

Evolution and Diversity of Assembly-Line Polyketide Synthases

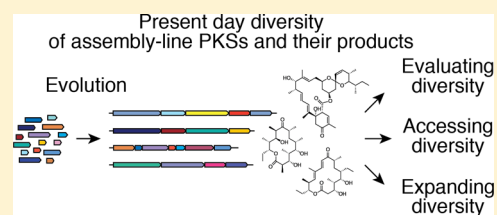
Focus Review

Aleksandra Nivina,^{†,‡} Kai P. Yuet,^{†,‡} Jake Hsu,^{‡,§} and Chaitan Khosla^{*,†,‡,§,¶}

[†]Department of Chemistry, [‡]Stanford ChEM-H, [§]Department of Chemical Engineering Stanford University, Stanford, California 94305, United States

S Supporting Information

ABSTRACT: Assembly-line polyketide synthases (PKSs) are among the most complex protein machineries known in nature, responsible for the biosynthesis of numerous compounds used in the clinic. Their present-day diversity is the result of an evolutionary path that has involved the emergence of a multimodular architecture and further diversification of assembly-line PKSs. In this review, we provide an overview of previous studies that investigated PKS evolution and propose a model that challenges the currently prevailing view that gene duplication has played a major role in the emergence of multimodularity. We also analyze the ensemble of orphan PKS clusters sequenced so far to evaluate how large the entire diversity of assembly-line PKS clusters and their chemical products could be. Finally, we examine the existing techniques to access the natural PKS diversity in natural and heterologous hosts and describe approaches to further expand this diversity through engineering.



CONTENTS

1. Introduction	12525		
1.1. Polyketide Synthases (PKSs)	12525		
1.2. Why Study PKS Evolution?	12526		
2. Evolution of Assembly-Line PKSs	12526		
2.1. Evolutionary Origins of Assembly-Line PKSs	12527		
2.2. Phylogeny of Catalytic Domains from Assembly-Line PKSs	12527		
2.2.1. Ketosynthase (KS) Domains	12527		
2.2.2. Acyltransferase (AT) Domains	12527		
2.2.3. Other Domains	12527		
2.3. Processes Involved in Assembly-Line PKS Diversification	12529		
2.3.1. Gene Duplication	12529		
2.3.2. Horizontal Gene Transfer (HGT)	12529		
2.3.3. Gene Conversion	12529		
2.3.4. Recombination	12529		
2.4. Models for Evolution of Assembly-Line PKSs	12529		
2.4.1. Current Model	12529		
2.4.2. Discordances in the Current Model	12529		
2.4.3. Alternative Model	12530		
2.4.4. Model for Evolutionary Unit of an Assembly-Line PKS	12530		
2.5. Factors Influencing the Evolution of Assembly-Line PKS Diversity	12530		
2.5.1. Environmental Factors	12530		
2.5.2. Genetic Factors	12531		
2.5.3. Evolutionary Advantages	12531		
3. Diversity of Orphan PKSs	12531		
3.1. Catalogues of Assembly-Line PKSs	12531		
3.2. Updated Catalogue of Orphan Assembly-Line PKSs	12532		
3.3. Evaluating the Product Diversity of Orphan Assembly-Line PKSs	12532		
3.4. Similarity Network of Assembly-Line PKSs	12533		
3.5. Eukaryotic PKS Clusters	12533		
3.6. Prioritizing PKS Clusters for Further Study	12535		
4. Accessing PKS Diversity	12535		
4.1. Expressing Assembly-Line PKSs in Heterologous Hosts	12535		
4.1.1. Phage Recombination-Assisted Cloning	12535		
4.1.2. Transformation-Associated Recombination Cloning	12536		
4.1.3. pCAP-Based Transformation-Associated Recombination Cloning	12536		
4.1.4. Cas9-Assisted Targeting of Chromosome (CATCH) Cloning	12536		
4.1.5. Site-Specific Recombination Cloning	12536		
4.2. Activating Assembly-Line PKSs in Native Hosts	12536		
5. Expanding PKS Diversity through Engineering	12537		
5.1. Combinatorial and Rational Engineering of Assembly-Line PKSs	12537		
5.2. Engineering Inspired by Evolution	12538		
5.2.1. Using Natural Splice Points	12538		
5.2.2. Using Natural Recombination Mechanisms	12539		
6. Future Directions	12539		

Received: August 24, 2019

Published: December 15, 2019

6.1. Understanding Assembly-Line PKS Evolution	12539
6.2. Expanding Access to Natural Diversity	12539
6.3. Overcoming Technical Challenges in Deorphanizing PKS Clusters	12540
6.4. Harnessing the Knowledge of PKS Evolution	12540
Associated Content	12540
Supporting Information	12540
Author Information	12540
Corresponding Author	12540
ORCID	12540
Notes	12540
Biographies	12540
Acknowledgments	12541
References	12541

1. INTRODUCTION

1.1. Polyketide Synthases (PKSs)

Polyketide synthases (PKSs) are multifunctional enzymes responsible for the biosynthesis of numerous natural products, many of which are currently used as antibiotics (e.g., erythromycin), antiparasitic drugs (e.g., avermectin), cholesterol-lowering agents (e.g., lovastatin), immunosuppressants (e.g., FK506), and cancer chemotherapy (e.g., epothilone). PKSs are classified into three types: type I PKSs are large multifunctional proteins comprised of several functional domains and found in both bacteria and fungi, type II PKSs are formed by discrete catalytic domains and are typically found in bacteria, type III PKSs are simpler chalcone synthase-type enzymes that catalyze the formation of the product within a single active site, mainly in plants and bacteria. Type I PKSs are subdivided into iterative PKSs (reviewed in ref 1) and assembly-line PKSs, also called modular PKSs (reviewed in ref 2). Whereas an iterative PKS catalyzes multiple chain elongation cycles using the same set of enzymatic domains, the nascent polyketide chain is channeled from one module to another within an assembly-line PKS such that each module typically catalyzes only one elongation cycle. Iterative type I PKSs are primarily found in fungi, while the assembly-line architecture predominates in bacteria, although several eukaryotic assembly-line PKSs have also been identified.

This review focuses on assembly-line PKSs, which are among the most complex biosynthetic protein machineries known in nature. The structure and mechanism of assembly-line PKSs have been a subject of numerous studies (reviewed in refs 3–6). The catalytic chemistry of a prototypical assembly-line PKS is schematically outlined in Figure 1. Within each module of the assembly-line, polyketide acyl chain elongation is catalyzed collaboratively by a ketosynthase (KS), an acyltransferase (AT), and an acyl carrier protein (ACP) domain. The ACP domain is post-translationally modified with a phosphopantetheinyl (P-pant) “swinging arm” by a P-pant transferase (PPTase). The KS receives the growing polyketide chain from the ACP of the previous module, while the AT trans-esterifies an α -carboxyacyl extender unit from an appropriate acyl-CoA metabolite onto the ACP. The KS then catalyzes a decarboxylative Claisen-like condensation between the polyketide intermediate and the extender unit. Before being translocated onto the KS of the next module, the newly synthesized, ACP-bound β -keto thioester intermediate can be modified by additional domains, such as a ketoreductase

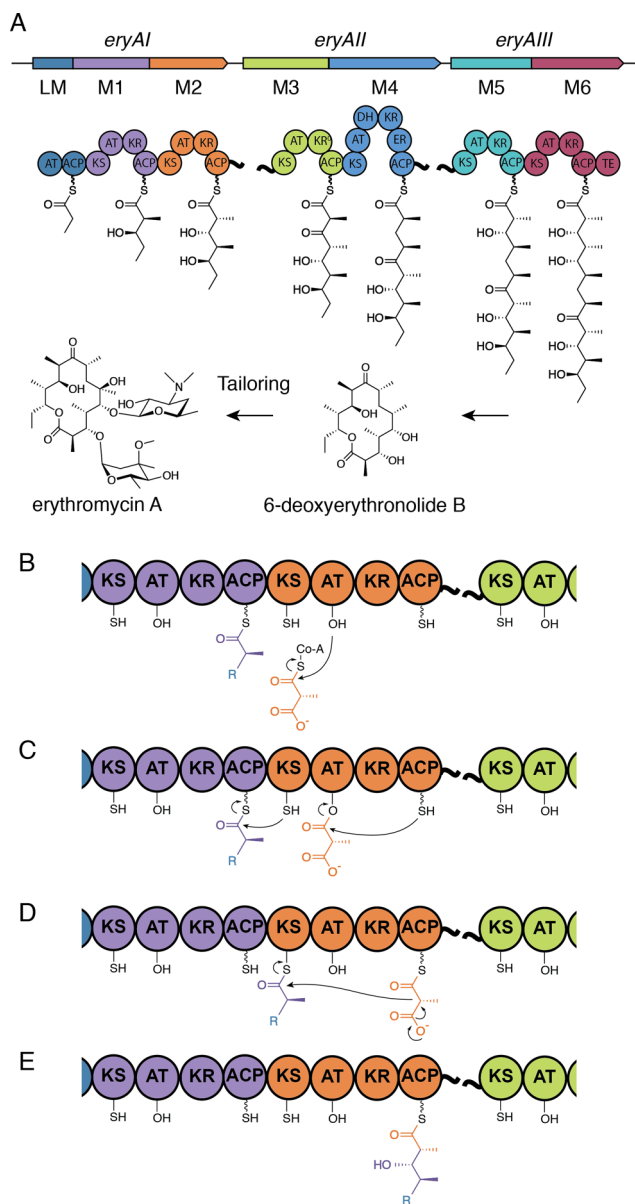


Figure 1. (A) The 6-deoxyerythronolide B synthase (DEBS), a prototypical assembly-line PKS, synthesizes 6-deoxyerythronolide B, the precursor of erythromycin A. (B–E) Reactions catalyzed by module 2 (M2) of DEBS. (B,C) Transacylation of the electrophilic and nucleophilic substrates of M2 from the ACP of module 1 (M1) and (2S)-methylmalonyl-CoA, respectively. (D,E) Polyketide chain elongation and ketoreduction. KS, ketosynthase; AT, acyltransferase; ACP, acyl carrier protein; KR, ketoreductase; KR⁰, redox-inactive KR with epimerase activity; DH, dehydratase; ER, enoylreductase; TE, thioesterase.

(KR), dehydratase (DH), enoylreductase (ER), methyltransferase (MT), or others. KR, DH, and ER successively and stereospecifically reduce the extended product into a β -hydroxyl, alkene, and methylene functionality, respectively; KR domains establish the stereoconfiguration of both the α - and β -carbon atoms of their products. These domains are usually encoded within the module but can also be present as free-standing proteins in *trans*.⁷ Ultimately, the full-length polyketide is released from the PKS by hydrolysis or macrocyclization catalyzed by a thioesterase (TE) domain or reductive cleavage.

General model of PKS evolution

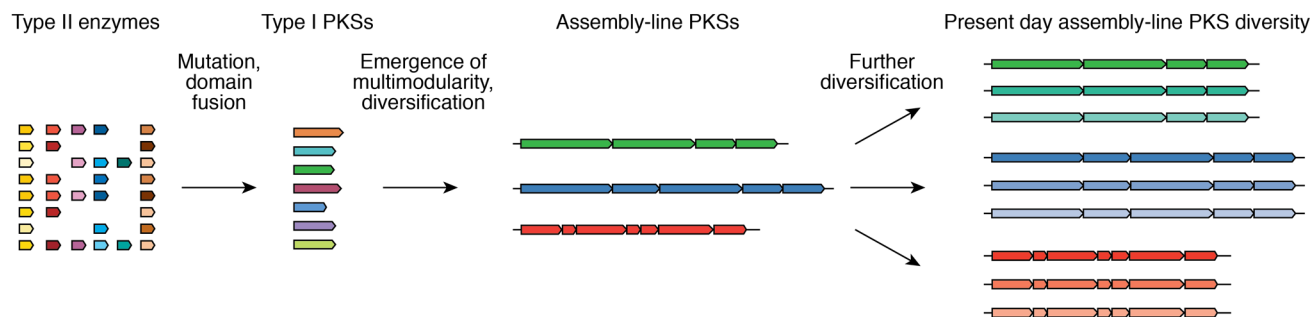


Figure 2. General model for PKS evolution. The multidomain architecture of type I PKS modules evidently arose through the fusion of genes encoding single-domain proteins of type II systems. The processes that led to the emergence of the multimodular architecture are less well understood. For instance, it is unclear whether assembly-line systems evolved from iterative PKSs that lost their ability to perform several consecutive condensation reactions on the same polyketide chain or from a separate subset of type II proteins. Once a set of assembly-line PKSs emerged, other processes allowed further diversification of these modular enzymes and their products.

