthesis is glycosylation. To investigate relationships between PP glycosyltransferases, a phylogenetic tree was constructed using 40 glycosyltransferase genes found in sequenced PP gene clusters and additional glycosyltransferase genes found in gene clusters that encode a diverse collection of non-PP aromatic polyketides (Figure S1). This maximum likelihood phylogenetic tree shows that clades do not correlate strongly with aromatic polyketide aglycone substrate structures or donor sugar types. Among PP glycosyltransferse genes, Clx40, Arx9, Arx3, LlpU and XanP do form a monophyletic clade, suggesting that they have been vertically transferred through PP scaffold B gene cluster evolution. Other PP glycosyltransferases, including Arn11, Arn14, PnxGT1, PnxGT2, PdmS and PdmQ, do not clade based on their respective PP min-PKS lineages, suggesting that they have likely entered the PP biosynthetic gene clusters in distinct horizontal gene transfer events.

For tailoring genes that to date are exclusively found in secondary metabolism (e.g., BVO and GBM), we see limited introduction of these genes into PP biosynthesis. Within the examined set of gene clusters, both BVO and GBM genes appear to have entered PP biosynthesis only once. However, for tailoring gene families that are common across the global metagenome (e.g., Asn-synthetase homologue and glycosyltransferase genes), we see many more "introduction events" resulting in their more varied use in PP structural diversification. The more frequent introduction of Asn-synthetase homologue and glycosyltansferase genes into PP biosynthesis is likely due to both their more frequent appearance in the environment and their more relaxed substrate specificity.

3. CONCLUSIONS

Soil metagenomes represent the products of nature's ongoing combinatorial biosynthetic efforts. Here, we show that sequence tag-based metagenomic screening methods provide an opportunity to identify collections of related natural product gene clusters that not only encode novel metabolites, but can also provide insights into the natural combinatorial biosynthetic process. More specifically, we show that the functional characterization of gene clusters whose sequence tags fall between clades associated with two related but distinct chemotypes ("intermediate" sequence tags) can be used to increase the chemical and the biosynthetic diversity associated with a targeted class of natural products.

On the basis of the differences in min-PKS gene phylogeny and the PP chemical structures we show that pentangular polyphenols actually arise from four distinct polyketide lineages. Our phylogenetic analysis of PP tailoring genes indicates that the horizontal transfer of at least a subset of PP tailoring genes has likely been restricted to gene clusters that encode closely related chemical structures. If true, this would suggest that nature has sampled only a fraction of the "natural product-like" chemical space that can theoretically be encoded by secondary metabolite tailoring genes. These observations provide a more detailed picture of the nature's combinatorial biosynthetic process, which can help guide future laboratory-based combinatorial biosynthetic efforts and help ensure the construction of a set of molecules that is orthogonal to those produced naturally. In particular, our study suggests that combinatorial biosynthetic experiments focused on widely distributed enzymes (e.g., glyocsyl- or amino-transferases) are likely to yield metabolites already sampled by evolution. On the other hand, experiments focused on increasing the substrate promiscuity of tailoring genes that are unique to secondary metabolism (e.g., BVO or GBM) would be much more likely to yield metabolites not yet encoded by natural biosynthetic pathways.

4. MATERIALS AND METHODS

General Experimental Procedures. All PCR was performed using BioRad Tetrad thermocycler. All recovered cosmids and TAR constructs were sequenced using IonTorrent Personal Genome Machine (PGM). Optical rotations of isolated compounds were measured on a JASCO P-1020 polarimeter and IR spectra were recorded on a Bruker TENSOR27 IR spectrometer. All NMR data used for structural characterization were obtained on a Bruker Avance DMX 600 MHz NMR spectrometer equipped with a cryoprobe. ^{1}H and ^{13}C NMR chemical shifts were referenced to the DMSO-d_6 solvent signals (δ_H 2.50 and δ_C 39.51, respectively). HRESIMS data was acquired on the LCT Premier time-of-flight mass spectrometer.

