A 3-hydroxy-3-methylglutaryl-CoA synthase-based probe for the discovery of the acyltransferase-less type I polyketide synthases

Haoyu Liang,[†, #] Lin Jiang,[†, #] Qiyun Jiang,[‡] Jie Shi,[†] Jingxi Xiang,[†] Xiaohui Yan,[†, Δ] Xiangcheng Zhu,[†, ⊥, Δ] Lixing Zhao,[▽] Ben Shen,[£, ¶, §] Yanwen Duan[*, †, ⊥, Δ] and Yong Huang[*, †, Δ]

[†]Xiangya International Academy of Translational Medicine at Central South University, Changsha, Hunan, China;

[‡]School of Geosciences and Info-physics at Central South University, Changsha, Hunan, China;

[Δ]National Engineering Research Center of Combinatorial Biosynthesis for Drug Discovery, Changsha, Hunan, China;

[⊥]Hunan Engineering Research Center of Combinatorial Biosynthesis and Natural Product Drug Discovery, Changsha, Hunan, China;

[▽]Yunnan Institute of Microbiology, Yunnan University, Kunming, Yunan, China;

[£]Departments of Chemistry and [¶] Molecular Medicine, and [§]Natural Products Library Initiative at The Scripps Research Institute, The Scripps Research Institute, Jupiter, FL, USA.

[#] These authors contributed equally.

Correspondence:
*Email: Prof. Yong Huang, jonghuang@csu.edu.cn or Prof. Yanwen Duan, ywduan66@sina.com.

ABSTRACT

Acyltransferase (AT)-less type I polyketide synthases (PKSs) produce complex natural products due to the presence of many unique tailoring enzymes. The 3-hydroxy-3-methylglutaryl coenzyme A synthases (HCSs) are responsible for *β*-alkylation of the growing polyketide intermediates in AT-less type I PKSs. In this study, we discovered a large group of HCSs, closely associated with the characterized and orphan AT-less type I PKSs through *in silico* genome mining, sequence and genome neighborhood network analysis. Using HCS-based probes, the survey of 1207 in-house strains and 18 soil samples from different geological locations revealed the vast diversity of HCS-containing AT-less type I PKSs. The presence of HCSs in many AT-less type I PKSs suggests their co-evolutionary relationship. Our study should inspire future efforts to discover new polyketide natural products from AT-less type I PKSs.

INTRODUCTION

Many pharmacologically important natural products, such as antibiotic erythromycin, immunosuppressant rapamycin and anthelmintic avermectin, are biosynthesized by type I polyketide synthases (PKSs). Acyltransferase (AT)-less type I PKSs (also known as trans-AT PKSs) belong to a rapidly emerging family of novel PKSs, which are characterized by the absence of AT domain in each PKS module. The transfer of acyl CoA building block to the acyl carrier protein (ACP) domain in the elongation step is provided by one or more ATs *in trans*. Since the discovery and establishment of the biosynthetic logic for AT-less type I PKSs in leinamycin (LNM, **1**) and pederin (**2**), the study of over 40 different polyketides have resulted in new insights into their biosynthesis beyond the canonical type I PKSs (Cheng et al., 2003; Helfrich and Piel, 2016; Piel, 2002, 2010). For example, the KSs from AT-less type I PKSs could be extremely useful for predicting the sub-structure of the corresponding genetic-coded natural products, based on their structural and evolutionary relationships (Lohman et al., 2015; Nguyen et al., 2008). In addition, 3-hydroxy-3-methylglutaryl coenzyme A synthases (HCSs) and the associated HCS cassette proteins are responsible for the β-branching in many AT-less type I PKSs-derived polyketides, including LNM, pederin, and bacillaenes (BAEs, **3** and **4**) (Calderone, 2008; Calderone et al., 2006; Calderone et al., 2007; Liu et al., 2009) (Figures 1A and S1). Interestingly, HCSs are only observed in three canonical type I PKSs for curacin, jamaicamide and cylindrocyclophane biosynthesis, all of which were from photosynthetic bacteria (Chang et al., 2004; Edwards et al., 2004; Nakamura et al., 2012). The installation of

various *β*-branches imposes an additional level of structure diversity in these natural products, critical for their conformation and individual biological activity (Haines et al., 2013).

The HCSs are responsible for generating structure diversity for polyketides, by attaching methyl, ethyl, or carboxylic acids to their *β*-carbons, in a similar manner with mevalonate biosynthesis (Calderone, 2008; Maloney et al., 2016). HCSs catalyze the Claisen condensation between acyl-*S*-ACP and the ACP-tethered polyketide to afford an HMG-*S*-ACP intermediate, followed by dehydration and/or decarboxylation by the action of enoyl-CoA hydratases (ECH1 and ECH2), resulting in the *β*-alkylated intermediates (Figure 1B). The enzymes involved in the above sequential catalysis are called HCS gene cassettes, which are integral parts of many AT-less type I PKS gene clusters.

A recent genome scale analysis suggested that AT-less type I PKSs may constitute about 38% of bacterial modular PKSs among the sequenced bacterial genomes (O'Brien et al., 2014). Our discovery of the LNM family of natural products through *in silico* and genome mining also revealed that there is huge potential for structure diversity from AT-less type I PKSs (Pan et al., 2017). Although KSs in either type I or II PKSs have been used to survey the structure diversity of PKSs in different geological locations (Charlop-Powers et al., 2015; Charlop-Powers et al., 2016), the diversity of AT-less type I PKSs in the environment, to our knowledge, has not been explored in a systematic manner. In this study, we surveyed the diversity of AT-less

4

type I PKSs containing HCSs. Our studies identified 13 HCS-containing bacterial strains from 1207 in-house bacterial strain collection. Two of them have been confirmed to contain the respective AT-less type I PKS gene clusters. The survey of 18 soil samples revealed the presence of numerous HCSs in soil, which might be associated with AT-less type I PKSs. Our study suggests the wide-spread of AT-less type I PKSs and should inspire future efforts to discover more polyketide natural products from this important family of PKSs.

RESULTS

A specific type of HCSs mainly clustered with AT-less type I PKSs

In order to understand the evolution of HCSs in AT-less type I PKSs, a total of 6,712 HCS protein sequences from GenBank (as of September 27, 2018) with > 25% sequence identity of LnmM HCS from LNM biosynthetic gene cluster (BGC) were obtained. These HCSs were clustered into 2,250 representative sequences with 90% identity cutoff using CD-HIT (Huang et al., 2010). Sequence similarity network (SSN) analysis of the 2,250 HCSs were conducted using multiple E-value cutoffs ranging from 1×10^{-10} to 1×10^{-150} (Figures 2 and S2). An E-value threshold of 1×10^{-135} showed ~70% sequence identity for HCSs, resulting in 11 major clusters. The SSN analysis revealed that most HCSs are grouped to mevalonate biosynthetic pathways, which reflects their house-keeping roles in these organisms. In contrast, cluster C contains 251 HCSs, which belong to several bacterial phyla, including Actinobacteria, Firmicutes and Proteobacteria. Cluster C was identified as HCSs for $β$-branching in polyketide biosynthesis based on the following reasons: (1) this cluster contains all of the 32 HCSs from the 29 characterized AT-less type I PKSs, as well as JamH (AAS98779.1), CurD (AAT70099.1) and CylF (ARU81120.1) from canonical type I PKSs (Figures 2B and S1); (2) 200 out of 219 orphan HCSs are clustered with the 170 putative AT-less type I PKS gene clusters, based on the co-localization and similarity of their flanking open reading frames (ORFs) with proteins from characterized AT-less type I PKSs (Tables S1-S170). Interestingly, only two HCSs, AJC59428.1 and CP001344.1, both from *Cyanothece* sp. PCC 7425, are clustered with canonical type

6

I PKSs (Tables S171 and S172), while seven HCSs may be associated with type II PKSs (Tables S173-S179). The role of the remaining 10 HCSs cannot be determined due to the incomplete genome sequencing (Tables S180-S189).