Individual modules of assembly-line PKSs are classified as *cis*-AT and *trans*-AT modules. *cis*-AT modules contain all three essential domains (KS, AT, and ACP) comprising a PKS, whereas in *trans*-AT modules the extender unit is transacylated onto the ACP domain by a free-standing AT that is often shared across multiple modules (reviewed in ref 8). Modules are connected either by intermodular linkers⁹ or, if a PKS spans several polypeptides, by docking domains that establish specific noncovalent interactions between successive modules.¹⁰ The architectures of *cis*-AT PKSs are often colinear to their genetic encoding (i.e., the order in which modules are encoded on the DNA level corresponds to the order in which they operate), whereas the modules of *trans*-AT assembly lines are often not colinear.¹¹

A number of tailoring enzymes can further modify the backbone, either while the intermediates are still bound to the assembly line or after they are released.^{12–14} Typically, all genes involved in the biosynthesis of the final product are colocalized within the bacterial genome, forming a biosynthetic gene cluster (BGC).

1.2. Why Study PKS Evolution?

Assembly-line PKSs can contain up to 30 modules, distributed over several polypeptide chains. Together with nonribosomal peptide synthetases (NRPSs), they comprise two related classes of megasynthases attaining up to several MDa in size and responsible for the biosynthesis of numerous secondary metabolites. In addition to their remarkable catalytic mechanisms, their multimodular architecture also provides a unique example of studying the evolution of genes that encode multiple homologous but functionally distinct units. From a fundamental standpoint, there is a compelling correlation between genotypic and phenotypic diversity within this family of enzymes.¹⁵

From a practical perspective, the study of assembly-line PKS evolution and diversity could help us expand our therapeutic arsenal. On one hand, exploration of natural PKS diversity holds the potential of discovering assembly lines that synthesize new bioactive polyketides. On the other hand, a better understanding of mechanisms used by nature for polyketide diversification could open new avenues for PKS engineering. Evolutionary-inspired approaches have already started to find use in guiding the assembly of chimeric PKSs that produce novel biomolecules.

In the next sections, we will summarize the current models of assembly-line PKS evolution and their impact on enzyme engineering, evaluate the diversity of natural PKSs, provide an overview of methods that allow accessing this diversity through the activation of BGCs, and discuss the broader implications of this evolutionary analysis for the field of natural products research.

2. EVOLUTION OF ASSEMBLY-LINE PKSS

There is compelling evidence to suggest that all PKSs are evolutionarily related: despite the differences in their architectures and mechanisms, their domains belong to the same protein families and catalyze similar reactions. However, the precise evolutionary relationships between different PKSs are unclear and therefore present an outstanding challenge. The multimodular architecture of assembly-line PKSs is uncommon among proteins, meaning that the selective pressures and molecular mechanisms involved in their evolution could be distinct from those operating in most protein families. Nonuniform distribution of different PKS types among bacterial and eukaryotic phyla further complicates the challenge. If one assumes that iterative PKSs predated assembly lines, then evolution of the present-day diversity of assembly-line PKSs likely involved genetic processes such as mutation, gene fusion to establish module architecture, gene duplication to yield multimodular PKSs, and further diversification of assembly lines via mutation, recombination, and interspecies horizontal gene transfer (HGT) (Figure 2A).

The most profound difference between iterative and assembly-line PKSs lies in the chemistry of the chain translocation reaction involving a KS and an ACP domain. Specifically, whereas KSs of iterative PKSs operate multiple times on the same polyketide chain,¹⁶ the KS domains of assembly-line PKSs must release their β -ketoacyl-ACP product before the newly vacated active site Cys residue can attack the reactive thioester linkage in this product (Figure 1C). While the precise mechanism by which such back-transfer of the growing polyketide chain is precluded remains unclear, chain elongation by assembly-line PKS modules is energetically coupled to intermodular chain translocation via a “turnstile” mechanism: a module is precluded from accepting a new chain until the product of previous chain elongation cycle has been passed down to the downstream module.^{17,18} This avoids KS reacylation by the downstream ACP and consequent iterative

chain elongation. The existence of the “turnstile” mechanism suggests that chain translocation between different modules is an evolutionarily acquired feature, i.e., a gain-of-function mutation as opposed to a loss-of-function trait.

In this section, we will first place the evolution of assembly-line PKSs in the context of related enzymes such as iterative PKSs and fatty acid synthases. This will be followed by a phylogenetic analysis of the domains of assembly-line PKSs as well as a brief review of genetic processes believed to play important roles in assembly-line PKS evolution. Both of these concepts are critical to understanding current models for assembly-line PKS evolution (and limitations thereof), which is the principal focus of this section.

2.1. Evolutionary Origins of Assembly-Line PKSs

Assembly-line PKSs are evolutionarily related to a number of other multifunctional enzyme families. Even though models for evolutionary relationships have been previously proposed based on phylogenetic studies (Supporting Information, Figure S1), the origins of diverse PKS types and subtypes are not yet fully understood. Relatively close homologues include type I iterative PKSs, such as those found in fungi,^{19–21} certain lipid biosynthetic pathways of mycobacteria,^{22,23} enediene synthases in actinobacteria,²⁴ polyunsaturated fatty acid (PUFA) synthases,²⁵ and heterocyst glycolipid synthases in nitrogen-fixing cyanobacteria.²⁶

Assembly-line PKSs are also evolutionarily related to fatty acid synthases (FASs). Their modular architectures differ significantly from bacterial and fungal type I FASs²⁷ and more closely resemble vertebrate FASs instead, although it is unclear whether these relationships are products of divergent or convergent evolution.^{20,21} The evolutionary relatedness to type II PKS and FAS systems is even more distant.

Even though most enzymatic domains that form PKSs and NRPSs belong to different protein families, the two assembly-line systems use very similar biosynthetic strategies and often form hybrid assemblies: about one-third of biosynthetic gene clusters encode both types of enzymes.²⁸ These assembly lines appear to have evolved to facilitate translocation of hybrid products between individual PKS and NRPS modules, and their carrier proteins are serviced by the same PPTases and TEs with broad substrate specificity (reviewed in refs 29 and 30). Surprisingly, hybrid assemblies have a wide array of architectures including nonmodular, iterative, assembly line, or mixed type.²⁸ Given their prevalence and the presence of specialized domains and interfaces to ensure intermodular interactions, it is tempting to speculate that these hybrid assemblies appeared early in the evolutionary history of polyketide and nonribosomal peptide natural products. However, this subject is beyond the scope of the current review.

2.2. Phylogeny of Catalytic Domains from Assembly-Line PKSs

Most evolutionary relationships of PKSs to related enzymes were deduced from the overall biosynthetic enzyme architecture and the alignment of KS domains, which show the highest degree of amino acid sequence conservation. However, when exploring the emergence of multimodularity within PKSs, an analysis of KS domains alone is insufficient; it does not reflect the entire evolutionary history of assembly-line PKSs, as shown by phylogenetic studies of other domains.

2.2.1. Ketosynthase (KS) Domains. Within assembly-line PKSs, KS domains fall into two clades corresponding to *cis*-AT

and *trans*-AT enzymes.³¹ The phylogenetic tree of KS domains of *cis*-AT PKSs typically follows the phylogeny of the host organisms, with higher sequence identities within a single BGC and, to a lesser extent, different assembly lines within the species.^{32,33} The two exceptions are KS domains from mixed NRPS/PKS systems, which ligate a peptide intermediate from the upstream NRPS module to a polyketide extender unit, and the decarboxylative KS_Q (KS⁰) domains, whose active site Cys residue is replaced by Gln. These two groups form separate branches that are very close to corresponding domains of *trans*-AT PKSs.³³

In contrast to the KS domains of *cis*-AT PKSs, KS domains of *trans*-AT PKSs are not phylogenetically grouped with other KSs from the same BGC. Instead, the closest KS relatives almost always elongate structurally similar polyketide intermediates.^{34,35} It has been noted that KS domains from *trans*-AT PKSs are less promiscuous than *cis*-AT KSs³⁶ and form evolutionarily conserved units with ACPs from the upstream modules rather than ACPs from the same module.³⁷ A similar phylogenetic pattern has been observed for a group of aminopolyol synthases,³⁸ suggesting that it also may apply to some subsets of *cis*-AT PKSs.³⁹

2.2.2. Acyltransferase (AT) Domains. In *cis*-AT PKSs, AT domains comprise two clades based on their substrate specificity. Apart from a small number of exceptions, one clade contains AT domains utilizing malonyl-CoA, while the other corresponds to AT domains utilizing methylmalonyl-CoA and rarer substrates.^{20,40}

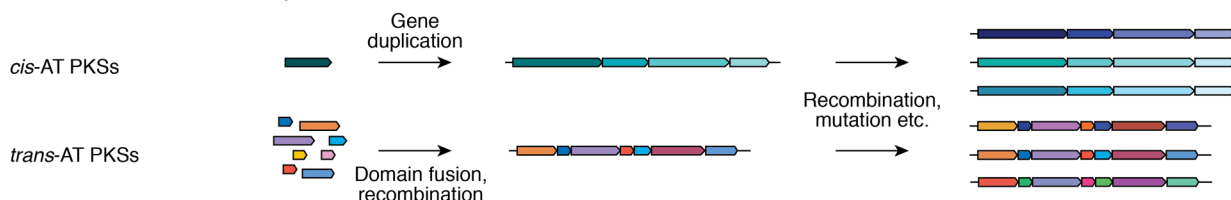
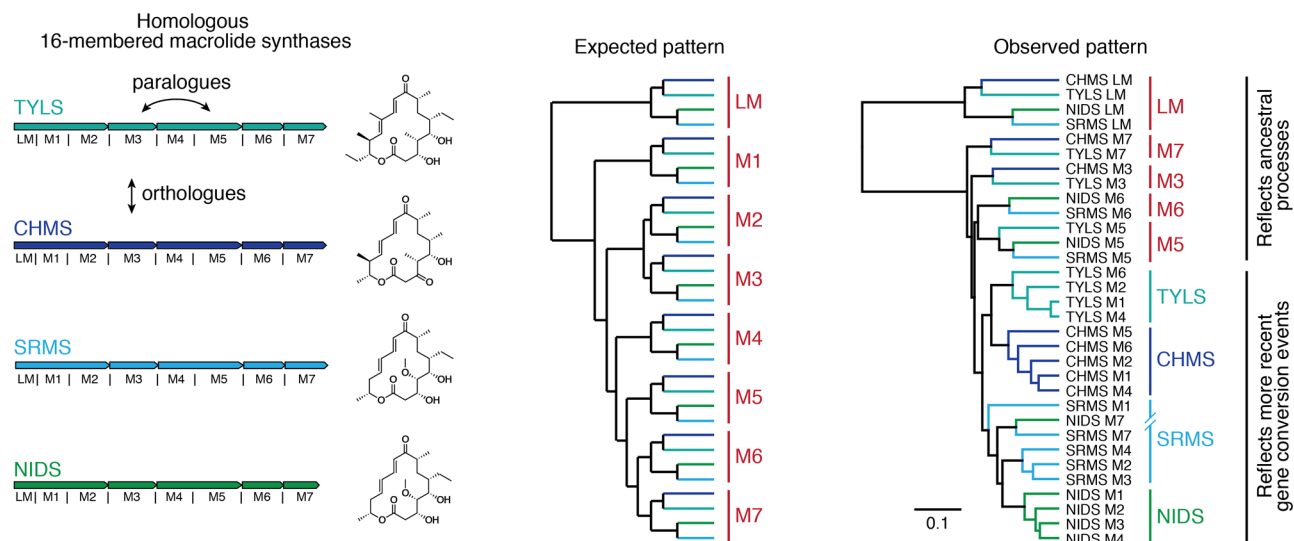
In *trans*-AT PKSs, free-standing AT proteins comprise a distinct clade from their counterparts in *cis*-AT PKSs.⁴¹ These ATs also distribute across two subclades: one that includes catalytically relevant acyltransferases (nearly all of which utilize malonyl extender units) and another that includes enzymes with acyl hydrolase activity and are therefore capable of hydrolyzing acetyl groups that are erroneously *trans*-acylated from acetyl-CoA onto an ACP.^{42,43}

2.2.3. Other Domains. Like AT domains, KRs also cluster based on their catalytic properties. In *cis*-AT PKSs, KRs comprise two clades that segregate based on alcohol stereochemistry.⁴⁴ In *trans*-AT PKSs, KRs are distributed across four clades that are distinguished by not just alcohol stereochemistry but also the presence of other enzymes within the module, including methyltransferases (MTs) and dehydratases.³⁷