Library Screening and Phylogenetic Analysis of KS_α and KS_β Amplicons. Our AB (California), AZ (Arizona) and TX (Texas) desert soil eDNA libraries were used for screening with KS_α and KS_β degenerate primers.[12b,24] Each library contains $>10^6$ clones, which is predicted to approach saturation of the genetic diversity present in soil microbiomes. These libraries are arrayed as 5000 membered subpools to facilitate screening and gene cluster recovery studies.[25] DNA aliquots from the 5000 membered subpools found in each arrayed library were screened using degenerate primers designed to amplify partial KS_α and full-length KS_β genes. Each 25 μL PCR reaction contained 50 ng of cosmid DNA, 2.5 μM of each primer, 2 mM dNTPs, 1X ThermoPol reaction buffer (New England Biolabs), 0.5 units Taq DNA polymerase

and 5% DMSO. PCR was conducted using the following touchdown protocol: denaturation (95 °C, 2 min), 8 touchdown cycles [95 °C, 45 s; 65 °C (−1 °C per cycle), 1 min; 72 °C, 1 min], 35 standard cycles (95 °C, 45 s; 58 °C, 1 min; 72 °C, 1 min) and a final extension step (72 °C, 2 min). The resulting PCR amplicons were gel purified and Sanger-sequenced. Approximately 500 bp of KS_α or 600bp of KS_β gene (corresponding to nucleotides 562−1061 of *benA* or 25−621 of *benB*) fragments from each amplicon were aligned using ClustalW. Phylogenetic analysis and pairwise distant calculation were performed using MEGA5.1.[26]

Gene Cluster Recovery and Bioinformatics Analysis. The eDNA clones from which the AB1692 and AB1414 amplicon sequences were amplified were recovered from the AB1692 and AB1414 subpools using serial dilution method. Briefly, overnight cultures of the subpools were inoculated into 96-well microtiter plates at a dilution of 50 cells/well. After overnight shaking at 37 °C, the diluted cultures were screened by whole-cell PCR using specific primers designed to recognize the AB1692 and AB1414 amplicon sequences. The PCR positive wells were replated at 5 cells/well and rescreened. PCR-positive wells from the second round were plated onto LB agar media to give distinct colonies that were screened individually using colony PCR to identify the clones AB1692 and AB1414. The AB library was screened for overlapping eDNA clones using clone-specific primers designed to recognize the sequence at each end of the eDNA clones AB1692 and AB1414. Hits were identified in subpools AB916 for the AB1692 clone, and AB1442 and AB561 for the AB1414 clone. The overlapping eDNA clones were recovered from these subpools using the serial dilution method outlined above. All three cosmids were PGM-sequenced and annotated using MetaGeneMark,[27] in combination with BLAST search and Pfam domain analysis. Annotation results indicated that the biosynthetic gene cluster associated with the AB1414 KS_β sequence was completely captured in three overlapping clones, AB1442, Ab1414 and AB561, but the gene cluster associated with AB1692 KS_β sequence was not complete in the AB916, therefore the additional overlapping clone was recovered using the same method described above. Sequencing of the additional overlapping clone AB170 suggested the complete recovery of the AB1692 KS_β sequence-associated PP gene cluster over three overlapping eDNA clones, AB1692, AB916 and AB170.

TAR Assembly of Three Overlapping eDNA Clones. The three overlapping eDNA cosmid clones (AB1692, AB916 and AB170; AB1442, AB1414 and AB561) that are predicted to contain the AB1692 and AB1414 amplicons-associated pentangular polyphenol gene clusters were assembled into BAC clones in yeast using TAR. The AB1692 and AB1414 pentangular polyphenol pathway-specific TAR capture vectors were constructed using InFusion cloning methodology (Clontech). Five hundred bp upstream (UPS) and downstream (DWS) homology arms were amplified from the AB1692 and AB170 cosmids, and the AB1442 and AB561 cosmids, respectively, using the primers listed in the supplementary protocol 1. Gel-purified amplicons (Qiagen) and BmtI/SphI linearized pTARa vector were added into a standard InFusion cloning reaction (Clontech) to yield the pathway-specific capture vectors.