Since our mining of AT-less type I PKS using HCS probes *in silico* reveals 170 orphan AT-less type I PKSs, we performed their genome neighborhood network (GNN) analysis (Figure 3). In order to understand the evolutionary relationships of the HCS cassettes and AT-less type I PKSs, the characterized AT-less type I PKSs containing HCS cassettes were also included. Using E-value cutoff of 1×10^{-6}, the GNN analysis revealed not only the clustering of HCSs in the context of the entire AT-less type I PKSs, but the close association of the other accessory enzymes in HCS cassettes with AT-less type I PKSs, including ACPs, ECHs, ATs, ketosynthases (KSs) and acyltransferase/decarboxylases (AT/DCs). Under the current E-value cutoff, HCSs in cluster C are likely grouped, based on their respective substrates propionyl-S-ACP or acetyl-S-ACP (Figures 1B and S1) (Pan et al., 2017). In addition, phylogenetic analysis suggested that the ACPs from HCS cassettes were also clustered into propionyl-S-ACP or acetyl-S-ACP specific groups (Figure S3). Therefore, the association of these proteins with the AT-less type I PKSs might reflect their co-evolutionary relationships, which is consistent with the recent observation of strong protein-protein interaction between CurD and its donor ACP in canonical type I PKSs, as well as the unique acceptor ACP domains present within the PKS modules for $β$-branching in mupirocin AT-less type I PKS (Haines et al., 2013; Maloney et al., 2016).

Survey 1207 bacterial strains using HCS-specific probes

HCSs have been used to identify and clone the BGCs of potent antitumor compound FR901464 in *Burkholderia*, cylindrocyclophane with in vitro activities towards 20S proteasome from cyanobacteria and protein phosphatase 1 inhibitor calyculin from a metagenomic library of a marine sponge (Nakamura et al., 2012; Wakimoto et al., 2014; Zhang et al., 2011). We therefore hypothesized that HCS-based PCR probe could be used to discover AT-less type I PKS BGCs in our strain collection (Hindra et al., 2014; Yan et al., 2016; Pan et al., 2017). Because these HCSs are highly conserved, degenerate primers were designed using their conserved region in characterized AT-less type I PKSs BGCs (Figures 4 and S4). The primers were first used to clone LnmM, the HCS gene from *lnm* gene cluster (Figure S4). They were subsequently used in the PCR detection of HCS genes from 1207 bacterial strains, resulting in 13 strains with the expected 0.2- to 0.3-kb DNA fragments. These DNA fragments were sequenced and confirmed to be partial HCS genes, which were used to construct an evolutionary tree with known HCSs associated with AT-less type I PKSs using Mega 7 (Figure 4C). The HCS genes in those 13 strains are clustered into six branches, using the HCS from the mevalonate biosynthesis pathway of *Staphylococcus aureus* as an out group. Six HCS genes from CB00144, CB00158, CB01229, CB01357, CB01456, and CB01509, are in the same cluster with SnaI (CBW45745.1) and VirC (BAF50725.1) involved in the biosynthesis of streptogramin A-type antibiotics. Three HCSs from CB01882, CB01883 and CB02980 are similar to LnmM, suggesting their involvement in the biosynthesis of the recently identified leinamycin-family of natural products, in which a novel polyketide weishanmycin was

8

identified from CB01883 using an alternative probe of DUF-SH domain (Pan et al., 2017). The remaining HCSs from CB00111, CB01828 and CB02999 resemble CalT, CorE and BaeG, with sequence identities of 67%, 62% and 98%, respectively.

Because GNN analysis suggests that a sub-group of HCSs are clustered with AT-less type I PKS BGCs, strain CB02999 and CB01881 were further selected to determine if they could produce specific polyketides biosynthesized by AT-less type I PKSs or contain their BGCs. The HCS from CB02999 shares high sequence similarity with the BaeG HCS, involved in the biosynthesis of BAEs in *Bacillus subtilis* and *Bacillus amyloliquefaciens* (Chen et al., 2006). CB02999 was first determined to resemble *Bacillus amyloliquefaciens* based on its house-keeping genes 16s rRNA and *rpoB* (NR_041455.1 and CP003332.1, 99% and 99% identity). The production of BAE, dihydro-BAE, and the glycosylated BAE derivative in CB02999 were confirmed by comparing their UV spectra and high resolution mass spectrometry (HRMS) with those previously reported (Figures 4D and S5) (Butcher et al., 2007; Moldenhauer et al., 2010; Qin et al., 2014). These studies suggested that *B. amyloliquefaciens* CB02999 contains the prototype AT-less type I PKS responsible for the biosynthesis of BAEs.

In order to identify the AT-less type I PKS BGC in CB01881, shotgun genome sequencing data were generated using the Illumina HiSeq 2000. Analysis of the genome sequence of CB01881 revealed the presence of an AT-less type I PKS BGC. The encoded PKS enzymes are similar to the PKSs for the biosynthesis of calyculin (~36% identity) and a few orphan AT-less type I PKSs from *S. griseoflavus* Tu4000 (GG657758.1), *S. oceani* DSM 16646 (CP002131.1) and *S.* sp. CNS654 (KL370899.1)

9

(Figure 4E). The AT-less type I PKS in CB01881 contains multiple PKS modules without AT domains in six ORFs and four discrete ATs. Its HCS containing the previously cloned HCS gene fragment was also identified (Tables S190-191). Interestingly, four contiguous ACP di-domains were also present in the PKS modules, which were the putative sites to attach the *β*-branches (Haines et al., 2013).

The survey of 18 soil samples using HCS probes revealed the presence of diverse HCSs in the environment

The discovery of more than 200 HCSs in AT-less type I PKS BGCs *in silico* and the identification of strains containing these PKSs in our in-house strain collection, encouraged us to study the diversity of HCSs and AT-less type I PKSs in the environment. The biosynthetic potential of polyketides in environmental samples is typically surveyed using conserved KS domains in type I or type II PKSs. One limitation is that a single bacteria strain in a typical soil sample may contain dozens of KS domains, due to the presence of several PKS gene clusters and multiple PKS modules in each type I PKS. Therefore the accurate identification of KSs in soil samples would rely on the deep-sequencing of environmental DNAs, in order to reduce the false negative results (Crits-Christoph et al., 2018). Because only one or two HCSs are present in AT-less type I PKS BGCs, we envisioned the potential to survey environmental HCS genes specific for AT-less type I PKSs.

We first determined if the current HCS probes could be used to survey soil samples. Eighteen soil samples from various places in China were collected and the environmental DNAs were extracted based on the reported protocol (Figure 5)

(Charlop-Powers et al., 2015). Clear DNA fragments with the expected size of 0.2- to 0.3-kb were amplified from all of the soil samples, which were gel-purified and used for high-throughput sequencing using Illumina MiSeq platforms, along with the mixture of the 13 in-house strains containing HCSs (CB). The obtained DNA sequences were then clustered into operational taxonomic units (OTUs) at a sequence distance of 10% (> 90% sequence identity), and analyzed by BlastX. Based on the analysis of the HCS genes from the known AT-less polyketide gene clusters and the HCS genes cloned from the in-house strains, an E-value above 10^{-28} was deemed to be HCS homologues for β-alkylation (Figure 2).