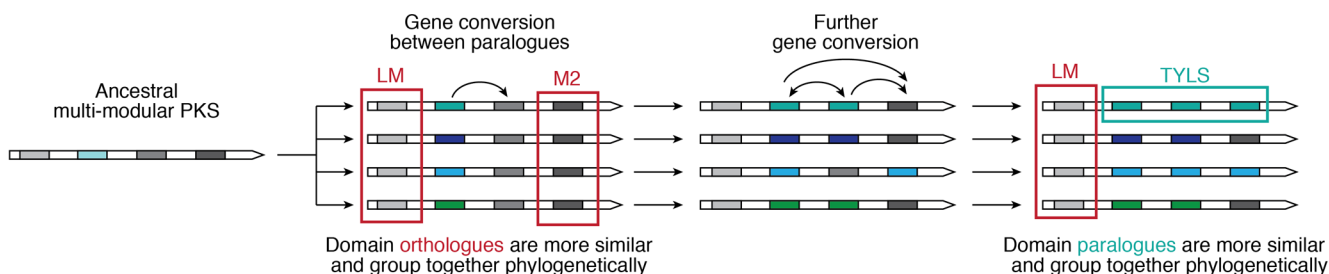
MT domains are relatively rare in *cis*-AT PKSs; they usually present an alternative mechanism for introducing an α -C substituent into the polyketide backbone by modules with malonyl-specific AT domains. The phylogeny of MT domains reflects the identity of the methyl acceptor, i.e., C- versus O-methyltransferases.⁴⁵ Notably, the N-MTs from NRPSs are also closely related to their homologues from *cis*-AT PKSs, albeit in a clade of their own. In contrast, MT domains in *trans*-AT PKSs cluster more variably, likely based on module composition as well as substrate specificity.³⁴

ACP domains are relatively short and variable, which complicated phylogenetic analysis until recently, when more sequences became available. In *trans*-AT PKSs, ACP clades track with those of their downstream KSs.³⁷ The ACPs from *cis*-AT PKSs comprise their own clade, although no clear clustering principle can be gleaned from this clade. Nonetheless, the ACPs of giant aminopolyol synthases appear to evolutionarily comigrate together with their downstream KSs, similar to the *trans*-AT ACPs.^{38,39}

A. Current model of assembly-line PKS evolution

B. Discordance in the phylogeny of KS domains from *cis*-AT PKSs

C. Gene conversion in assembly-line PKSs explains the discordance



D. Alternative model of assembly-line PKS evolution

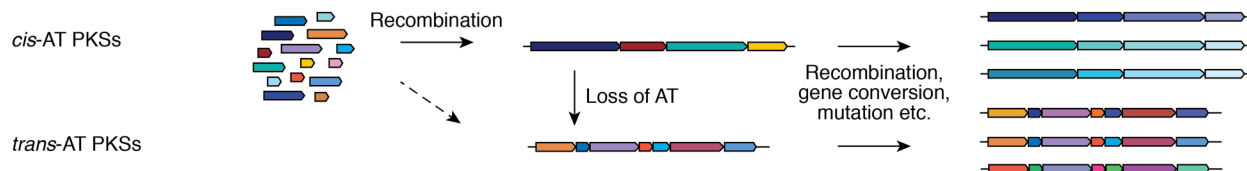


Figure 3. Models of *cis*-AT and *trans*-AT PKS evolution. (A) It has been hypothesized that evolution of *cis*- versus *trans*-AT PKSs took distinct paths.^{21,32,34} However, this dichotomy has some discordances. It does not explain the absence of iterative *trans*-AT PKSs, the convergence toward strikingly similar architectures despite different evolutionary paths, the presence of AT domain vestiges in *trans*-AT modules,³⁴ or (B) the inconsistency of the phylogenetic tree of *cis*-AT KS domains with this hypothesis.⁶⁴ The last inconsistency is exemplified by KS domains from four homologous 16-membered macrolide synthases (left; TYLS, tylectone synthase; CHMS, chalomycin synthase; SRMS, spiramycin synthase; NIDS, niddamycin synthase). Under the current model, their KS domains would be expected to form groups of orthologous domains (center). In fact, most KS domains are grouped with paralogues from the same PKS (right). Protein sequence alignment was performed with ClustalOmega,⁸⁴ and the dendrogram was constructed using UPGMA hierarchical clustering. (C) The discordance in KS sequence alignment is a result of concerted evolution and can be explained by gene conversion events between KS domains.^{64,82} Gene conversion leads to high sequence similarity between paralogous domains, causing them to cluster closer to each other than to their orthologues (e.g., teal square). Because gene conversion need not affect all domains within a PKS (e.g., red square), some of them maintain a phylogenetic pattern reflecting ancestral events that had led to the separation of homologous assembly-line PKSs. (D) An alternative model for assembly-line PKS evolution builds on the hypothesis that *trans*-AT PKSs evolved from *cis*-AT PKSs through loss of AT domains. In this model, the high sequence identity of KS domains in *cis*-AT PKSs would be explained by subsequent gene conversion events rather than ancestral gene duplications.

2.3. Processes Involved in Assembly-Line PKS Diversification

Not only does phylogenetic analysis of individual domains from assembly-line PKSs provide insight into evolutionary relationships within this PKS family, but it also highlights the genetic processes that may have led to their diversification. In this section, we review the genetic processes that are thought to have played important roles in the evolution of assembly-line PKSs.

2.3.1. Gene Duplication. Early in the study of assembly-line PKSs, it was noted that some modules within the same PKS share exceptionally high levels of sequence similarity.⁴⁶ The clustering of KS domains derived from the same assembly line has been observed for many *cis*-AT PKSs and has led to speculation that their multimodularity arose mainly through repeated gene duplication, followed by further diversification through mutation.^{32,47} The fact that in most *cis*-AT PKSs, modules operate in the same order in which they are encoded on the DNA level, known as the principle of colinearity, also supported the role of gene duplication and deletion in their evolution.⁴⁸ As evidence for other processes, such as horizontal gene transfer, recombination, domain loss and acquisition, and gene conversion, accumulated, the gene duplication model was modified to include these processes.^{15,20,34,48,49}

2.3.2. Horizontal Gene Transfer (HGT). In bacteria, genetic diversity is often acquired through horizontal gene transfer (HGT), a process during which genetic information is transmitted laterally to other neighboring bacteria rather than vertically to their descendants.⁴⁹ There is ample evidence of its role in the evolution of assembly-line PKS clusters; in fact, it appears to have played a particularly strong role in PKS evolution in proteobacteria.^{20,21} This inference is based on the observation of phylogenetic incongruencies between PKS genes and host species, anomalous distribution of genes among bacterial groups and atypical nucleotide compositions, and is especially notable in gene clusters encoding the biosynthesis of streptomycin,^{50,51} epothilone,³² and lagriamide,⁵² among others, as well as PKS clusters of bacterial origin found in sponges,⁵³ filamentous fungi,⁵⁴ and other taxa. In one instance, HGT has even been observed experimentally.⁵⁵

The high frequency of HGT of PKS genes could be due to multiple factors. Some PKSs are encoded on plasmids^{56–58} or located within pathogenicity islands,⁵⁹ which facilitates gene transfer through conjugation, transposition, or transduction. Additionally, transposon-like sequences are often observed proximal to KS domains, highlighting the potential for transfer of these PKS genes through transposition,⁵² although no direct evidence of such events has been found. It has also been suggested that the high rate of HGT in actinomycetales could be due to the linearity or instability of their chromosomes.⁶⁰

2.3.3. Gene Conversion. Gene conversion is a process by which two homologous sequences are homogenized, where one sequence becomes a copy of another through unidirectional sequence replacement. It is widespread and well-described in eukaryotes.⁶¹ Examples of gene conversion have also been described in prokaryotes, where it is responsible for antigenic variation or the evolution of multigene families, but the extent of its importance in bacterial genomes is not well understood.⁶²

Gene conversion is thought to play a role in the evolution of *cis*-AT PKSs. For example, it may explain the almost identical sequences of modules comprising the mycolactone synthase⁶³

and also rationalize changes in the structures of some macrolide antibiotics.⁶⁴

2.3.4. Recombination. Recombination undoubtedly plays a major role in the evolution of assembly-line PKSs; indeed, gene duplication, transposition, and gene conversion all rely on recombination processes. However, recombination by itself is an important mechanism of PKS evolution and diversification, especially in the cases of *trans*-AT PKSs.^{34,65}

In *cis*-AT PKSs, the lack of sequence conservation in docking domain pairs that flank adjacent modules suggests that modules comprising this class of assembly-line PKSs underwent recombinational shuffling.⁶⁶ The rate of homologous recombination differs between bacterial taxa and is particularly high in *Streptomyces*, which harbor a significant fraction of known PKSs. These bacteria undergo extensive HGT and recombination between species; these processes more recognized as being more important in sequence divergence than point mutation.⁶⁷ Homologous recombination within the same species is even higher,⁶⁸ and its importance for the diversification of PKS clusters has been demonstrated in the case of the avermectin producer, *Streptomyces avermitilis*.⁶⁵

2.4. Models for Evolution of Assembly-Line PKSs

2.4.1. Current Model. In large part to account for the differences in the phylogenetic clustering of KS domains between *cis*-AT and *trans*-AT PKSs (section 2.2.1), the prevailing view states that assembly-line PKSs have evolved via two independent and fundamentally different mechanisms. For *cis*-AT PKSs, gene duplication within the same PKS gene cluster is thought to be the driver of their evolutionary diversification, whereas for *trans*-AT PKSs, recombination is the dominant process (Figure 3A).^{21,32,34} However, the necessity to evoke these distinct mechanisms leads to several discordances.

2.4.2. Discordances in the Current Model. The above two-model hypothesis implies that multimodularity of assembly-line PKSs evolved independently at least twice and converged to an almost identical architecture. While not inconceivable, a single origin of multimodularity in *cis*-AT and *trans*-AT PKSs would be more parsimonious. Indeed, recent studies suggest that the two classes of assembly-line PKSs are more closely related than previously thought. Even though *trans*-AT PKS modules lack an AT domain, they usually contain a region called ATd, a subdomain nested between the KS and downstream domains.⁶⁹ It is structurally similar to the rigid KS-AT linker of *cis*-AT PKSs, a region that plays an important role in ACP docking during chain elongation and translocation.⁷⁰ The ATd subdomains of *trans*-AT PKSs often contain two additional helices, which have been proposed to facilitate lateral interactions between PKSs.⁷¹ However, in some cases ATd subdomains also include a large fragment of the AT domain or even entire KS-AT didomains.⁷² These KS-AT regions of various lengths may represent evolutionary intermediates between bacterial *cis*-AT and *trans*-AT PKSs. The evolution of *trans*-AT PKSs through AT domain loss would explain the absence of iterative *trans*-AT PKSs, which should have existed if the evolution of the two PKS groups was independent, from two respective groups of iterative PKS. Intriguingly, iterative *cis*-AT PKSs exist not only as stand-alone enzymes but are sometimes present as “stuttering” modules within an assembly-line PKS (reviewed in refs 1,73). For example, the stigmatellin,⁷⁴ borrelidin,⁷⁵ aureothin,⁷⁶ and neo-aureothin⁴⁸ synthases each harbor a module that performs

more than one round of programmed chain elongation.^{77,78} In other cases, module iterations are stochastic, leading to minor byproducts. For example, certain modules of DEBS and the epothilone synthase have been shown to iterate at measurable frequencies.^{79,80} Although mechanisms have evolved to preclude back-transfer of polyketides in assembly-line PKSs (such as the “ratchet”⁸¹), these remnants of iterative functions could reflect the evolutionary origins of assembly-line PKSs.

If *cis*-AT PKSs originated through module duplication, then it is also unclear why only the phylogeny of KS domains supports this model. One would expect other domains of duplicated modules to also be closely related. However, the non-KS domains are phylogenetically grouped by catalytic properties such as substrate specificity or stereospecificity rather than by the assembly line of origin (section 2.2). While additional recombination events could explain this incongruity, in some cases (e.g., the avermectin synthase), the constituent modules would have had to undergo large-scale recombination in order for these assembly lines to have evolved by module duplication followed by recombination.⁶⁵

Finally, under the hypothesis that multimodularity of *cis*-AT PKSs evolved through module duplication, the phylogenetic tree of KS domains itself is discordant.⁶⁴ This is exemplified by a set of homologous PKSs producing 16-membered macrolides (Figure 3B). If module duplication preceded the diversification of the resulting assembly-line PKS into different homologous clusters, one would expect KS domains to be more distant from paralogous KS domains within the same PKS than from their orthologues. The phylogenetic tree shows a different pattern: for many PKSs, their paralogous KS domains have the highest sequence similarity. This discordance can be explained by extensive gene conversion between paralogous KSs: this rate has been estimated at 27%, and has been shown to result in a concerted evolution of PKS modules (Figure 3C).^{64,82} If that is indeed the case, then the high sequence similarity between paralogous KS domains is the result of recent gene conversion events, rather than ancestral gene duplication that occurred during the emergence of assembly-line architecture.