For TAR assembly, 200 ng of each DraI-cut cosmid (AB1692, AB916 and AB170; AB1442, AB1414 and AB561) were mixed with 100 ng of the *Hpa*I-digested pathway-specific capture vector and then transformed into 200 μL of *Saccharomyces cerevisiae* CRY1−2 spheroplasts prepared according to published protocols.[7a] Transformed spheroplasts were mixed with synthetic complete (SC) top agar (1 M sorbitol, 1.92 g/L synthetic complete uracil dropout supplement, 6.7 g/L yeast nitrogen base, 2% glucose and 2.5% agar) and overlaid onto SC dropout plates without uracil. Plates were incubated at 30 °C for approximately 72 h, until colonies appeared. Twelve yeast colonies were picked and cultured overnight in SC dropout liquid media without uracil. DNA was isolated using a zymolyase lysis protocol (ZYMO RESEARCH) and screened by PCR with primers designed to recognize sequences from each cosmid. A bacterial artificial chromosome (BAC) clones that amplified with all the primer sets were electroporated into *E. coli* EPI300 (Epicenter). DNA isolated from the resulting *E. coli* EPI300 was PGM-sequenced to confirm the correct reassembly of the three overlapping cosmids into approximately 90 kb insert BAC clones (BAC-1692/916/170 and BAC-AB1442/1414/561).

Heterologous Expression of Arenimycins C (1) and D (2). BAC-AB1692/916/170 and BAC-AB1442/1414/561 were transformed into *E. coli* S17.1 and transferred into *S. albus* via intergenic conjugation[28] to generate the strains *S. albus* BAC-AB1692/916/170 and *S. albus* BAC-AB1442/1414/561 (apramycin selection). Calixanthomycin A (**1**), and arenimycins C (**2**) and D (**3**) were isolated from 1L cultures (200 rpm at 30 °C) of *S. albus* BAC-AB1692/916/170 and *S. albus* BAC-AB1442/1414/561, respectively, in 125 mL baffled flasks, each containing 50 mL of R5A[29] media. The cultures were harvested after 9 days, extracted with ethyl acetate [1:3 (v:v), ethyl acetate to the culture] and dried in vacuo. Calixanthomycin A (**1**), and arenimycins C (**2**) and D (**3**) were isolated from the resulting ethyl acetate extracts using two rounds of reversed-phase HPLC (C_{18} column, 10 mm × 250 mm, 3.5 mL/min). The first round of HPLC (isocratic, 60% acetonitrile for **1**; 40% acetonitrile for **2** and **3**) yielded crude samples, from which **1**−**3** were purified. The second rounds of HPLC used 60% aqueous methanol for calixanthomycin A (**1**, 6.0 mg/L), and 70 and 60% aqueous methanol with 0.1% trifluoroacetic acid for arenimycins C (**2**, 4.6 mg/L) and D (**3**, 2.5 mg/L), respectively.

*Calixanthomycin A (**1**).* Yellow brownish powder; $[\alpha]^{20}_D$ 68; UV (MeOH) λ_{max} 226, 295, 329 nm; IR (neat) ν_{max} 3393, 2927, 2736, 1679, 1604, 1557 cm^{-1}; ^1H and ^{13}C NMR, COSY, and HMBC data, see Table S1; HRESIMS *m*/*z* 695.2327 $[M + H]^+$ (calcd for $C_{36}H_{39}O_{14}$, 695.2340).

*Arenimycin C (**2**).* Dark reddish powder; $[\alpha]^{20}_D$ −77; UV (MeOH) λ_{max} 216, 256, 421, 477 nm; IR (neat) ν_{max} 3384, 3070, 2929, 2856, 1682, 1620, 1511 cm^{-1}; ^1H and ^{13}C NMR, COSY, and HMBC data, see Table S2; HRESIMS *m*/*z* 812.2770 $[M + H]^+$ (calcd for $C_{40}H_{46}NO_{17}$, 812.2766).

*Arenimycin D (**3**).* Dark reddish powder; $[\alpha]^{20}_D$ −133; UV (MeOH) λ_{max} 215, 254, 416, 474 nm; IR (neat) ν_{max} 3392, 3072, 3050, 2931, 2858, 2738, 1686, 1641, 1619 cm^{-1}; ^1H and ^{13}C NMR data, see Table S2; HRESIMS *m*/*z* 654.1826 $[M + H]^+$ (calcd for $C_{32}H_{32}NO_{14}$, 654.1823).