Rarefaction analysis suggested that the sequence space of these HCS genes was saturated (Figure 5C). Nonmetric multidimensional scaling (NMDS) ordination analysis revealed a tremendous beta diversity for the 18 soil samples (Figure 5D). For example, 12 out 18 soils have more than 6 different types of HCSs, and there are 34 different HCS genes in soil S12. Based on these pioneering studies to survey the secondary metabolite biosynthetic genes diversity, various factors of the soil samples, such as elevation, annual precipitation, temperature, pH, latitude and longitude of sampling sites, may contribute to the high beta diversity for soil samples S3-S9 (Charlop-Powers et al., 2015; Charlop-Powers et al., 2016). Interestingly, soil samples S1-S2 and S10-S18, collected from an area of around 120-km^2 region within the Yunnan-Kweichow plateau, still possessed great beta diversity, in comparison to samples collected 1,000-km away. This was attributed to the microenvironment difference in the individual sampling sites (Charlop-Powers et al., 2016).

11

Interestingly, the HCS genes of streptogramin-A type compounds were not widely distributed, which were different from our in-house screening and previous reports (Genilloud et al., 2011). Several other HCS genes, including EZH71961.1, WP_004936929.1, WP_033101211.1, WP_037576445.1, and WP_059300318.1, are abundant in the surveyed soil samples (Figures 5E and S6). This result implied that there might be certain differences for the HCS diversity between our strain collection and the collected soil samples. The OTUs of the HCS genes in soil samples S3-S7, S9-S13 and S17-18 were significantly more than those in S1-S2, S8 and S14-S16 (Figure S7). Putative new HCSs were also identified in soil samples S4-S5 and S9-S13 (Figure 5E). The survey of the diversity of HCSs in soil samples and the presence of specific HCS genes in all the samples suggested the wide-spread of AT-less type I PKSs.

The presence of diverse AT-less type I PKS specific KS domains in soil samples

Canonical type I PKSs share high homology and mainly clade with members of their biosynthetic pathways, while the KS domains from AT-less type I PKS possess substrate specificity and fall into phylogenetic clades that correlate with their substrates. These differences have been used to dissect these two types of PKSs (Piel et al., 2004; Jenke-Kodama et al., 2005; Nguyen et al., 2008; Lohman et al., 2015). In order to verify the presence and the diversity of AT-less type I PKSs in the soil samples, we next determined the presence of KSs belonging to AT-less type I PKSs, using 11 out of the previous 18 soil samples with specific KS-based PCR probes (Figure 6).

12

The expected KS gene fragments were amplified by PCR from the soil DNAs, using the previously reported primers to clone the conserved KS domains in type I PKSs (Figure S8) (Savic et al., 2006). The cloned KS DNA fragments were subjected to high-throughput sequencing, and then clustered into OTUs at a sequence distance of 10% (> 90% sequence identity), and compared to 224 KS domains from identified AT-less type I PKSs in GenBank by BlastX, using cylindrocyclophane as a control from canonical type I PKSs. The E-value ranging from 5.82 $\times$ 10^{-31} to 2.37 $\times$ 10^{-53} were deemed appropriate to assign the KS sequences from the soil DNA to each AT-less type I PKS. Our analysis indicated that there were abundant KS domains belonging to AT-less type I PKS gene clusters containing HCSs. Consistent with the presence of multiple HCSs in specific soil samples, various KS fragments belonging to AT-less type I PKSs were identified from the corresponding soils. For example, soil samples S5 and S10 contained KS fragments with high sequence similarity to more than 20 different types of AT-less type I PKSs, suggesting the huge biodiversity even in a single soil sample (Figure 6).

DISCUSSION

Microbial natural products remain one of the most important sources of new small molecule drugs, while the discovery of novel scaffolds become increasingly difficult in well-studied bacterial phyla (Newman and Cragg, 2016; Cameron et al., 2017). Polyketides produced by AT-less type I PKSs represent a growing family of natural products with complex molecular structures and promising biological activities (Piel, 2010; Helfrich and Piel, 2016). Compared to canonical type I PKSs, AT-less type I PKSs possess not only KS domains specific for various alky substrates in individual modules, but various tailoring enzymes for online modification of the growing polyketide chains, such as the imbedded methyltransferase, DUF-SH domains and HCSs (Pan et al., 2017; Xie et al., 2017). HCSs play important roles to diversify polyketide structures for modular PKSs, especially AT-less type I PKSs (Calderone et al., 2006). The current study revealed that a unique cluster of HCSs of various bacteria phyla are closely associated with characterized and orphan AT-less type I PKSs. HCS-specific primers were then used to screen 1207 microbial strains and 18 soil samples collected in representative geological locations in China, which reveals the vast biodiversity of HCS and AT-less type I PKS genes in the environment (Figures 5 and 6).

The specific KS probes have been widely used to survey the biogeographic diversity of type I PKSs and type II PKSs, which reveal different biosynthetic richness of specific soil types in various geographic locations (Charlop-Powers et al., 2015; Charlop-Powers et al., 2016). In this study, specific HCS gene-based probe was used

for the first time to survey the biogeographic distribution of HCSs and the AT-less type I PKSs. One unique feature of the HCS survey is that HCS is only responsible for the β-alkylation of the growing polyketide intermediates on ACPs, while distantly related to the overall polyketide structures, which could lead to discover new polyketides with diverse molecular scaffolds. In addition, there are only limited HCSs present in one bacterial strain, compared to dozens of KSs or KS domains in many bacteria (Crits-Christoph et al., 2018). Together with the KS survey in the 11 soil samples and BlastX analysis with KS domains from AT-less type I PKSs, our study revealed that soil samples are rich in HCSs and AT-less type I PKSs. For example, 57 unique HCS genes were distributed in 18 different soil samples, while one soil sample may contain dozens of HCS genes. This is in line with the recent survey of the biosynthetic potential in soils and therefore should be useful for the discovery of new polyketides (O'Brien et al., 2014).

The HCSs in AT-less type I PKSs may be acquired through horizontal gene transfer. Based on the phylogenetic and comparative genome analyses, the genes for mevalonate biosynthesis in gram-positive cocci were proposed to be horizontally transferred from a primitive eukaryotic cell (Wilding et al., 2000). Our results suggested that HCSs in "HCS cassette" would likely be recruited to AT-less type I PKSs from the mevalonate pathway, in a later evolution point. After the HCSs were incorporated into PKS BGCs, HCSs may further co-evolve with type I PKSs to facilitate the β-alkylation, with the presence of the acceptor ACPs containing the conserved Tyr residue in the PKS module (Haines et al., 2013). For example, the LnmKLM HCS cassette proteins

15

are not only co-transcribed with the LNM PKSs in LNM biosynthesis, inactivation of either of the *lnmK*, *lnmL* or *lnmM* generated a series of shunt metabolites from the premature release of early biosynthetic intermediates, probably due to disruption of specific protein-protein interaction between HCS cassette proteins with LNM PKSs (Huang et al., 2011). There are several other discrete proteins in HCS cassette, including ACPs, ECHs and AT/DCs or ATs. How these proteins were gradually incorporated into the AT-less type I PKS remains a fascinating evolutionary question.

The AT-less type I PKSs have been proposed to evolve independently from canonical type I PKSs mainly because those KS domains only form certain clades based on their individual acyl-*S*-ACP substrates, while KSs from canonical type I PKSs form biosynthetic pathway-specific clades. Recent structural and evolutionary relationships study of KSs in AT-less type I PKSs captured several of their evolutionary traits from canonical type I PKSs, including the presence and transformation of KS-AT didomain OzmQ-KS1 towards a KS domain in AT-less type I PKSs for oxazolomycin biosynthesis (Lohman et al., 2015). Interestingly, only three HCS cassettes are present in known canonical type I PKSs, including curacin, jamaicamide and cylindrocyclophane, while two HCS cassettes are identified in orphan canonical type I PKSs (Tables S171 and S172). All of the five HCS cassettes for canonical type I PKSs are from cyanobacteria. In contrast, there are 32 HCSs in the characterized AT-less type I PKSs, and 217 orphan HCSs for AT-less type I PKSs. The preference of *β*-alkylation in AT-less PKSs might attribute to the presence of a standalone AT responsible for priming the ACP with the malonyl-CoA, while canonical type I PKSs

typically lack this activity. In addition, *β*-alkylation in AT-less type I PKSs greatly expands their structural diversity and rigidity, due to the malonyl-CoA specific AT activity in most AT-less type I PKSs, comparing to the diverse acyl-CoA substrates of ATs in canonical type I PKSs (Chan et al., 2006). Interestingly, a few HCSs associated with orphan type II PKS gene clusters were also identified and their role in morphing aromatic polyketides remains to be deciphered.