2.4.3. Alternative Model. To resolve these discordances, we propose an alternative model for assembly-line PKS evolution that applies to both *cis*-AT and *trans*-AT PKSs (Figure 3D). Our model is based on the premise of extensive gene conversion between paralogous KS domains within the same *cis*-AT PKS, leading to repetitive regions of abnormally high sequence similarity within the same assembly line.^{64,82} This would allow for an evolutionary process that is entirely analogous to the mosaic-like assembly proposed for *trans*-AT PKSs without the need to invoke extensive gene duplications.³⁴ In addition to presenting a simpler logic for *trans*-AT PKS evolution from *cis*-AT PKSs via the loss of AT domains, this model would also explain the absence of iterative *trans*-AT PKSs, the presence of AT domain remnants in many *trans*-AT PKSs, and the existence of assembly-line PKSs (e.g., the NOCAP synthase discussed below) that contain modules of both classes. The hypothesis of *trans*-AT PKS evolution through displacement of *cis*-AT PKS domains is also supported by phylogenetic evidence in algae.⁸³ Of course, further support for such a model would require clearer evidence for the role of gene conversion mechanisms in the evolution of assembly-line PKSs.

2.4.4. Model for Evolutionary Unit of an Assembly-Line PKS. Historically, the functional unit of a PKS was called a module: a polypeptide containing KS-AT-(DH-KR-ER)-ACP

domains and able to perform one round of polyketide chain elongation and elaboration.^{85,86} It is also an architectural, and hence genetic, unit: this domain order is conserved across vertebrate FASs, iterative PKSs, and *cis*-AT PKSs. However, it is unclear whether this genetic unit also corresponds to an evolutionary unit that has been preserved in multimodular PKSs. Each KS domain of an assembly-line PKSs must interact with the ACP domain of its upstream module during chain translocation as well as the ACP domain of its own module during chain elongation; both reactions require specific protein–protein interactions (Figure 1A).^{70,87} Genetic recombination between homologous modules can be expected to scramble one of these interfaces while preserving the other.

The KS domains of *trans*-AT PKSs appear to have coevolved with their ACP partners from upstream modules.³⁷ Their evolutionary relationships also appear to be correlated to structural similarities between their substrates, as defined by the enzymatic domains observed in the reductive loops of upstream modules.^{34,37} This suggests that the canonical evolutionary unit of *trans*-AT PKSs is the (DH-KR-ER)-ACP-KS domain sequence, which would preserve the chain translocation interface.

In contrast, the evolutionary history of KS domains of *cis*-AT PKSs is obscured by two factors. First, they show lower specificity toward their substrates.⁸⁸ Second, gene conversion events discussed above mask some of the evolutionary history of *cis*-AT PKSs. Nonetheless, a recent analysis of aminopolyol PKSs has revealed coevolutionary relationships between KS domains and processing enzymes from upstream modules, suggesting that a typical evolutionary unit is either (DH-KR-ER)-ACP-KS-AT or AT-(DH-KR-ER)-ACP-KS.³⁸ While it remains unclear whether the evolutionary comigration of KS domains and ACP domains of upstream modules generalizes to all *cis*-AT PKSs, this hypothesis is supported by the observation that the post-AT linker may be a functionally effective splice point for natural recombination as well as evolutionarily inspired PKS engineering^{89,90} (discussed in section 5.2.1).

These observations have led to a proposed redefinition of module boundaries from the “classical” KS-AT-(DH-KR-ER)-ACP toward “alternative” AT-(DH-KR-ER)-ACP-KS.^{37,39} While these boundaries most likely correspond to the evolutionary unit of assembly-line PKSs, they are different from the functional, architectural, and genetic unit defined by the “classical” module boundaries. More research is warranted before this new definition can be universally accepted.

2.5. Factors Influencing the Evolution of Assembly-Line PKS Diversity

While the emergence of the earliest functional assembly-line PKSs undoubtedly set the stage for their subsequent diversification through mutation, HGT, gene conversion, and recombination, a general understanding of these molecular processes cannot explain the tremendous phenotypic diversification that subsequently emerged within this PKS family. To do so more satisfactorily, these processes have to be put into the context of environmental and genetic factors and considered from the perspective of evolutionary advantages that they provide.

2.5.1. Environmental Factors. Many microorganisms produce a vast array of secondary metabolites whose biological roles in nature are not yet understood.^{91–96} For example, polyketide natural products are produced by organisms dwelling in diverse environments ranging from soil to marine

and fresh water, from free-living to symbiotic or parasitic systems.^{97–100} These environmental factors presumably contributed to shaping the structural diversity and biological activity of polyketide natural products; however, our understanding of the connections between microbial ecology and natural product biosynthesis is still emerging and will therefore not be discussed here.

2.5.2. Genetic Factors. The genetic factors influencing the evolution of assembly-line PKSs are also not well understood. In prokaryotes, assembly-line PKSs are mainly confined to actinobacteria, proteobacteria, firmicutes, and cyanobacteria, with an uneven distribution among bacterial groups within each phylum.^{28,101,102} The distribution of the two types of PKS assembly lines is also nonhomogeneous: *cis*-AT PKSs are most common in actinobacteria, cyanobacteria, and proteobacteria, whereas *trans*-AT PKSs are more widespread in proteobacteria and firmicutes.³⁴ The evolutionary rationale for this uneven distribution is also unclear.

Actinobacteria and especially *Streptomyces* are by far the most prolific producers and often harbor multiple PKS clusters in their genomes. The study of their genomes revealed several key points that have likely contributed to the diversity of their natural products. First, *Streptomyces* contain numerous plasmids, integrative and conjugative elements, and genomic islands that carry biosynthetic clusters and can increase the rate of their horizontal gene transfer.^{103–105} Identification of gene clusters on these mobile genetic elements highlights their biological relevance in horizontal gene transfer.¹⁰⁶ Second, their genomes favor the formation and recombination of multiple biosynthetic gene clusters: *Streptomyces* chromosomes are large (6–12 Mb), linear, and unstable. PKS clusters can span several hundreds of kilobases, and genome size scales almost linearly with the number of PKS clusters, suggesting that larger genomes are more likely to contain multiple clusters.¹⁰⁷ The linear structure and the instability of *Streptomyces* chromosomes contribute to the overall genomic plasticity that involves frequent HGT, recombination, gene duplication, and deletion.^{108,109} Third, the GC content of DNA is highly correlated with recombination frequency in different organisms, even though the causality of these effects is not entirely clear.¹¹⁰ *Streptomyces* are no exception to this rule, and their high GC content (>70%) is matched by a high recombination rate.

PKS diversification in cyanobacteria has also been attributed to HGT, recombination, gene duplication and deletion, but no specific genetic trait can explain the observed diversity of secondary metabolites in this phylum.^{111,112} Even less is known about the genetic factors that contribute to PKS diversification in other bacteria, and more research would be needed to elucidate the underlying molecular mechanisms.

2.5.3. Evolutionary Advantages. Two conceptually different perspectives exist on the role and diversification of natural products.²¹ According to a more traditional viewpoint, the evolutionary advantage conferred by the function of the molecule constitutes the trait under selection.¹¹³ Here, every molecule produced by a biosynthetic cluster must have an advantageous biological activity to justify the metabolic cost of its production and to be selected for. The alternative model, also referred to as the “screening hypothesis”,^{114,115} presumes that the selected trait is the adaptability itself, i.e., the capacity to generate and maintain the chemical diversity of secondary metabolites that can be screened for advantageous properties when needed. This model does not require all molecules to

have a beneficial function, so long as a few molecules provide enough advantage to maintain the entire system.

The practical implications of the two models for natural product chemistry are quite distinct. The first model implies that screening for bioactive molecules holds great promise for the discovery of novel molecules of therapeutic interest. On the other hand, the second model anticipates that most natural products do not have measurable bioactivity and that a large library would be needed to screen for new therapeutics. While the available body of knowledge is insufficient to provide conclusive evidence, the difficulty in finding compounds with measurable bioactivity suggests that the screening hypothesis may be more realistic. However, this hypothesis also suggests that the mechanisms that create diversity are remarkable and that their success rate is sufficient for their presence to be selected for in bacterial genomes. This would imply that these same mechanisms can be leveraged to generate diversity in the laboratory and open new avenues for assembly-line PKS engineering (discussed in section 5). Determining the precise order of events leading to the appearance of contemporary assembly-line PKSs would be an extremely challenging task. However, insights gained from this research can inform us about the best strategies to pursue in the future evolutionary-inspired engineering approaches.

3. DIVERSITY OF ORPHAN PKSS

As discussed above, the mechanisms and selective pressures involved in the evolution of assembly-line PKSs have led to astounding polyketide diversity. In this section, we will present computational approaches for estimating this natural diversity, including an updated catalogue of assembly-line PKSs found in the NCBI database. The number of novel clusters sequenced every year reflects the vastness of the PKS sequence space and the extent to which polyketide structural diversity is underexplored.

3.1. Catalogues of Assembly-Line PKSs

Biosynthetically characterized PKSs have been catalogued in a variety of databases. For example, CSDB,¹¹⁶ ClusterMine360,¹¹⁷ SBSPKS v2,¹¹⁸ and DoBISCUIT¹¹⁹ include manually curated lists of 150–300 known microbial PKSs and NRPSs, including many assembly-line PKSs. The more recent MIBiG repository is the result of a community effort to facilitate the standardized deposition and retrieval of BGCs responsible for making known natural products. As of August 2018, MIBiG includes over 250 assembly-line PKSs, including PKS-NRPS hybrids.¹²⁰

However, BGCs that make known polyketides only offer a narrow glimpse into the diversity of assembly-line PKSs. A powerful approach to evaluate the actual diversity of PKS clusters and their products is through computational analysis of sequence databases. Algorithms such as antiSMASH,¹²¹ ClusterFinder,¹²² PRISM,¹²³ and others (reviewed in refs 124,125) allow users to mine sequenced data for microbial BGCs and predict the biosynthesized product. Additional algorithms can improve predictions for certain types of clusters: for example, NaPDoS uses domain phylogeny to predict PKS and NRPS products,¹²⁶ while the TransATor allows a more accurate prediction of *trans*-AT PKS products based on the substrate specificity of their KS domains.¹²⁷ Despite significant advances of in silico prediction algorithms, determining the structure of a polyketide from its BGC sequence alone remains an elusive goal.

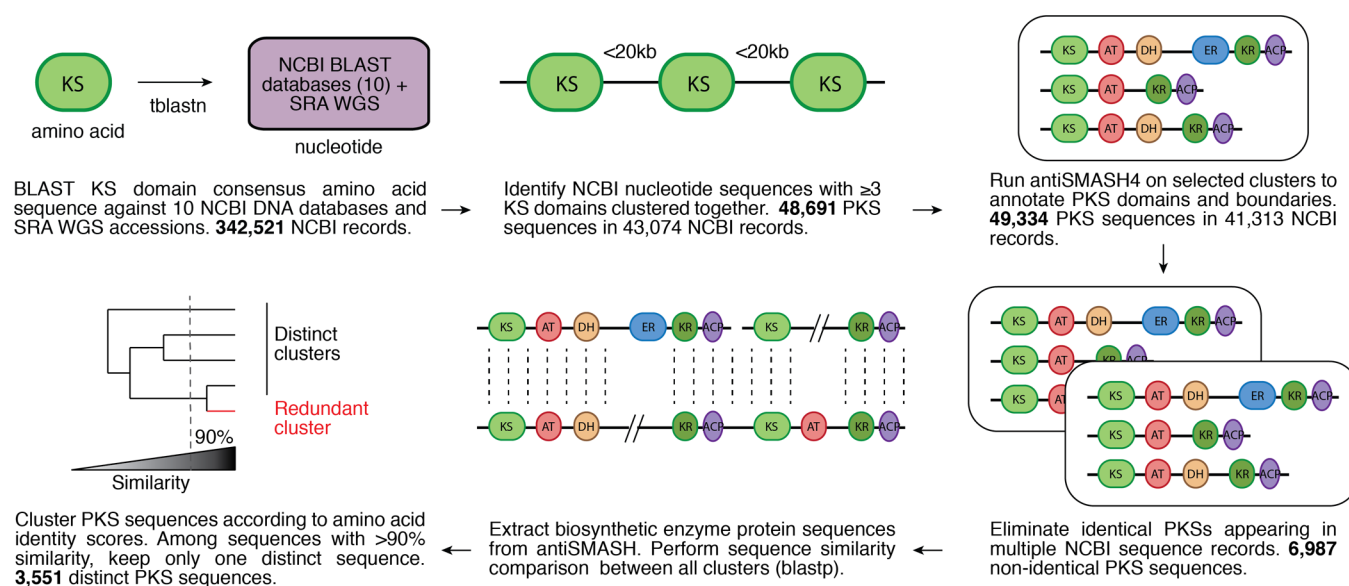


Figure 4. Summary of the workflow to generate the catalogue of distinct assembly-line PKSs. In the final clustering schematic, the red line represents a PKS sequence that scored higher than 90% in amino acid similarity to another sequence and was thus removed from the catalogue of distinct clusters.