Antibacterial Assay. Minimal inhibitory concentrations (MIC) for calixanthomycin A (**1**), and arenimycins C (**2**) and D (**3**) against *Staphylococcus aureus* USA-300 (MRSA), *Bacillus subtilis* RM125, *Enterococcus faecalis* (VRE) and *Escherichia coli* DRC39 were determined by liquid dilution methods.[30] All assays were performed in duplicate. Each bacterial strain was grown overnight in either brain heart infusion (MRSA and VRE) or LB (*B. subtilis* RM125 and *E. coli* DRC39) liquid media. Overnight cultures were diluted 1000-fold and transferred to 96-well microtiter plates (75 μL/well). Compounds were dissolved in DMSO, added to the media in the first well of row to give a concentration of 50 μg/mL and serially diluted 2-fold/well across the plate. These solutions (75 μL) were then added to the corresponding well in an assay plate and incubated at 30 °C overnight without shaking. Bacterial growth was visually inspected and the last well with no bacterial growth was reported as the MIC.

Antiproliferative Assay. The antiproliferative activity of calixanthomycin A (**1**), and arenimycins C (**2**) and D (**3**) was evaluated in duplicate using the colon carcinoma cell line HCT-116 (ATCC; CCL-247).[31] HCT-116 cells were grown in McCoy's 5A Modified Medium (Gibco) supplemented with 10% (v/v) FBS. Cells in log phase growth were harvested by trypsinization. Trypsinized cells were seeded into 96-well plates (1000 cells/well) and incubated overnight at 37 °C in the presence of 5% CO_2. Compounds **1**−**3** (in DMSO) were sequentially diluted in culture media (2 or 3-fold dilutions starting at 50 μg/mL) across a 96-well plate and 100 μL was transferred to the appropriate well in an assay plate. The plates were incubated at 37 °C for 3 days and then evaluated for viability using a crystal violet-based colorimetric assay.[32] Cell viability was recorded based on the A_{590} of stain present in each well relative to no drug DMSO control wells.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The phylogenetic tree of glycosyltransferase genes found in diverse type II PKS biosynthetic gene clusters; the structures of

structurally related known compounds; structure determination of compounds **1**−**3**; NMR tables for compounds **1**−**3**; gene annotation tables for the *Clx* and *Arn* gene clusters; primers used for library screening and TAR cloning; the 1D and 2D NMR spectra of compounds **1**−**3**. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

sbrady@rockefeller.edu

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by NIH GM077516. S.F.B. is an HHMI Early Career Scientist. We thank Alex Milshteyn for valuable comments on the manuscript.

REFERENCES

(1) (a) Yamanaka, K.; Reynolds, K. A.; Kersten, R. D.; Ryan, K. S.; Gonzalez, D. J.; Nizet, V.; Dorrestein, P. C.; Moore, B. S. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 1957−1962. (b) Owen, J. G.; Reddy, B. V. B.; Ternei, M. A.; Charlop-Powers, Z.; Calle, P. Y.; Kim, J. H.; Brady, S. F. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 11797−11802. (c) Biggins, J. B.; Kang, H.-S.; Ternei, M. A.; DeShazer, D.; Brady, S. F. *J. Am. Chem. Soc.* **2014**, *136*, 9484−9490.

(2) (a) Banik, J. J.; Brady, S. F. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 17273−17277. (b) King, R. W.; Bauer, J. D.; Brady, S. F. *Angew. Chem., Int. Ed.* **2009**, *48*, 6257−6261. (c) Feng, Z.; Kallifidas, D.; Brady, S. F. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 12629−12634. (d) Kang, H. S.; Brady, S. F. *Angew. Chem., Int. Ed.* **2013**, *52*, 11063−11067. (e) Chang, F. Y.; Brady, S. F. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 2478−2483. (f) Kang, H. S.; Brady, S. F. *ACS Chem. Biol.* **2014**, *9*, 1267−1272. (g) Reddy, B. V.; Milshteyn, A.; Charlop-Powers, Z.; Brady, S. F. *Chem. Biol.* **2014**, *21*, 1023−1033.

(3) (a) Hertweck, C.; Luzhetskyy, A.; Rebets, Y.; Bechthold, A. *Nat. Prod. Rep.* **2007**, *24*, 162−190. (b) Hertweck, C. *Angew. Chem., Int. Ed.* **2009**, *48*, 4688−4716.

(4) Lackner, G.; Schenk, A.; Xu, Z.; Reinhardt, K.; Yunt, Z. S.; Piel, J.; Hertweck, C. *J. Am. Chem. Soc.* **2007**, *129*, 9306−9312.