MATERIALS AND METHODS

Bacterial strains, culture conditions and general methods

All chemical and biological reagents, including antibiotics, inorganic salts, carbon and nitrogen sources used in this study were obtained from common commercial sources, unless otherwise specified. Tryptic soy broth (TSB) with 0.2% glycine were used for mycelium growth, containing (per liter): glycine 2 g, tryptone 17 g, soy peptone 3 g, NaCl 5 g, K_2HPO_4 2.5 g, dextrose 2.5 g, pH 7.3 ± 0.2. *S.* sp. CB01881 was grown at 30 °C on R2A agar for sporulation. *B. amyloliquefaciens* CB02999 was cultured in 50 mL TSB in 250-mL shaking flask and grown for 24 h at 30 °C and 220 rpm. Then 5 mL of seed culture was inoculated to Landy medium, containing (per liter): glucose 4 g, yeast extract 4 g, malt extract 10 g, and 0.5 ml of 1 M NaOH. It was then cultured for 3 days at 30 °C and 220 rpm.

BAE and its analogues were analyzed on a Waters Ultra Performance Liquid Chromatography (UPLC) system equipped with a photo-diode array (PDA) detector, using a C-18 column (2.7 μm, 4.6 mm × 50 mm, Waters). The mobile phase consisted of buffer A (ultrapure H_2O containing 0.1 % HCOOH and 0.1 % CH_3CN) and buffer B (chromatographic grade CH_3CN containing 0.1 % HCOOH) was applied at a flow rate of 0.4 mL min^{-1}. A linear gradient program (60 % buffer A and 40 % buffer B to 50 % buffer A and 50 % buffer B for 1 min, 50 % buffer A and 50 % buffer B to 0 % buffer A and 100 % buffer B for 2 mins, followed by 60 % buffer A and 40 % buffer B for 1 min) was applied to detect BAEs at 360 nm.

18

Total DNA extraction

Our strain collection consists of strains isolated from various unexplored and underexplored ecological niches. The cultivation of bacterial strains and genomic DNA preparation followed standard protocols (Kieser et al., 2000). The concentrations of genomic DNA samples were estimated using a micro spectrophotometer (Analytic Jena).

Sequence similarity network analysis of HCS genes in GenBank

Data set source. The HCS sequences were downloaded from the NCBI GenBank protein database on September 27, 2018. LnmM (AAN85526.1) was used as a query ID for BlastP search in non-redundant protein sequences database and 6, 712 sequences with > 25% sequence identity with LnmM were obtained.

Data set curation. Because some of the 6, 712 HCS sequences are duplicates or share very high sequence similarity, it is necessary to remove those sequences for further bioinformatic analysis. Using HCS sequences from characterized AT-less type I PKSs as a training set, 90% sequence identity was considered as an appropriate threshold. If two HCSs have > 90% sequence identity, they were treated as a single input HCS seqeunce. Using this strategy, the 6,712 HCS sequences were winnowed into 2,250 representative sequences by filtering to a maximum identity of 90% using CD-HIT (Huang et al., 2010).

Construction of networks. A database containing the 2,250 representative HCS

sequences was built using formatdb command in Blast+. To construct the network, each individual sequence was used as a query to search the whole database using BlastP. The Blast *E*-value was used to define similarity between the query HCS and the hits. An *E*-value threshold of 1.0×10^{-135} was established based on the lowest sequence similarity that all of the characterized HCS sequences were grouped together. The edges formed from self-loops between each node itself, and those from duplicates between two nodes were deleted from the network. Finally, the sequence similarity network was visualized using the organic layout in Cytoscape 4.1 (Kohl et al., 2011).

High-throughput PCR screening of HCS genes from in-house strains

Design of the PCR primers. Two conserved regions of amino acids of HCSs were chosen and a pair of degenerated primers were designed according to CODEHOP strategy (Rose et al., 2003). Forward primer: 5'- CCGTTCGGCGGCATG GTCAAGGGNGCNCAYCG-3', Reverse primer: 5'- GAAGAACTCGGAGGARCANCC GGTNCCRTA-3' (R=A/G, M=A/C, N=A/T/G/C). PCR reactions were performed in 20 μl reaction mixtures consisting of 200–500 ng of genomic DNA, 10 pmol of each primer, 4 mmol of dNTP, $1 \times$ GC buffer II and 1 U of LA Taq DNA polymerase (TaKaRa). The PCR cycles consisted of 95 °C for 5 min, followed by 10 cycles of denaturation (95 °C for 0.5 min), annealing (65 °C to 55 °C for 0.5 min), and extension (72 °C for 0.3 min), and 25 cycles of denaturation (95 °C for 0.5 min), annealing (55 °C for 0.5 min), and extension (72 °C for 0.3 min), ended after a final extension (72 °C for 10 min). The

resultant PCR products were cloned and sequenced.

Phylogenetic analysis. Phylogenetic analysis was performed using MEGA 7 software. The phylogenetic tree was constructed using the neighbor-joining method and 1000 bootstrap replications.

Genome scanning to identify an AT-less type I PKS gene cluster from *S*. sp. CB01881

The shotgun genome sequence of *S*. sp. CB01881 was generated on an Illumina HiSeq 2000 instrument (BGI, Shengzhen, China). The data sets consisted of 11,420,954 single 125-bp reads. The assembly revealed 478 sequence contigs, which were combined into 41 scaffolds. Secondary metabolite biosynthetic gene cluster prediction of *S*. sp. CB01881 was performed using antiSMASH 4.0 (Blin et al., 2017). The identified AT-less type I PKS gene cluster was further analyzed using Blast. The sequencing data of *S*. sp. CB01881 can be accessed at GenBank under accession number RSFG00000000.

Survey AT-less type I PKSs in soil

Soil DNA isolation. The 18 soil samples at 20 cm depth of soil were collected in various places in China. The geological location for all the samples was listed in Table S192. DNA was extracted from soil using previously published DNA isolation protocol with modifications (Charlop-Powers et al., 2015). Crude environmental DNA was passed through one round of column purification using the Ezup Column Soil DNA

Purification Kit (Sangon Biotech, Shanghai, China).

PCR amplification and high-throughput sequencing. The HCS genes in the eDNA were amplified using the same PCR condition in PCR screening of HCS genes from in-house strains. The KS gene fragments of the AT-less type I PKS KSs were amplified from the same eDNA samples, using primers MAK1 (5'-GACACSGCSTGY TCBTCGTCG-3') and MAK3 (5'-CASTTGGTCCTGCCRCGCAGSTTGCC-3'). The following PCR conditions were used (Analytik Jena FlexCycler2): 95 °C for 5 min, followed by 10 cycles of denaturation (95 °C for 0.5 min), annealing (65 °C to 55 °C for 0.5 min), and extension (72 °C for 0.3 min), and 25 cycles of denaturation (95 °C for 0.5 min), annealing (55 °C for 0.5 min), and extension (72 °C for 0.3 min), and a final extension step at 72 °C for 10 min. The primers for HCS and KS were designed to contain a unique 12 bp barcode that allowed simultaneous sequencing different samples in a single run. PCR reactions were examined by agarose gel electrophoresis to determine the concentration and purity of each amplicon. Amplicons of the appropriate size were recovered using SanPrep Column DNA Gel Extraction Kit (Sangon Biotech, Shanghai, China). The purified HCS and KS amplicons were pooled in equal molar ratios and were then sequenced using Illumina MiSeq platform. The resulting fasta files with demultiplexed and quality-filtered sequences for each soil sample were used for downstream analysis.