Nonetheless, computational analysis of BGCs can be used to estimate the diversity of assembly-line PKSs and their products. AntiSMASH is a particularly powerful and widely used tool for identifying and annotating bacterial BGCs,¹²¹ with many additional functionalities becoming available in each new release. (antiSMASH 5.0 is the most recent one.¹²⁸) Several other databases contain data from large-scale genome mining, such as IMG-ABC (~150 PKSs sequenced at the Joint Genome Institute)¹²⁹ and antiSMASH database 2.0 (over 3000 PKSs from publicly available microbial genomes).¹³⁰ In 2013, we catalogued all nonredundant assembly-line PKSs available in the NCBI databases¹⁰² and identified 885 nonredundant PKSs, most of which produced unknown compounds. (These uncharacterized PKSs were referred to as “orphans.”) Given the rapidly increasing number of genomes deposited into sequence archives, we have updated this catalogue to obtain a snapshot of assembly-line PKSs sequenced to date.

3.2. Updated Catalogue of Orphan Assembly-Line PKSs

The general strategy for compiling and phylogenetically analyzing all assembly-line PKSs has been described previously.¹⁰² Briefly, a consensus ketosynthase (KS) sequence was aligned using BLAST against nine NCBI DNA databases as well as the archive for whole-genome shotgun sequences available as of May 2018. To select for multimodular PKSs, BLAST hits were refined by requiring a minimum of 3 KS domains located within 20kb of each other, and the PKS gene clusters that met this criterion were further analyzed by antiSMASH 4.0.¹³¹ Identical PKSs were eliminated based on either an identical sequence or an identical domain architecture in the same species. From the remaining PKSs, the sequences of individual PKS and NRPS proteins were extracted and subjected to comparative pairwise analysis using BLAST, calculated as described in ref 102. PKSs that scored more than 90% in amino acid similarity were considered redundant, yielding the final catalogue of distinct assembly-line PKSs (Figure 4). As before, cluster similarity scores were visualized in the form of a dendrogram.

A total of 3551 distinct clusters from 1662 species were catalogued, representing a 4-fold increase over the data set from five years prior. Among these, 1692 clusters were annotated as *cis*-AT PKS clusters, 975 as *cis*-AT PKS/NRPS hybrids, 293 as *trans*-AT PKSs, 343 as *trans*-AT PKS/NRPS hybrids, and 248 as other hybrids. The full list of non-redundant assembly-line PKS clusters and the dendrogram visualizing their distances are available online at http://web.stanford.edu/group/orphan_pks/. It should be noted that, although our number of PKS clusters closely matches the number listed in the antiSMASH 2.0 database (3302 type I PKSs and 623 *trans*-AT PKSs),¹³⁰ the two catalogues are complementary, not identical, because the analyses differed in terms of NCBI databases, PKS cluster types, and sequence similarity cutoffs. Nonetheless, both of them reflect the vast numbers of assembly-line PKS clusters present in nature.

3.3. Evaluating the Product Diversity of Orphan Assembly-Line PKSs

On the basis of the date when each PKS sequence in our catalogue was deposited in the NCBI database, it appears that the number of distinct assembly-line PKSs continues to grow exponentially, doubling every 2.5 years (Figure 5A, blue bars). This rate of discovery is consistent with the overall growth of NCBI sequencing data in GenBank. The vast majority of these clusters are orphan; by using the MIBiG database and NCBI annotations, we estimate that only around 10% of assembly-line PKSs in our catalogue have been linked to the production of a known molecule (Figure 5A, red bars).

A major challenge in traditional natural product discovery is the high rate of rediscovery of a given molecule. Even among sequences deposited into NCBI databases, the number of redundant assembly-line PKS clusters has been increasing, reaching 51% by mid-2018 (Figure 5B). As one continues exploring PKS diversity, this will likely become even more problematic.

The number of distinct assembly-line PKSs is astonishing in itself. However, it is even more interesting to consider the diversity of these clusters and their products. By eliminating

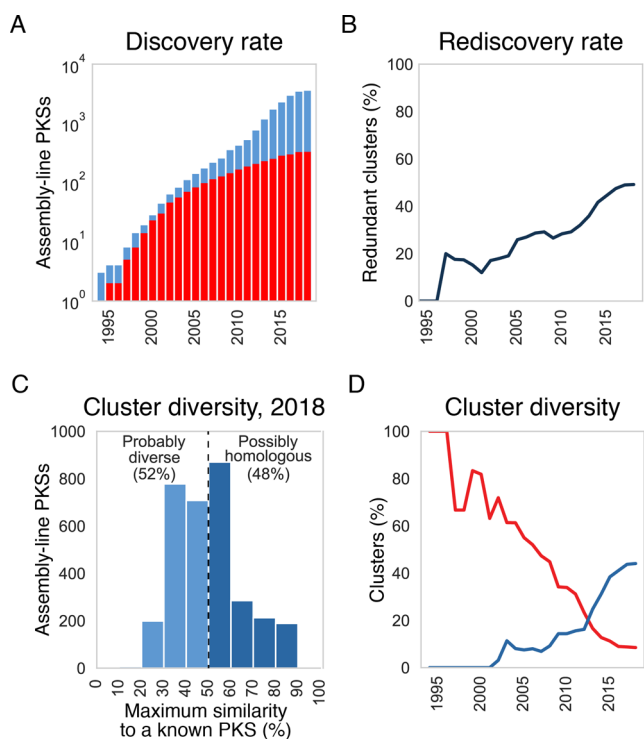


Figure 5. (A) The discovery rate of distinct clusters is shown (blue; having less than 90% amino acid sequence similarity score to any other cluster). Also shown (in red) is the number of clusters with known products, determined using MIBiG database and NCBI annotations. For years 1994–2017, numbers reflect sequences deposited by December of that year. For 2018, only sequences deposited by May were taken into account. (B) Rediscovery rate among nucleotide sequences deposited to NCBI, determined as the percentage of redundant clusters (having more than 90% amino acid sequence similarity score to a previously sequenced cluster). (C) Distribution of sequence similarity scores between an orphan assembly-line PKS and its closest neighbor whose product has been characterized. PKSs with pairwise similarity scores above 50% probably make structurally similar polyketides, while orphan PKSs whose sequences show greater differences from those of any known PKS most likely produce novel chemotypes. (D) The red line plots the percentage of all distinct assembly-line PKSs that are chemically decoded. The blue line plots the percentage of orphan PKSs that are more than 50% similar to a chemically decoded assembly-line PKS.

redundant clusters (above 90% similarity score) from the catalogue, we sought to estimate the number of clusters producing different molecules. On the basis of similarities between nine 16-membered macrolide PKSs (46–89% similar, with a mean of 56%), we assume that assembly lines with higher than 50% similarity could make identical or very similar molecules. (As a point of reference, the tylactone and rosamicin synthases are 72% similar and produce the same polyketide backbone.) It should be noted that the tailoring enzymes associated with a given biosynthetic cluster differ even for PKSs with high sequence similarity and give rise to distinct natural products. Nonetheless, on the basis of the above arguments, we assume that PKSs that are less than 50% similar most likely produce polyketide products that could be regarded as distinct chemotypes.

By evaluating the maximum sequence similarity of orphan assembly-line PKSs to any previously characterized PKS, it is possible to estimate how diverse their products are from known polyketide natural products. Remarkably, more than

one-half of all orphan assembly lines show less than 50% sequence similarity to any known PKS (Figure 5C). Although the rate of chemically decoding orphan assembly-line PKSs cannot possibly keep up with their discovery (Figure 5D, red line), it appears that the fraction of orphan PKSs making polyketides whose structures are related to known natural products is increasing (Figure 5D, blue line). This fraction, however, is likely an overestimate because only ca. 20% of the emerging orphan PKSs preserve the same architecture over the entire assembly line. (Most of the orphan PKSs that comprise the blue line statistics in Figure 6B share their assembly-line architecture with a substantial portion, but not all, of a characterized PKS.) Nonetheless, the overall upward trend suggests that, while modern genomics-driven natural products discovery may be steadily sampling the actual diversity of assembly-line PKSs in nature, the major part of this diversity has not yet been explored.

3.4. Similarity Network of Assembly-Line PKSs

The diversity of assembly-line PKSs can be visualized as a network (Figure 6). Sequence similarity networks are a useful tool for analyzing relationships within a protein family.¹³² Individual PKS sequences are represented as nodes (circles), while pairs of PKSs with sequence similarity above a certain threshold are shown as edges (lines) where an edge's length correlates with the relative dissimilarity between the PKS pair. (The relative position of disconnected groups has no meaning.) Unlike dendrograms that only show optimal connections, networks allow visualization of all relationships above a threshold. In our analysis, the threshold of pairwise cluster similarity was 50%, and networks of distinct assembly-line PKS were visualized using Cytoscape 3.7.2.¹³³ Orphan PKS nodes (smaller circles) not connected to any node corresponding to a characterized PKS (larger circle) highlight the unexplored diversity of assembly-line PKSs.

From this data, it is apparent that *cis*-AT PKSs (red) and PKS/NRPS hybrids (orange) separate from *trans*-AT PKSs (dark blue) and PKS/NRPS hybrids (light blue). This may be due to their nonuniform distribution among bacterial phyla. Indeed, the main group in the top left corner almost exclusively comprises actinobacterial PKSs (regardless of subclass), whereas the two large groups to its right is comprised of cyanobacterial and firmicute PKSs, respectively (Supporting Information, Figure S2). A few examples of PKSs with >50% similarity are found in species belonging to different phyla, supporting the theory that even though HGT has played an important role in assembly-line PKS evolution, it does not occur frequently between phyla.²⁰

Overall, these networks reveal promising opportunities for the exploration of polyketide diversity in nature. On one hand, orphan PKSs belonging to a large, tightly connected network that includes at least one known PKS may warrant investigation, as they could yield natural products with related properties. On the other hand, by exploring a disconnected group of orphan PKSs, one could discover truly novel polyketide structures and bioactivities. Such disconnected groups include, for example, PKSs that biosynthesize the DNA chelator colibactin,¹³⁴ the antimetabolic agent rhizoxin,¹³⁵ and the pre-mRNA splicing inhibitor FR901464.¹³⁶

3.5. Eukaryotic PKS Clusters

While a vast majority of chemically decoded assembly-line PKSs are from bacterial sources, it has become clear within the past decade that eukaryotic genomes also encode a number of

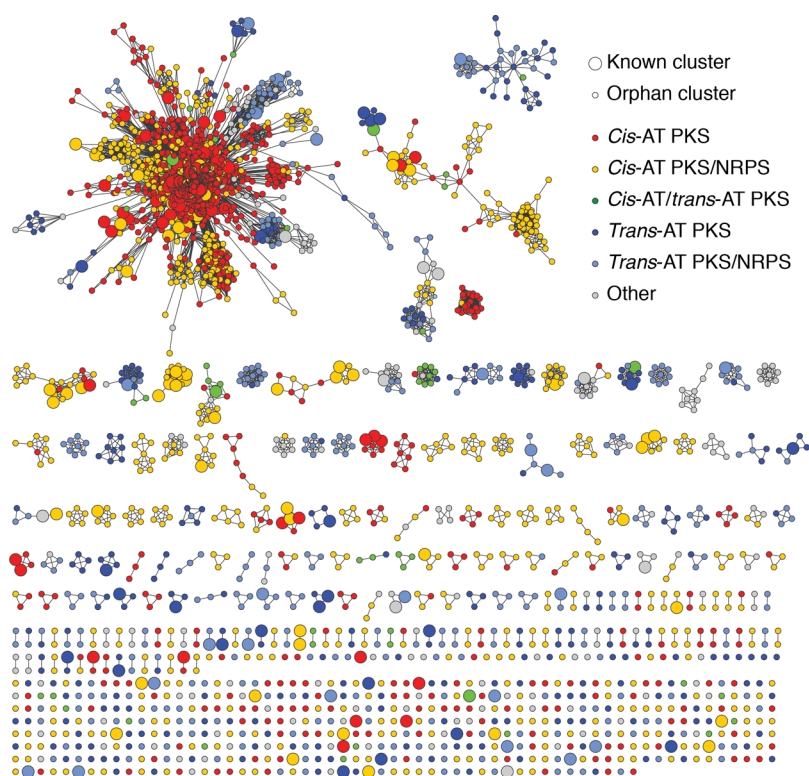


Figure 6. Network of 3551 distinct assembly-line PKS clusters, visualized by Cytoscape 3.7.2.¹³³ Nodes correspond to known (larger circles) and orphan (smaller circles) PKSs and are color-coded according to antiSMASH predictions (legend). Edges represent >50% sequence similarity between two clusters, calculated as described in ref 102.