(5) Sullivan, J. *Methods Enzymol.* **2005**, *395*, 757−779.

(6) (a) Kim, B. C.; Lee, J. M.; Ahn, J. S.; Kim, B. S. *J. Microbiol. Biotechnol.* **2007**, *17*, 830−839. (b) Kudo, F.; Yonezawa, T.; Komatsubara, A.; Mizoue, K.; Eguchi, T. *J. Antibiot.* **2011**, *64*, 123−132. (c) Lopez, P.; Hornung, A.; Welzel, K.; Unsin, C.; Wohlleben, W.; Weber, T.; Pelzer, S. *Gene* **2010**, *461*, 5−14. (d) Saito, H.; Bruenker, P.; Sterner, O.; Bailey, J. E.; Wolfgang, M. *Nippon Nogei Kagakkai Taikai Koen Yoshishu* **2002**, 299. (e) Wendt-Pienkowski, E.; Huang, Y.; Zhang, J.; Li, B.; Jiang, H.; Kwon, H.; Hutchinson, C. R.; Shen, B. *J. Am. Chem. Soc.* **2005**, *127*, 16442−16452. (f) Xu, Z.; Schenk, A.; Hertweck, C. *J. Am. Chem. Soc.* **2007**, *129*, 6022−6023. (g) Zhang, W.; Wang, L.; Kong, L.; Wang, T.; Chu, Y.; Deng, Z.; You, D. *Chem. Biol.* **2012**, *19*, 422−432. (h) Li, A.; Piel, J. *Chem. Biol.* **2002**, *9*, 1017−1026.

(7) (a) Kallifidas, D.; Brady, S. F. *Methods Enzymol.* **2012**, *517*, 225−239. (b) Kim, J. H.; Feng, Z.; Bauer, J. D.; Kallifidas, D.; Calle, P. Y.; Brady, S. F. *Biopolymers* **2010**, *93*, 833−844.

(8) (a) Malet-Cascon, L.; Romero, F.; Espliego-Vazquez, F.; Gravalos, D.; Fernandez-Puentes, J. L. *J. Antibiot.* **2003**, *56*, 219−225. (b) Rodriguez, J. C.; Fernandez Puentes, J. L.; Baz, J. P.; Canedo, L. M. *J. Antibiot.* **2003**, *56*, 318−321.

(9) (a) Gomi, S.; Sasaki, T.; Itoh, J.; Sezaki, M. *J. Antibiot.* **1988**, *41*, 425−432. (b) Kersten, R. D.; Ziemert, N.; Gonzalez, D. J.; Duggan, B. M.; Nizet, V.; Dorrestein, P. C.; Moore, B. S. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, E4407−E4416. (c) Takeda, U.; Okada, T.; Takagi, M.; Gomi, S.; Itoh, J.; Sezaki, M.; Ito, M.; Miyadoh, S.; Shomura, T. *J. Antibiot.* **1988**, *41*, 417−424.

(10) (a) Kunimoto, S.; Lu, J.; Esumi, H.; Yamazaki, Y.; Kinoshita, N.; Honma, Y.; Hamada, M.; Ohsono, M.; Ishizuka, M.; Takeuchi, T. *J. Antibiot.* **2003**, *56*, 1004−1011. (b) Nakagawa, A.; Omura, S.; Kushida, K.; Shimizu, H.; Lukacs, G. *J. Antibiot.* **1987**, *40*, 301−308. (c) Omura, S.; Nakagawa, A.; Kushida, K.; Lukacs, G. *J. Am. Chem. Soc.* **1986**, *108*, 6088−6089. (d) Ratnayake, R.; Lacey, E.; Tennant, S.; Gill, J. H.; Capon, R. J. *Chemistry* **2007**, *13*, 1610−1619. (e) Winter, D. K.; Sloman, D. L.; Porco, J. A., Jr. *Nat. Prod. Rep.* **2013**, *30*, 382−391.

(11) Ridley, C. P.; Lee, H. Y.; Khosla, C. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 4595−4600.

(12) (a) Wawrik, B.; Kerkhof, L.; Zylstra, G. J.; Kukor, J. J. *Appl. Environ. Microbiol.* **2005**, *71*, 2232−2238. (b) Seow, K. T.; Meurer, G.; Gerlitz, M.; Wendt-Pienkowski, E.; Hutchinson, C. R.; Davies, J. *J. Bacteriol.* **1997**, *179*, 7360−7368.