Data analysis. Operational taxonomic units (OTUs) were generated using CD-HIT based on the following thresholds: the HCS nucleotide sequences were clustered

with sequence identity cut-off of 90%. The KS nucleotide sequences were first clustered at 97% sequence identity, and the non-redundant sequences were further clustered at 95% sequence identity. The formed OTUs of HCS sequences containing less than 10 reads were removed. The remaining OTUs of HCS sequences were used to do rarefaction analysis based on Chao1 diversity. The OTUs of HCS sequences containing at least 100 reads were selected to analyze high abundance HCS genes in soil samples.

HCS and KS centroid sequences were assigned to AT-less type I PKS gene clusters using BlastX command in Blast+ at E-value cutoff of 10^{-28} and 10^{-30}, respectively. Based on the BlastX results, the ratio of HCS gene hits in the AT-less type I PKS gene clusters in each soil sample were manually counted and calculated. To compare HCS gene beta diversity of the soil samples, nonmetric multidimensional scaling ordination analysis was performed using R package - vegan. Phylogenetic analysis of HCS gene hits was performed using MEGA 7. The heat map of HCSs or AT-less type I PKS KSs of each soil sample was generated using software HemI (Deng et al., 2014).

ASSOCIATED CONTENT

SUPPORTING INFORMATION

The supporting information is available free of charge on the Publication website.

MAJOR DATASET

The 16S rRNA sequence of *B. amyloliquefaciens* CB02999 (MK351256), The shotgun genome sequence of *S.* sp. CB01881 (RSFG00000000), The HCS gene fragments of the 13 in-house strains (Bioproject NO: PRJNA506854), and the high-throughput sequencing dataset of HCS and KS domains amplified from the soil DNA (Biorproject NO: PRJNA511514) were submitted to GenBank and may be accessed.

NOTES

The authors declare no competing financial interest.

REFERENCES

Blin K, Wolf T, Chevrette MG, Lu X, Schwalen CJ, Kautsar SA, Suarez Duran HG, de Los Santos ELC, Kim HU, Nave M, Dickschat JS, Mitchell DA, Shelest E, Breitling R, Takano E, Lee SY, Weber T, Medema MH. 2017. antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res* **45**: W36-41. doi: 10.1093/nar/gkx319.

24

Buchholz TJ, Rath CM, Lopanik NB, Gardner NP, Håkansson K, Sherman DH. 2007. Polyketide β-branching in bryostatin biosynthesis: identification of surrogate acetyl-ACP donors for BryR, an HMG-ACP synthase. *Chem Biol* **17**: 1092-1100. doi: 10.1016/j.chembiol.2010.08.008.

Butcher RA, Schroeder FC, Fischbach MA, Straight PD, Kolter R, Walsh CT, Clardy J. 2007. The identification of bacillaene, the product of the PksX megacomplex in *Bacillus subtilis*. *Proc Natl Acad Sci U S A* **104**: 1506-1509. doi: 10.1073/pnas.061050 3104.

Calderone CT. 2008. Isoprenoid-like alkylations in polyketide biosynthesis. *Nat Prod Rep* **25**: 845-853. doi: 10.1039/b807243d.

Calderone CT, Kowtoniuk WE, Kelleher NL, Walsh CT, Dorrestein PC. 2006. Convergence of isoprene and polyketide biosynthetic machinery: isoprenyl-S-carrier proteins in the pksX pathway of *Bacillus subtilis*. *Proc Natl Acad Sci U S A* **103**: 8977-8982. doi: 10.1073/pnas.0603148103.

Calderone CT, Iwig DF, Dorrestein PC, Kelleher NL, Walsh CT. 2007. Incorporation of nonmethyl branches by isoprenoid-like logic: multiple beta-alkylation events in the biosynthesis of myxovirescin A1. *Chem Biol* **14**: 835-846. doi: 10.1016/j.chembiol.200 7.06.008.

Chan YA, Boyne MT II, Podevels AM, Klimowicz AK, Handelsman J, Kelleher NL, Thomas MG. 2006. Hydroxymalonyl-acyl carrier protein (ACP) and aminomalonyl-ACP are two additional type I polyketide synthase extender units. *Proc Natl Acad Sci U S A.* **103**:14349-14354. doi: 10.1073/pnas.0603748103.

Chang Z, Sitachitta N, Rossi JV, Roberts MA, Flatt PM, Jia J, Sherman DH, Gerwick WH. 2004. Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacterium *Lyngbya majuscula*. *J Nat Prod* **67**: 1356-1367. doi: 10.1021/np0499261.

Charlop-Powers Z, Owen JG, Reddy BV, Ternei MA, Guimarães DO, de Frias UA, Pupo MT, Seepe P, Feng Z, Brady SF. 2015. Global biogeographic sampling of bacterial secondary metabolism. *Elife* **4**: e05048. doi: 10.7554/eLife.05048.

Charlop-Powers Z, Pregitzer CC, Lemetre C, Ternei MA, Maniko J, Hover BM, Calle PY, McGuire KL, Garbarino J, Forgione HM, Charlop-Powers S, Brady SF. 2016. Urban park soil microbiomes are a rich reservoir of natural product biosynthetic diversity. *Proc Natl Acad Sci U S A* **113**: 14811-14816. doi: 10.1073/pnas.1615581113.

Chen XH, Vater J, Piel J, Franke P, Scholz R, Schneider K, Koumoutsi A, Hitzeroth G, Grammel N, Strittmatter AW, Gottschalk G, Süssmuth RD, Borriss R. 2006. Structural and functional characterization of three polyketide synthase gene clusters in *Bacillus*

amyloliquefaciens FZB 42. *J Bacteriol* **188**: 4024-4036. DOI: 10.1128/JB.00052-06.

Cheng YQ, Tang GL, Shen B. 2003. Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis. *Proc Natl Acad Sci U S A* **100**: 3149-54. doi: 10.1073/pnas.0537286100.

Crits-Christoph A, Diamond S, Butterfield CN, Thomas BC, Banfield JF. 2018. Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis. *Nature* **558**: 440-444. doi: 10.1038/s41586-018-0207-y.

Deng W, Wang Y, Liu Z, Cheng H, Xue Y. 2014. HemI: a toolkit for illustrating heatmaps. *PLoS One* **9**: e111988. doi: 10.1371/journal.pone.0111988.

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. doi: 10.1016/j.chembiol.2004.03.030.

Ferguson JJ Jr, Rudney H. 1959. The biosynthesis of beta-hydroxy-beta-methylglutaryl coenzyme A in yeast. I. Identification and purification of the hydroxymethylglutaryl coenzymecondensing enzyme. *J Biol Chem* **234**: 1072-1075.

Genilloud O, González I, Salazar O, Martín J, Tormo JR, Vicente F. 2011. Current approaches to exploit actinomycetes as a source of novel natural products. *J Ind Microbiol Biotechnol* **38**: 375-389. doi: 10.1007/s10295-010-0882-7.

Green MR. Sambrook J. 2012. Molecular Cloning: A Laboratory Manual (Fourth Edition). Cold Spring Harbor Laboratory Press.

Haines AS, Dong X, Song Z, Farmer R, Williams C, Hothersall J, Płoskoń E, Wattana-Amorn P, Stephens ER, Yamada E, Gurney R, Takebayashi Y, Masschelein J, Cox RJ, Lavigne R, Willis CL, Simpson TJ, Crosby J, Winn PJ, Thomas CM, Crump MP. 2013. A conserved motif flags acyl carrier proteins for *β*-branching in polyketide synthesis. *Nat Chem Biol* **9**: 685-692. doi: 10.1038/nchembio.1342.

Helfrich EJ, Piel J. 2016. Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat Prod Rep* **33**: 231-316. doi: 10.1039/c5np00125k.