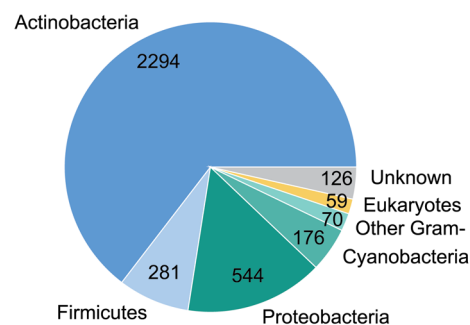
these megasynthases (Figure 7A). When evolutionary relationships were visualized on a dendrogram or in a network, 59 distinct eukaryotic assembly-line PKSs clustered across several groups (Supporting Information, Figure S2 and the dendrogram available online).

One such group includes assembly-line PKS from nematodes, originally identified as an orphan PKS.¹⁰² More recently, the hybrid PKS-NRPS from *Caenorhabditis elegans* has been decoded as a producer of the nemamide family of natural products (Figure 7B).¹³⁷ These remarkable molecules are regulators of starvation-induced larval arrest.

Another cluster of eukaryotic assembly-line PKSs is found in soil-dwelling social amoeba from the *Dictyostelium* genus. Genome sequencing has revealed more than 40 PKSs in *Dictyostelium discoideum*.¹³⁸ So far, only iterative PKSs have been characterized from these species,^{139–141} and the chemistry and biology of polyketide products of *Dictyostelium* assembly-line PKSs remain unknown.

Assembly-line PKSs are also found in various eukaryotic protists. Apicomplexan parasites such as *Cryptosporidium*, *Toxoplasma*, and *Eimeria* contain assembly-line PKSs that appear to produce fatty acid components of the rigid wall of their oocysts, thereby ensuring transmission of the pathogen between hosts.^{142–144} These protists also encode assembly-line PKSs that appear to produce more oxygenated metabolites of unknown structure.¹⁴⁵ Related PKSs are found in the dinoflagellate *Gambierdiscus*. Dinoflagellates possess some of the largest genomes among eukaryotes, and very few whole-genome sequences are available in the NCBI database. However, transcriptomic analyses have revealed numerous PKSs in dinoflagellates and suggest that these marine protists are a large reservoir of these enzymes.^{146,147} Their expression

A. Cluster distribution among phyla



B. Nemamide biosynthetic cluster in nematodes

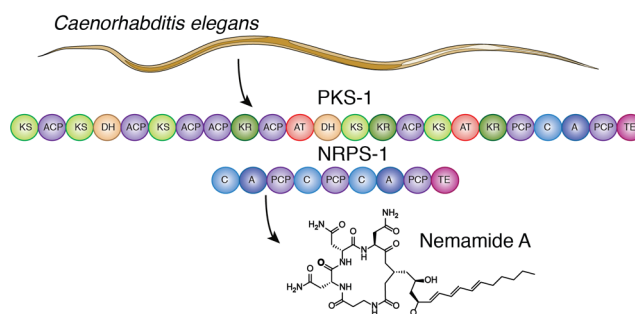


Figure 7. (A) Distribution of assembly-line PKSs among the different phyla. (B) The nemamide PKS from *C. elegans*, described in ref 137. KS, ketosynthase; AT, acyltransferase; KR, ketoreductase; DH, dehydratase; C, condensation domain; A, adenylation domain; ACP, acyl carrier protein; PCP, peptidyl carrier protein; TE, thioesterase.

in *Gambierdiscus* has been linked to polyether toxins released during algal blooms.¹⁴⁸

Other eukaryotic species harboring assembly-line PKSs include phytopathogenic fungi, fish, arthropods, and mollusks. So far, the evolutionary history of eukaryotic assembly-line PKSs remains cryptic. It is possible that their patchy occurrence reflects a loss of this PKS family from most eukaryotic lineages. Alternatively, eukaryotic PKSs could have been acquired from prokaryotes during secondary endosymbiosis or resulted from more recent interkingdom HGT events.⁸³ Regardless, the diversity of molecules synthesized by eukaryotic assembly-line PKSs and their relevance to host development or pathogenicity suggest that they represent an underexplored source of bioactive natural products.

3.6. Prioritizing PKS Clusters for Further Study

In general, bioinformatic decoding of orphan assembly-line PKS chemistry is outside the realm of feasibility today, making experimental analysis a necessity. Given the abundance and the diversity of orphan PKSs, methods for prioritizing clusters for deorphanization are crucial.¹⁴⁹

The classic method, consisting of screening for biological activity in the native host, remains laborious and technically challenging¹⁵⁰ but has nonetheless benefitted enormously from state-of-the-art untargeted metabolomics approaches. A vivid example can be found in the work leading to the discovery of nemamides (Figure 7B).¹³⁷ Alternatively, culture-independent methods are also being used for compound prioritization,¹⁵¹ and computational tools are becoming increasingly helpful in such pursuits.¹⁵² Ultimately, given the enormous gap between the pace of discovering orphan PKS assembly lines and their deorphanization, selecting a target orphan PKS for further analysis is a subjective exercise.

To help genome mining approaches navigate this diversity in search of the most interesting and novel clusters, several computational approaches are being developed. On one hand, cluster prioritization can be based on the novelty of the product's chemical structure, often reflected by the orphan cluster's evolutionary relationships with known BGCs. On the protein level, EvoMining reconstructs evolutionary histories of biosynthetic enzymes in an attempt to find clusters that produce molecules with novel chemical structures, the so-called "chemical dark matter".^{153,154} At the cluster level, the combination of BiG-SCAPE and CORASON tools opens the possibility to analyze BGC similarity networks and cluster group phylogenies to direct genome mining approaches either toward the discovery of molecules with novel structures or explore known compound analogues.¹⁵⁵ On the other hand, cluster prioritization can be based directly on the nature of enzymes present in the cluster. For instance, the ARTS tool predicts clusters that are more likely to produce a molecule with an antibiotic activity based on the presence of self-resistance genes: BGC-encoded genes that are homologous to the antibiotic target gene yet harbor mutations that confer resistance.¹⁵⁶

However, in silico prioritization of clusters is only the first step toward deorphanizing a BGC through experimental approaches. We recently identified a distinct clade of NOCAP (nocardiosis-associated polyketide) synthases only observed in 12 clinical strains of *Nocardia* isolated from nocardiosis-affected patients. Using both direct in vitro reconstitution from purified proteins and *Escherichia coli* as a heterologous host for polyketide biosynthesis, we characterized

an unprecedented set of polyketides (ref 157, and Yuet et al., in preparation). While this example validates the utility of a combined in vitro and in vivo approach to deorphanize assembly-line PKS clusters identified through in silico analysis, it also highlights the importance of a careful choice of targets.

4. ACCESSING PKS DIVERSITY

In the previous section, we discussed the diversity of orphan assembly-line PKSs and introduced potential strategies for prioritizing these promising sources of new natural products for further analysis. However, the process of producing a novel polyketide and determining its structure heavily relies on wet laboratory techniques. Genetic manipulation and chemical analyses are at the core of the efforts to explore the natural diversity of polyketide compounds. In this section, we delve deeper into state-of-the-art methods for connecting orphan assembly-line PKSs to their natural products.

If an organism encoding an orphan PKS can be cultured and genetically manipulated, then promoter mutagenesis followed by metabolic profiling can enable natural product discovery. Alternatively, heterologous expression of an entire biosynthetic pathway in a well-established host, as in the case of the NOCAP synthase, can achieve the same goal. Here, we briefly review these two approaches.

4.1. Expressing Assembly-Line PKSs in Heterologous Hosts

Heterologous hosts such as *E. coli* have significant genetic and growth advantages over native hosts, thus allowing expression of BGCs from unculturable organisms. The most widespread approach to transferring BGCs in a heterologous host is direct cloning (Figure 8). Because of the difficulty of handling large DNA fragments, researchers have developed tools based on homologous recombination to precisely capture the BGC of interest.

4.1.1. Phage Recombination-Assisted Cloning. Direct cloning of PKS gene clusters can be accomplished in *E. coli* with the assistance of phage recombination systems such as phage lambda-derived Red¹⁵⁸ and phage Rac-derived RecET.¹⁵⁹ For example, *Photorhabdus luminescens* TT01 harbors 10 unexplored secondary metabolic pathways. By

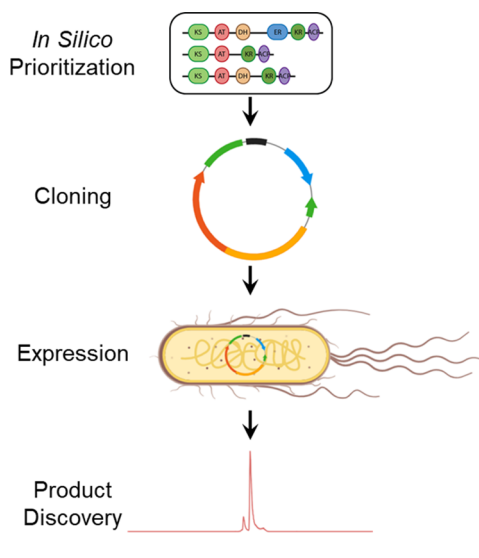


Figure 8. A general workflow for expressing assembly-line PKSs in heterologous hosts.

using the full length RecET in *E. coli*, these BGCs (10–52 kb) were recombined onto pSC101-based expression vectors. Seven gene sets were cloned successfully, two of which were expressed in *E. coli* Nissle 1917, leading to the identification of new metabolites luminmycin A and luminmide A/B.¹⁶⁰ The use of RecET also enabled heterologous production of disorazol in *Myxococcus xanthus*¹⁶¹ and salinomycin in *Streptomyces coelicolor*.¹⁶² A cryptic hybrid PKS-NRPS from *Paenibacillus lavae* was cloned and activated in *E. coli*, leading to the production of a novel compound, sevadicin.¹⁶³

4.1.2. Transformation-Associated Recombination Cloning. Transformation-associated recombination (TAR) techniques exploit homologous recombination in *Saccharomyces cerevisiae* to rapidly “capture” large gene clusters directly from genomic DNA.^{164–166} For example, researchers studying marinopyrrole biosynthesis in *Streptomyces* sp. CNQ418 found that TAR could enable direct cloning within days while phage-mediated homologous recombination methods such as λ Red/ET recombineering have turnaround times of months.^{167,168}

A shuttle vector pTARA was developed containing three components for shuttling among three organisms: yeast, *E. coli*, and *Streptomyces*. CEN6 (centromere in chromosome VI) and ARS4 (autonomously replicating sequence 4) sequences as well as a URA3 selection marker allow for gene cluster assembly and propagation in *S. cerevisiae*. Bacterial artificial chromosome elements and a chloramphenicol resistance cassette allow for maintenance and verification in *E. coli*. An apramycin resistance cassette and the phage ϕ C31 integration system enable site-specific chromosomal integration of the cluster in a number of different *Streptomyces* strains, including *Streptomyces toyocaensis*, *Streptomyces lividans*, and *Streptomyces albus*.¹⁶⁹ Using pTARA, these investigators directly cloned the 56 kb colibactin biosynthetic gene cluster from *Citrobacter koseri*, a gut bacterium. Multiple biosynthetic gene clusters, including an 89 kb orphan NRPS gene cluster, were also directly cloned or reassembled from cosmid DNA libraries.

4.1.3. pCAP-Based Transformation-Associated Recombination Cloning. More recently, the TAR cloning strategy has been adapted onto pCAP01, a shuttle vector.¹⁷⁰ Unlike pTARA, pCAP01 can be maintained at a higher copy number in *E. coli*, even with large (>50 kb) inserts. In addition, the ϕ C31 integration elements in pCAP01 allow its site-specific integration into chromosomes of a broader range of heterologous actinobacteria.¹⁷¹ Using λ -Red recombination-based methods, the 30 kb marinopyrrole and the transcriptionally silent 67 kb orphan *tar* BGC were cloned from *Streptomyces* sp. CNQ418 and expressed in *Streptomyces coelicolor* M152, leading to heterologous production of marinopyrrole and taromycin A, respectively.¹⁶⁸ Despite its relatively rapid workflow, this method requires a considerable amount of colony screening due to high levels of unproductive pCAP01 recircularization by nonhomologous end joining, resulting in capture rates below 2%.

To minimize plasmid recircularization by nonhomologous end joining, the URA3 gene encoding the *S. cerevisiae* orotidine 5'-phosphate decarboxylase was introduced into pCAP01 as a counter-selectable marker, yielding pCAP03.¹⁷² This vector was used to capture and express in *Streptomyces coelicolor* M1152 thiotetronic acid-producing 22 kb PKS/NRPS biosynthetic gene clusters from *Salinispora pacifica* CNS-863 after screening only 12 transformants (eight of which were positive; 75% capture rate) and *Streptomyces afghaniensis* after screening 10 transformants (two of which were positive; 20%

capture rate). The method has been extended to capture clusters associated with the production of amicoumacin¹⁷³ and colibactin.¹⁷⁴ Primary limitations to the use of pCAP01 and pCAP03 include a requirement for restriction enzymes that cut at sites flanking, but not within, a biosynthetic gene cluster of interest, and the need for high-quality high-molecular weight genomic DNA to capture clusters larger than 50 kb. Another limitation inherent in all TAR-based methods involves the relatively slow growth rates of yeast.