(13) Oki, T.; Konishi, M.; Tomatsu, K.; Tomita, K.; Saitoh, K.; Tsunakawa, M.; Nishio, M.; Miyaki, T.; Kawaguchi, H. *J. Antibiot.* **1988**, *41*, 1701−1704.

(14) Asolkar, R. N.; Kirkland, T. N.; Jensen, P. R.; Fenical, W. *J. Antibiot.* **2010**, *63*, 37−39.

(15) Aoyama, T.; Naganawa, H.; Muraoka, Y.; Nakamura, H.; Aoyagi, T.; Takeuchi, T.; Iitaka, Y. *J. Antibiot.* **1992**, *45*, 1391−1396.

(16) Kondo, K.; Eguchi, T.; Kakinuma, K.; Mizoue, K.; Qiao, Y. F. *J. Antibiot.* **1998**, *51*, 288−295.

(17) Misra, R.; Pandey, R. C.; Hilton, B. D.; Roller, P. P.; Silverton, J. V. *J. Antibiot.* **1987**, *40*, 786−802.

(18) (a) Fischer, C.; Lipata, F.; Rohr, J. *J. Am. Chem. Soc.* **2003**, *125*, 7818−7819. (b) Rix, U.; Wang, C.; Chen, Y.; Lipata, F. M.; Remsing Rix, L. L.; Greenwell, L. M.; Vining, L. C.; Yang, K.; Rohr, J. *ChemBioChem* **2005**, *6*, 838−845. (c) Faust, B.; Hoffmeister, D.; Weitnauer, G.; Westrich, L.; Haag, S.; Schneider, P.; Decker, H.; Kunzel, E.; Rohr, J.; Bechthold, A. *Microbiology* **2000**, *146*, 147−154.

(19) Lombo, F.; Blanco, G.; Fernandez, E.; Mendez, C.; Salas, J. A. *Gene* **1996**, *172*, 87−91.

(20) Xu, Z.; Schenk, A.; Hertweck, C. *J. Am. Chem. Soc.* **2007**, *129*, 6022−6030.

(21) Schenk, A.; Xu, Z.; Pfeiffer, C.; Steinbeck, C.; Hertweck, C. *Angew. Chem., Int. Ed.* **2007**, *46*, 7035−7038.

(22) Jakobi, K.; Hertweck, C. *J. Am. Chem. Soc.* **2004**, *126*, 2298.

(23) Kallifidas, D.; Kang, H. S.; Brady, S. F. *J. Am. Chem. Soc.* **2012**, *134*, 19552−19555.

(24) Reddy, B. V.; Kallifidas, D.; Kim, J. H.; Charlop-Powers, Z.; Feng, Z.; Brady, S. F. *Appl. Environ. Microbiol.* **2012**, *78*, 3744−3752.

(25) Brady, S. F. *Nat. Protoc.* **2007**, *2*, 1297−1305.

(26) Tamura, K.; Peterson, D.; Peterson, N.; Stecher, G.; Nei, M.; Kumar, S. *Mol. Biol. Evol.* **2011**, *28*, 2731−2739.

(27) Zhu, W.; Lomsadze, A.; Borodovsky, M. *Nucleic Acids Res.* **2010**, *38*, e132.

(28) Phornphisutthimas, S.; Thamchaipenet, A.; Panijpan, B. *Biochem. Mol. Biol. Educ.* **2007**, *35*, 440−445.

(29) Fernandez, E.; Weissbach, U.; Sanchez Reillo, C.; Brana, A. F.; Mendez, C.; Rohr, J.; Salas, J. A. *J. Bacteriol.* **1998**, *180*, 4929−4937.

(30) Li, C.; Zhang, F.; Kelly, W. L. *Mol. Biosyst.* **2011**, *7*, 82−90.

(31) Brattain, M. G.; Fine, W. D.; Khaled, F. M.; Thompson, J.; Brattain, D. E. *Cancer Res.* **1981**, *41*, 1751−1756.

(32) Zivadinovic, D.; Gametchu, B.; Watson, C. S. *Breast Cancer Res.* **2005**, *7*, R101−112.