Hindra, Huang T, Yang D, Rudolf JD, Xie P, Xie G, Teng Q, Lohman JR, Zhu X, Huang Y, Zhao LX, Jiang Y, Duan Y, Shen B. 2014. Strain prioritization for natural product discovery by a high-throughput real-time PCR method. *J Nat Prod* **77**: 2296-2303. doi: 10.1021/np5006168.

Huang Y, Niu B, Gao Y, Fu L, Li W. 2010. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* **26**: 680-682. doi: 10.1093/bioinformatics/btq003.

Huang Y, Huang SX, Ju J, Tang G, Liu T, Shen B. 2011. Characterization of the *lnmKLM* genes unveiling key intermediates for *β*-alkylation in leinamycin biosynthesis. *Org Lett* **13**: 498-501. doi: 10.1021/ol102838y.

Jenke-Kodama H, Sandmann A, Müller R, Dittmann E. 2005. Evolutionary implications of bacterial polyketide synthases. *Mol Biol Evol* **22**: 2027-39. doi: 10.1093/molbev/msi193.

Jenner M, Kosol S, Griffiths D, Prasongpholchai P, Manzi L, Barrow AS, Moses JE, Oldham NJ, Lewandowski JR, Challis GL. 2018. Mechanism of intersubunit ketosynthase-dehydratase interaction in polyketide synthases. *Nat Chem Biol* **14**: 270-275. doi: 10.1038/nchembio.2549.

Khosla C, Herschlag D, Cane DE, Walsh CT. 2014. Assembly line polyketide synthases: mechanistic insights and unsolved problems. *Biochemistry* **53**: 2875-2883. doi: 10.1021/bi500290t.

Kieser T, Bibb MJ, Buttner MJ, Chater KF, Hopwood DA. 2000. Practical *Streptomyces* Genetics. John Innes Foundation, Norwich, U.K.

Kohl M, Wiese S, Warscheid B. 2011. Cytoscape: software for visualization and analysis of biological networks. *Methods Mol Biol* **696**: 291-303. doi: 10.1007/978-1-60761-987-1_18.

Lange BM, Rujan T, Martin W, Croteau R. 2000. Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. *Proc Natl Acad Sci U S A* **97**: 13172-13177. doi: 10.1073/pnas.240454797.

Liu T, Huang Y, Shen B. 2009. Bifunctional acyltransferase/decarboxylase LnmK as the missing link for *β*-alkylation in polyketide biosynthesis. *J Am Chem Soc* **131**: 6900-6901. doi: 10.1021/ja9012134.

Lohman JR, Ma M, Osipiuk J, Nocek B, Kim Y, Chang C, Cuff M, Mack J, Bigelow L, Li H, Endres M, Babnigg G, Joachimiak A, Phillips GN Jr, Shen B. 2015. Structural and evolutionary relationships of "AT-less" type I polyketide synthase ketosynthases. *Proc Natl Acad Sci U S A* **112**: 12693-12698. doi: 10.1073/pnas.1515460112.

Maloney FP, Gerwick L, Gerwick WH, Sherman DH, Smith JL. 2016. Anatomy of the β-branching enzyme of polyketide biosynthesis and its interaction with an acyl-ACP substrate. *Proc Natl Acad Sci U S A* **113**: 10316-10321. doi: 10.1073/pnas.160721011

3.

Moldenhauer J, Götz DC, Albert CR, Bischof SK, Schneider K, Süssmuth RD, Engeser M, Gross H, Bringmann G, Piel J. 2010. The final steps of bacillaene biosynthesis in *Bacillus amyloliquefaciens* FZB42: direct evidence for β, γ dehydration by a trans-acyltransferase polyketide synthase. *Angew Chem Int Ed Engl* **49**: 1465-7. doi: 10.1002/anie.200905468.

Nakamura H, Hamer HA, Sirasani G, Balskus EP. 2012. Cylindrocyclophane biosynthesis involves functionalization of an unactivated carbon center. *J Am Chem Soc* **134**: 18518-21. doi: 10.1021/ja308318p.

Newman DJ, Cragg GM. 2016. Natural products as sources of new drugs from 1981 to 2014. *J Nat Prod* **79**: 629-661. doi: 10.1021/acs.jnatprod.5b01055.

Nguyen T, Ishida K, Jenke-Kodama H, Dittmann E, Gurgui C, Hochmuth T, Taudien S, Platzer M, Hertweck C, Piel J. 2008. Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection. *Nat Biotechnol* **26**: 225-233. doi: 10.1038/nbt1379.

O'Brien RV, Davis RW, Khosla C, Hillenmeyer ME. 2014. Computational identification and analysis of orphan assembly-line polyketide synthases. *J Antibiot (Tokyo)* **67**: 89-97. doi: 10.1038/ja.2013.125.

Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H. 2012. Vegan: Community Ecology Package.

Pan G, Xu Z, Guo Z, Hindra, Ma M, Yang D, Zhou H, Gansemans Y, Zhu X, Huang Y, Zhao LX, Jiang Y, Cheng J, Van Nieuwerburgh F, Suh JW, Duan Y, Shen B. 2017. Discovery of the leinamycin family of natural products by mining actinobacterial genomes. *Proc Natl Acad Sci U S A* **114**: E11131-11140. doi: 10.1073/pnas.17162451 15.

Piel J. 2010. Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat Prod Rep* **27**: 996-1047. doi: 10.1039/b816430b.

Pye CR, Bertin MJ, Lokey RS, Gerwick WH, Linington RG. 2017. Retrospective analysis of natural products provides insights for future discovery trends. *Proc Natl Acad Sci U S A* **114**: 5601-5606. doi: 10.1073/pnas.1614680114.

Helfrich EJ, Piel J. 2016 Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat Prod Rep* **33**: 231-316. doi: 10.1039/c5np00125k.

Piel J. 2002. A polyketide synthase-peptide synthetase gene cluster from an uncultured

bacterial symbiont of Paederus beetles. *Proc Natl Acad Sci U S A* **99**: 14002-14007. doi: 10.1073/pnas.222481399.

Piel J, Hui D, Fusetani N, Matsunaga S. 2004. Targeting modular polyketide synthases with iteratively acting acyltransferases from metagenomes of uncultured bacterial consortia. *Environ Microbiol* **6**: 921-927. doi: 10.1111/j.1462-2920.2004.005 31.x.

Qin W, Liu Y, Ren P, Zhang J, Li H, Tian L, Li W. 2014. Uncovering a glycosyltransferase provides insights into the glycosylation step during macrolactin and bacillaene biosynthesis. *Chembiochem* **15**: 2747-2753. doi: 10.1002/c bic.20140 2384.

Rose TM, Henikoff JG, Henikoff S. 2003. CODEHOP (COnsensus-DEgenerate Hybrid Oligonucleotide Primer) PCR primer design. *Nucleic Acids Res* **31**: 3763-3766. http://dx.doi.org/10.1093/nar/gkg524.

Savic M, Vasiljevic B. 2006. Targeting polyketide synthase gene pool within actinomycetes: new degenerate primers. *J Ind Microbiol Biotechnol* **33**: 423-430. doi: 10.1007/s10295-006-0080-9.

Vander Wood DA, Keatinge-Clay AT. 2018. The modules of trans-acyltransferase assembly lines redefined with a central acyl carrier protein. *Proteins* **86**: 664-675. doi: 10.1002/prot.25493.

Wakimoto T, Egami Y, Nakashima Y, Wakimoto Y, Mori T, Awakawa T, Ito T, Kenmoku H, Asakawa Y, Piel J, Abe I. 2014. Calyculin biogenesis from a pyrophosphate protoxin produced by a sponge symbiont. *Nat Chem Biol* **10**: 648-655. doi: 10.1038/nchembio.1573.