4.1.4. Cas9-Assisted Targeting of Chromosome (CATCH) Cloning. To address the above challenges, a Cas9-assisted targeting of chromosome (CATCH) cloning strategy was developed.^{175,176} In this method, bacteria are embedded in low-melting-temperature agarose gel, treated with lysozyme and proteinase K, and washed to yield high-quality high-molecular weight genomic DNA stabilized by agarose. The genomic DNA is then cleaved with CRISPR-Cas9 endonuclease directed by guide RNAs to digest specific sequences flanking the cluster of interest, bypassing the need for restriction enzymes. Avoiding homologous recombination in yeast altogether, the genomic DNA fragments are recovered by digestion with agarase, purified, and ligated into vectors with homologous 30 bp arms by Gibson assembly.¹⁷⁷ The reaction mixture is electrotransformed into *E. coli*. CATCH is rapid, taking ca. 8 h effort over several days, and yields positive clones varying from 20% (for 100 kb test inserts) to 60% (for 50 kb test inserts). The 78 kb bacillaene assembly-line PKS from *Bacillus subtilis* was successfully cloned after screening 102 transformants (12 of which were positive; 12% capture rate).

4.1.5. Site-Specific Recombination Cloning. Another direct cloning strategy based on the site-specific recombinase system Cre/loxP has been developed for assembly-line PKSs.¹⁷⁸ First, loxP sites are integrated flanking the gene cluster of interest with elements needed for plasmid replication. Then the Cre recombinase is expressed, and the whole region containing the gene cluster flanked by loxP is circularized as a plasmid. The resulting plasmid is isolated via transformation into *E. coli*. A 78 kb DNA fragment containing a siderophore biosynthetic gene cluster from *Agrobacterium tumefaciens* C58 was cloned with this strategy. An analogous method involving one less electroporation or conjugation step, based on Φ BT1 integrase-mediated recombination was used to clone the entire 55 kb erythromycin BGC.¹⁷⁹ Seven clones (out of a total of 20 *E. coli* colonies) selected for restriction enzyme verification harbored this BGC.

4.2. Activating Assembly-Line PKSs in Native Hosts

Some microorganisms harbor dozens of BGCs, many of which encode orphan assembly-line PKSs. For example, certain strains of *Streptomyces* are capable of producing as many as 50 distinct natural products.¹⁸⁰ However, many of these BGCs are tightly regulated.¹⁸¹ For organisms that are culturable and amenable to genetic manipulation, researchers rely on either overexpressing positive transcription regulators or deleting negative regulators to activate these normally silent BGCs.

For example, Bibb and co-workers identified a cryptic 29.5 kb gene cluster containing both modular type I and type III PKSs from *Streptomyces venezuelae* that was predicted to encode a biaryl metabolite, venemycin.¹⁸² However, both the native host and a heterologous *Streptomyces coelicolor* host harboring this cluster yielded insufficient venemycin for structural analysis. To overcome this challenge, they overexpressed *vemR*, a transcriptional activator from the ATP-binding

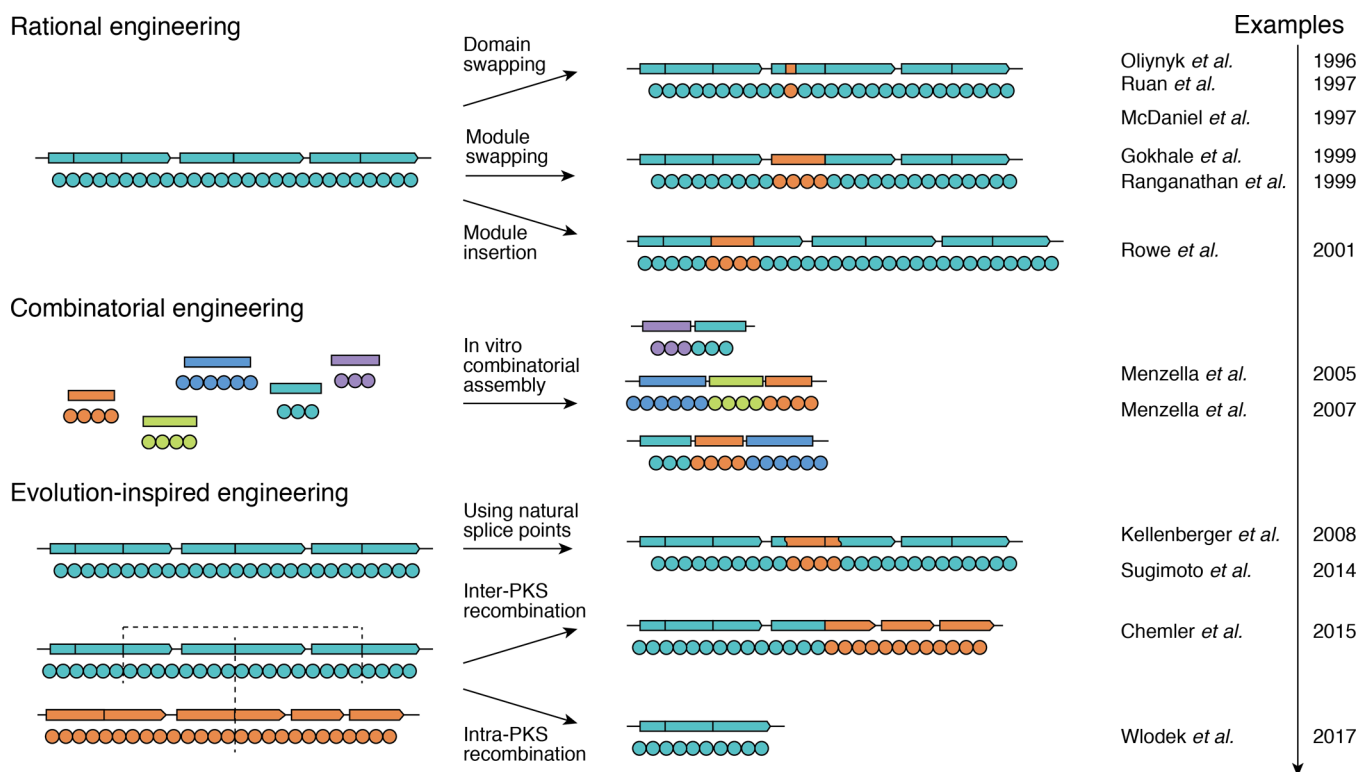


Figure 9. Over time, assembly-line PKS engineering has shifted from rational design (e.g., by domain swapping, module swapping, and module insertion) and combinatorial engineering (e.g., through in vitro combinatorial assembly) toward evolution-inspired approaches (e.g., use of natural splicing points, or inter- and intra-PKS recombination).

LuxR-like (LAL) family, with the constitutive promoter *ermE** in both strains, resulting in the production of adequate venemycin for structural characterization, confirming its unusual biaryl structure. Similarly, an orphan ansamycin PKS cluster was activated in *Streptomyces* sp. XZQH13 by constitutive expression of another LAL family regulator gene *astG1*, leading to the isolation of two known ansatrienins, hydroxymycotrienin A, and thiazinotrienomycin G.¹⁸³ Another orphan ansamycin PKS cluster was activated in *Streptomyces* sp. LZ35 by constitutive overexpression of a LuxR family transcriptional regulatory gene, leading to the discovery of three new naphthalenic ansamycins, neoansamycins A–C.¹⁸⁴ This approach can be further developed for high throughput activation of silent BGCs. In a step toward this direction, CRISPR/Cas9 methods have been used to delete genes¹⁸⁵ or knock-in promoters in *Streptomyces*.¹⁸⁶ Notably, a promoter knock-in strategy led to activate BGCs of different classes (type I, II, and III PKSs, NRPS, hybrid PKS-NRPS, and phosphonate) in multiple *Streptomyces* species.¹⁸⁶ Along similar lines, CRISPRi, which utilizes a catalytically dead Cas9 to interfere with gene expression in a sequence-specific manner, has been used to repress transcription of negative regulatory genes.¹⁸⁷ The efficacy of these strategies is pathway-specific. For example, if a BGC contains multiple operons, then overexpression of one activator or knock-in of a single promoter may not be sufficient for activation.

As an alternative to genetic manipulation, varying culture conditions such as media composition, aeration, culture vessel, and addition of enzyme inhibitors is sometimes sufficient to activate multiple biosynthetic gene clusters from a single strain. This “one strain–many compounds” (OSMAC) approach has been used to isolate more than 100 compounds (belonging to

more than 25 different structural classes) from only six different microorganisms: *Aspergillus ochraceus* DSM 7428, *Sphaeropsidales* sp. F-24'707, *Streptomyces* sp. Gö 40/14, *Streptomyces parvulus* Tü 64, *Streptomyces* sp. A1, and *Streptomyces* Tü 3634.¹⁸⁸ Ribosome engineering is another effective approach to globally activate BGCs.^{189–191} Strains of interest can also be cocultured with other organisms, resulting in interspecies crosstalk that acts to activate silent biosynthetic gene clusters.¹⁹² While the above methods are applicable to many hosts, the resulting physiological disturbances are global, making comparative metabolic profiling challenging.

5. EXPANDING PKS DIVERSITY THROUGH ENGINEERING

Since the discovery of assembly-line PKSs in the 1990s, numerous attempts to reprogram them have been explored, prompted by their modular architecture. PKS engineering promises to expand the polyketide diversity beyond the chemical landscape of natural compounds, for instance, to introduce small changes to the structure that would improve the molecule's bioactivity or bioavailability, much-sought results in medicinal chemistry. However, the task of PKS engineering is challenging. In this section, we will discuss how an understanding of the architecture and enzymatic reactions of assembly-line PKSs has gradually shifted engineering approaches from rational design toward evolutionary-inspired strategies (Figure 9).

5.1. Combinatorial and Rational Engineering of Assembly-Line PKSs

Early attempts at combinatorial assembly explored the possibility of generating libraries of “unnatural” natural products.^{193–195} However, it soon became clear that such

approaches are not straightforward: reordering domains or modules derived from naturally occurring PKSs often results in catalytically compromised assemblies.¹⁹⁶ Subsequent approaches in combinatorial engineering of PKSs using module swapping confirmed that most hybrid assemblies turned over poorly,^{197,198} which halted further efforts in this direction. More recently, a computational platform ClusterCAD was developed for streamlining the design of chimeric PKSs, potentially providing new opportunities for combinatorial polyketide biosynthesis.¹⁹⁹

In parallel to high-throughput combinatorial approaches, rational design strategies for accessing novel compounds were also explored. These included deletion, insertion, or replacement of intact domains and modules, engineering of substrate specificity, and metabolic supply of alternative precursors (reviewed in refs 200–202). However, these approaches also encounter difficulties in generating fully active PKSs.

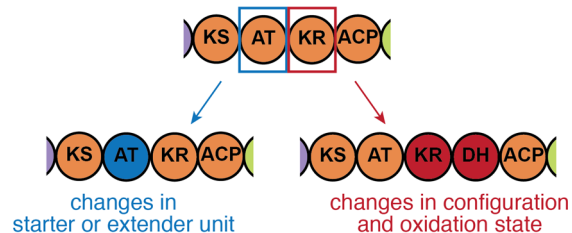
For nearly two decades, it has been clear that most of these difficulties stem from our inability to engineer essential protein–protein interactions involved in intramodule and intermodule chain processing and from insufficient knowledge about the specificity of different domains toward alternative substrates.²⁰³ Despite extensive structural and biochemical analysis,³ the mechanistic basis for the underlying dynamic protein–protein interactions remains poorly understood. Overcoming this challenge would be critical for further PKS engineering, both using rational and combinatorial approaches.

5.2. Engineering Inspired by Evolution

Given the difficulty of engineering PKSs in the laboratory, the catalytic diversity of natural assembly-line PKSs is all the more astonishing. The toolkit of molecular mechanisms and evolutionary strategies employed by nature appears to be much better suited for the challenge than the strategies typically used in the lab. The idea of taking inspiration from natural approaches for PKS engineering was proposed more than a decade ago⁶⁵ and was developed in two directions. One uses natural evolution to guide the choice of splice points for further engineering by traditional cloning techniques, while the other uses natural recombination mechanisms for generating novel PKSs (Figure 9).