Wilding EI, Brown JR, Bryant AP, Chalker AF, Holmes DJ, Ingraham KA, Iordanescu S, So CY, Rosenberg M, Gwynn MN. 2000. Identification, evolution, and essentiality of the mevalonate pathway for isopentenyl diphosphate biosynthesis in gram-positive cocci. *J Bacteriol* **182**: 4319-4327. doi:10.1128/JB.182.15.4319-4327.2000

Xie X, Khosla C, Cane DE. 2017. Elucidation of the stereospecificity of C-methyltransferases from trans-AT polyketide synthases. *J Am Chem Soc* **139**: 6102-6105. doi: 10.1021/jacs.7b02911.

Yan X, Ge H, Huang T, Hindra, Yang D, Teng Q, Crnovčić I, Li X, Rudolf JD, Lohman JR, Gansemans Y, Zhu X, Huang Y, Zhao LX, Jiang Y, Van Nieuwerburgh F, Rader C, Duan Y, Shen B. 2016; Strain prioritization and genome mining for enediyne natural products. *MBio.* **7**: e02104-2116. doi: 10.1128/mBio.02104-16.

Zhang F, He HY, Tang MC, Tang YM, Zhou Q, Tang GL. 2011. Cloning and elucidation of the FR901464 gene cluster revealing a complex acyltransferase-less polyketide synthase using glycerate as starter units. *J Am Chem Soc* **133**: 2452-2462. doi: 10.1021/ja105649g.

FIGURE LEGENDS

Figure 1. Selected polyketides biosynthesized by AT-less type I PKSs (**A**) and the mechanism of *β*-branch formation involving HCS cassette proteins (**B**). The *β*-branches are highlighted in red.

Figure 2. Sequence similarity network (SSN) analysis for HCSs. (**A**) SSN of the 2,215 HCS sequences. The SSN was displayed with an *E*-value threshold of 10^{-135}. Colors represent different clusters of HCSs. (**B**) A sub-group of HCSs responsible for *β*-branching in polyketide biosynthesis. The corresponding polyketide products and their corresponding biosynthetic enzymes were embedded.

Figure 3. GNN analysis of 170 orphan AT-less type I PKSs and 29 characterized AT-less type I PKSs. Blue square (■): orphan AT-less type I PKSs; red square (■): characterized AT-less type I PKSs; yellow square (■): leinamycin-type AT-less type I PKSs. The inserted legends represent each protein annotation, in which the HCS cassette proteins are colored in pink.

Figure 4. The survey of 1207 in-house strains using specific HCS probes. (**A**) Sequence alignment of selected HCS sequences from characterized AT-less type I PKS BGCs. Including BaeG (ABS74058.1), BatC (ADD82944.1), BonG (AFN27479.1), BryR (ABM63533.1), CalT (BAP05578.1), CorE (ADI59527.1), CorE (ADI59527.1),

CylF (ARU81120.1), DifL (CAG23983.1), ElaL (AEC04358.1), Fr9K (AIC32697.1), LnmM (AAN85526.1), MupH (AAM12922.1), NosK (WP_094329473.1), OnnA (AAV97869.1), OocE (AFX60327.1), PedP (AAW33975.1), SnaI (CBW45745.1), TaC (CAB46502.1), TaF (CAB46505.1), TmlH (CBK62727.1) and TstK (AGN11885.1). Aligned amino acid residues are colored based on the conservation level. The amino acid sequences used to design degenerated primers to clone HCS genes using CODHOP strategy; (**B**) PCR amplification of specific HCS gene fragments from in-house strains; (**C**) Phylogenic analysis of the cloned HCS gene fragments; (**D**) The HPLC profile of *B. amyloliquefaciens* CB02999 and the production of BAEs. ●, BAE (**3**) ; ○, dihydro-BAE; ■, BAE B (**4**); □, dihydro-BAE B. (**E**) The BGC of AT-less type I PKS in *S.* sp. CB01881.

Figure 5. Profiling of HCS abundance in 18 soil samples. (**A**) Geological location of the 18 soil samples. The number of different HCS OTUs only with over 100 reads from high-throughput sequencing in each soil sample were included for the analysis and shown in parentheses; (**B**) The determination of the richness of HCSs using Chao1-based rarefraction analysis; (**C**) NMDS analysis of HCSs. (**D**) The distribution of HCS genes in soil samples S1-S18.

Figure 6. Profiling of KS domains of AT-less type I PKS in 11 soil samples. (**A**) The heatmap of KS domains from AT-less type I PKSs in 11 soil samples. The rows were arranged to match the phylogenetic tree of HCSs in the left. The color of the heatmap

represents the similarity of the KS domains in each soil sample with the individual KS

domains in HCS-containing AT-less type I PKS BGCs. The similarity was calculated

using BlastX. (**B**) The heatmap of AT-less type I PKS KSs of each soil. The rows and

columns were clustered based on similarity of the elements of the corresponding
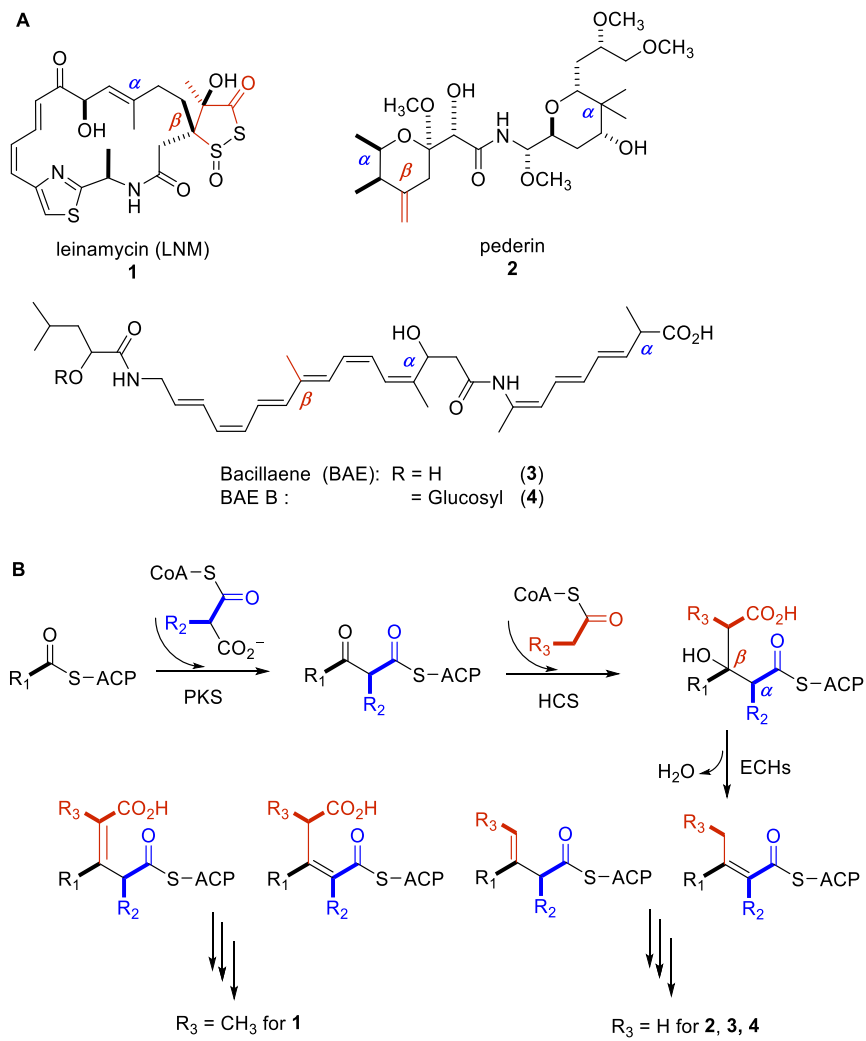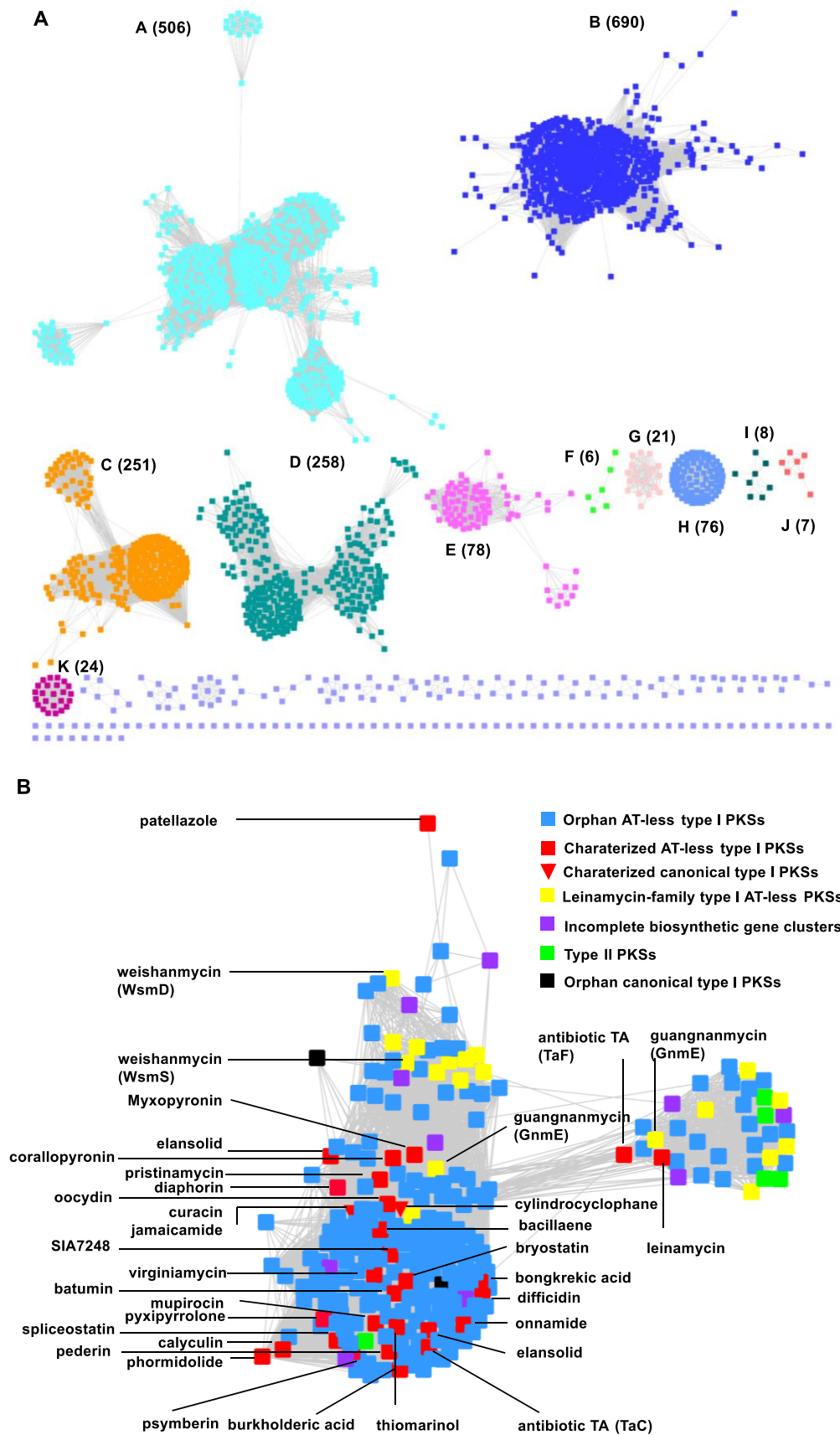
heatmap.

Figure 1.

leinamycin (LNM)
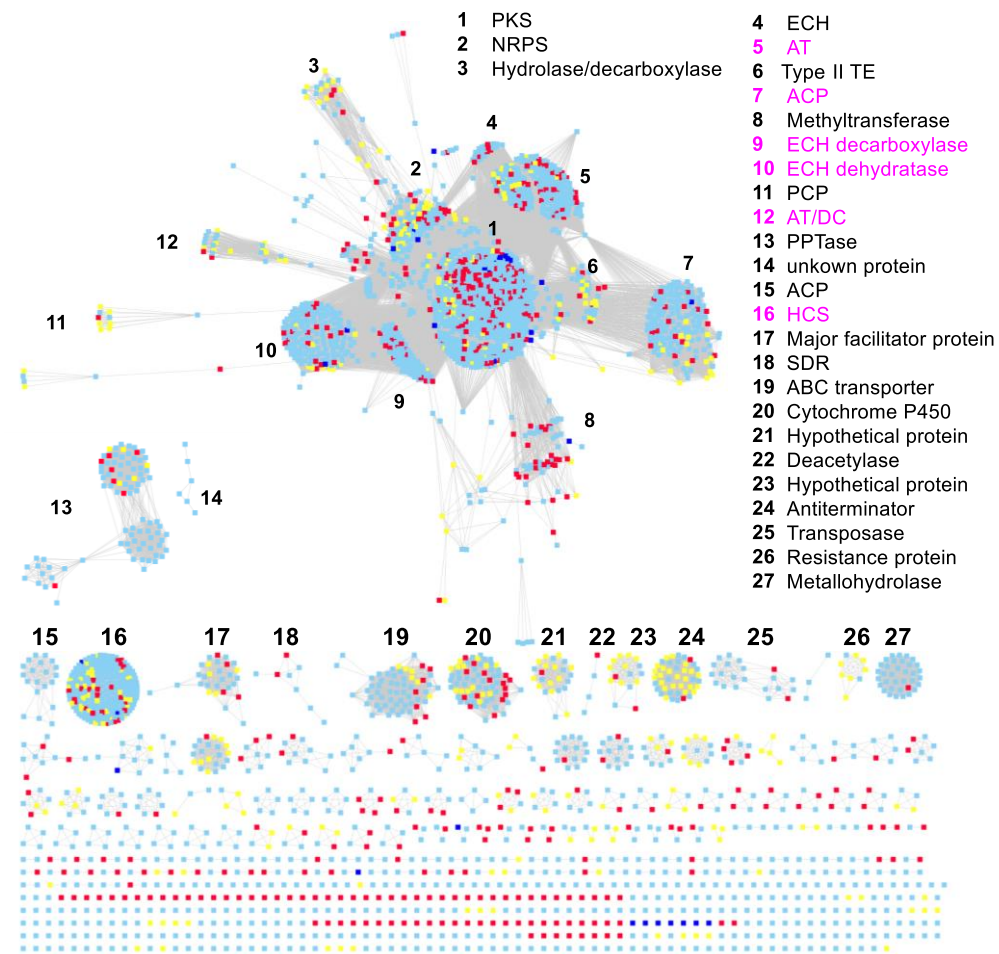**1**

pederin
2

Bacillaene (BAE): R = H   (**3**)
BAE B :   = Glucosyl   (**4**)

Figure 2.

Figure 3.

1 PKS		**4** ECH
2 NRPS		**5** AT
3 Hydrolase/decarboxylase		**6** Type II TE
		7 ACP
		8 Methyltransferase
		9 ECH decarboxylase
		10 ECH dehydratase
		11 PCP
		12 AT/DC
		13 PPTase
		14 unkown protein
		15 ACP
		16 HCS
		17 Major facilitator protein
		18 SDR
		19 ABC transporter
		20 Cytochrome P450
		21 Hypothetical protein
		22 Deacetylase
		23 Hypothetical protein
		24 Antiterminator
		25 Transposase
		26 Resistance protein
		27 Metallohydrolase

Figure 4.

Figure 5.

Figure 6.