5.2.1. Using Natural Splice Points. Because of the modular and colinear architecture of assembly-line PKSs, it is particularly appealing to engineer them by modifying single domains or modules to introduce small and predictable changes in the structure of the biosynthetic product. While point mutations can increase domain promiscuity or inactivate them, they rarely lead to altered domain function without compromising specificity or PKS turnover (reviewed in ref 204). It appears that PKS evolution did not rely on point mutations to change domain specificity either: domains with the same specificity are phylogenetically close (see section 2.2), suggesting that they originated from the same common ancestor rather than independently through point mutations. Instead, changes in domain specificity probably arise through domain swaps by recombination, which has prompted a search for natural splice points that can be exploited for engineering (Figure 10A, left). Because AT domains are responsible for selecting starter and extender units in polyketide biosynthesis, swapping them is an appealing strategy for product modification. AT domain swaps have been shown to yield functional chimeric PKSs as early as 1996.¹⁹⁴ Later studies revealed the presence of conserved regions in KS-AT

A. Domain swapping strategies



B. Module deletion strategies

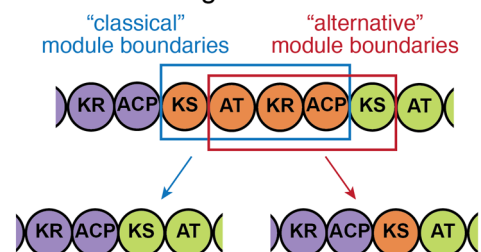


Figure 10. Alternative splice points for PKS engineering. (A) Two domain swapping strategies can lead to predictable changes in the structure of biosynthesized molecule: AT domain swaps affect the choice of starter or extender unit (malonyl-CoA, methylmalonyl-CoA, or other), whereas reductive loop swaps alter the configuration and oxidative state of the newly added extender unit. For both types of domain swaps, conserved regions were identified that can be used as splice points.^{205,210} (B) “Classical” module boundaries match the boundaries of unimodular proteins and correspond to the functional unit of chain elongation (KS and downstream ACP) and subsequent modification (reductive loop). “Alternative” module boundaries break the functional unit of chain elongation but preserve chain translocation unit (KS and upstream ACP), along with the reductive loop that determines the oxidative state of the translocated substrate.³⁹ Both “classical” and “alternative” module boundaries have been successfully used for module deletion (shown here), as well as module swapping and insertion.^{78,90,206} KS, ketosynthase; AT, acyltransferase; ACP, acyl carrier protein; KR, ketoreductase; DH, dehydratase.

interdomain linker (also called KAL) and post-AT linker, which most likely correspond to natural splice points and can be used for AT domain swapping.^{41,205} These linkers may be responsible for maintaining structural integrity of the module upon recombination, thus enhancing the evolutionary degrees of freedom of assembly-line PKSs.⁴¹ Conserved regions flanking AT domains were used as splice points to emulate natural recombination and exchange between modules: either from the same PKS as in the case of aureothin synthase, or from a homologous PKS as in the case of antimycin and antimycin-like synthases, leading to fully functional chimeric assembly-line PKSs.^{206,207}

The termini of reductive domains and multidomains have also been identified to harbor recombinational hotspots.⁶⁵ Although previous domain swaps at these interfaces did not always result in active PKSs, a few approaches were successful.^{208,209} On the basis of the phylogenetic analysis of these regions of PKS modules, a polylinker approach was developed that allowed testing of various splice sites and reductive domain donors while using presumed regions of natural recombination (Figure 10A, right).²¹⁰ This allows changing of the configuration and the oxidation state of the resulting polyketide.

Another strategy commonly used in PKS engineering involves deleting, inserting, or swapping entire modules,

leading to changes in the polyketide chain length. Under the current evolutionary model that considers gene duplication as a major step leading to the multimodular architectures of *cis*-AT PKSs, it is reasonable to assume that the unit of duplication corresponds to the KS-AT-(KR-DH-ER)-ACP module, which is the functional unit of chain elongation and matches the boundaries of unimodular proteins. This assumption has led to the use of intermodular ACP-KS regions (either linkers or docking domains) in the efforts of module rearrangement in the early 2000s (Figure 10B, left).^{197,198}

However, early on, it became apparent that modules can be deleted, swapped, or inserted at other splice points as well (Figure 10B, right). In 2004, the KS-AT linker region was used to delete two modules of amphotericin synthase, resulting in a functional PKS, producing high yields of the shortened polyene.²¹¹ Later, analysis of several PKS systems suggested that the KS-AT interface was a natural splice site for protein engineering via homologous recombination.^{41,65} Engineering of the aureothin and neo-aureothin synthase showed that splitting modules along the KS-AT interface rather than the “classical” ACP-KS interface was more productive for module deletions and insertions.^{78,90,206} This is particularly interesting in the light of recently proposed “alternative” module boundaries at the KS-AT interface, which are based on close evolutionary relationships between KS domains and the upstream processing domains (discussed in section 2.4.3). Such evolutionarily inspired strategies that alter homologous clusters through domain or module exchanges represent a powerful approach because they often result in chimeric PKSs that produce higher yields of new molecules.

A similar approach of using natural recombination points has been explored for engineering NRPSs. In one study, adenylation domains of hormaomycin synthase were successfully swapped at splicing points that show high sequence similarity.²¹² In another study, a more general strategy was proposed by introducing exchange units with a splice point located between the condensation and adenylation domains, which allows the assembly of chimeric NRPSs producing various new compounds.²¹³

5.2.2. Using Natural Recombination Mechanisms. In approaches where natural recombinational hotspots were used for generating chimeric PKSs, standard laboratory cloning techniques were used to perform the assembly. Another approach inspired by natural PKS evolution uses homologous recombination for the assembly and relies on naturally occurring regions of sequence similarity. The possibility of using homologous recombination was first assessed computationally and suggested numerous regions of sequence similarity that could potentially lead to chimeric assemblies and new molecules.^{214,215} The feasibility of this approach was later shown experimentally in two different studies. First, homologous DNA recombination between DEBS and pikromycin (PKS) clusters was shown to produce numerous functional chimeric assembly lines with splicing points located at various locations within modules, though preferentially in KS and AT domain-encoding regions.²¹⁶ This straightforward and versatile method relies on homologous recombination in yeast and has great potential for generating large libraries of PKS chimeras. A more recent study has demonstrated the possibility of generating chimeras by recombination within a single PKS cluster by harnessing the homologous recombination mechanisms of the host *Streptomyces* strain.²¹⁷ Using this method that accelerates the plausible mechanism of PKS evolution, 17

rapamycin synthase and nine tylactone synthase chimeras were generated, with splicing points mostly located in regions encoding KS, AT, and ACP domains. More strikingly, many of these chimeric PKSs were highly active, with titers comparable to those of the wild-type strain.

The described studies demonstrate the power of evolutionary-inspired engineering for producing active assembly-line PKSs. Homologous recombination-based techniques rely on splicing at arbitrary locations of sequence similarity and generate many chimeras that have to be tested for activity, which requires screening methods. However, the high success rate of producing active PKSs represents a major advance compared to previous engineering approaches and opens many possibilities for future developments.

6. FUTURE DIRECTIONS

6.1. Understanding Assembly-Line PKS Evolution

Even though natural products are widely used in medicine, the role of small molecules in natural environments is poorly understood.²¹⁸ At subinhibitory concentrations, many antibiotics trigger specific responses and may act as signaling molecules.²¹⁹ More broadly, secondary metabolites seem to play various roles in the development of their producing strains and their interactions with the environment.

Regardless of the ecological role that antibiotics play in the natural setting, antibiotic resistance genes have concomitantly coevolved in bacterial populations, forming the resistome.^{220,221} While modern day antibiotic treatment is responsible for the increased degree of resistance gene mobility, the growth of environmental reservoirs of antibiotic resistance and the major healthcare problem that we are currently facing, the antibiotic resistome itself is ancient.^{222,223}

With the increased awareness of the gravity of the antibiotic resistance crisis, the question of its evolution is now being investigated with computational, epidemiological, and molecular tools.²²⁴ The evolution of biosynthetic machines responsible for antibiotic production and diversification is the opposite side of the coin and has undoubtedly played a crucial role in shaping the resistome. However, our understanding of this process is less advanced; even though we have identified the global molecular processes involved (reviewed in section 2), we remain unaware of the molecular mechanisms, the evolutionary paths leading to chemical diversification, and the overall dynamics of these events. The evolution of PKSs is not an exception, and these questions need to be addressed. One of the future challenges for the field is to establish the evolutionary relationships between antibiotic producing and antibiotic resistance systems. Apart from the obvious scientific value of understanding the coevolution of these two systems, it could potentially inform us of interesting avenues for the development of novel antibiotics, such as using nature's toolkit for biosynthetic cluster diversification or exploring the chemical diversity not accessible through natural processes and thus less likely to have corresponding resistance mechanisms evolved and readily available.

6.2. Expanding Access to Natural Diversity

The exponential increase in the number of sequenced assembly-line PKSs and the high percentage of orphan clusters highlight the abundance of polyketide diversity that remains to be explored. Given the effort required to characterize the product of a single PKS, criteria for prioritizing orphan PKSs is essential. Development of computational and wet lab tools for

prioritization is a promising direction for the field. For example, sequence similarity to known clusters²²⁵ or the presence of a potential antibiotic self-resistance gene within the cluster¹⁵⁶ can be promising approaches to select PKSs for deorphanization. A single method is unlikely to solve this problem; instead, multiple approaches tailored to address different challenges must be exploited.

One of the challenges in predicting product structure from the sequence of a PKS is the fact that not all assembly lines follow the colinearity rule: the order in which proteins are encoded can differ from the order in which they operate. Recently, a solution to this problem was proposed. The KS domains of *trans*-AT PKSs typically coevolve with the upstream ACPs and modifying domains, enabling more precise predictions of biosynthesized molecule structures.¹²⁷ In contrast, the coevolution signal between KS domains of *cis*-AT PKSs and the ACP domains of their upstream modules is not strong enough; here, module interactions are more predictable based on docking domain coevolution.²²⁶

Evolutionary insights can also facilitate prioritization of experimental analysis of orphan assembly-line PKSs. A striking example is the discovery of several polyketides that share structural elements with the actin inhibitor misakinolide, most likely due to recombination between their BGCs.²²⁷ If combinatorial exchange of BGC regions is a common strategy for *trans*-AT PKS diversification, this approach could readily lead to the discovery of assembly-line PKSs that produce chimeric molecules.

Finally, it should be recognized that, although a tremendous diversity of assembly-line PKSs has been revealed through DNA sequencing, it is possible that biases in genomic and metagenomic sequencing have created a corresponding bias in our insights into PKS assembly lines. One way to assess the existence of such bias is by targeting underexplored bacterial phyla and environmental niches. For example, marine sponges have been found to harbor a large number of symbiotic bacteria from the genus *Entotheonella*, which produces natural products with a chemical richness that might be comparable to soil actinomycetes.^{97,228} More generally, unculturable bacteria is a promising source of new BGCs, but their DNA can be difficult to retrieve. Several pipelines have been developed that allow culture-independent sequencing, extraction, prioritization, and characterization of DNA encoding PKSs and NRPSs.^{151,229} As they become more versatile, the discovery of novel polyketides from metagenomic libraries may also become more powerful.

6.3. Overcoming Technical Challenges in Deorphanizing PKS Clusters

The techniques described in section 4 enable researchers to rapidly travel from gene to polyketide discovery. However, the path from polyketide discovery to polyketide deorphanization remains slow and painstaking, as there are a multitude of factors that govern a polyketide's production and stability. This challenge is the most significant barrier to quickly uncovering the chemical diversity of orphan polyketides. Without a complete structure, a polyketide, or its novel analogues, cannot be prepared by chemical synthesis, a route that can produce compounds at scales possibly unobtainable with either native or heterologous hosts. Notably, the absence of this information precludes detailed bioactivity experiments such as molecular-level structure–function analysis. To overcome this challenge, researchers must develop tools and strategies to analyze low

abundance products. Arguably, the most notorious example is colibactin. Colibactin is a genotoxic hybrid polyketide–nonribosomal peptide produced in some gut commensal *E. coli* strains and is intriguingly associated with colorectal cancer.¹³⁴ Recently, the Crawford and Herzon groups used an interdisciplinary approach that involved genetics, isotope labeling, tandem mass spectrometry, and chemical synthesis to finally elucidate the full structure of colibactin.²³⁰ For over a decade, no work, despite considerable efforts from several laboratories, had described the identification, isolation, and structural elucidation of the unstable final colibactin. In one herculean effort, 2000 L of culture were required to manufacture just 50 μg of a biosynthetic intermediate of colibactin for structural analysis using 1D and 2D NMR as well as tandem mass spectrometry.²³¹ Intriguing approaches like the one used for colibactin could be combined with yield optimization techniques in native or heterologous hosts to facilitate the deorphanization of polyketides in a timely manner.

6.4. Harnessing the Knowledge of PKS Evolution

Evolutionary analyses of assembly-line PKSs have proven their value not only in advancing the fundamental understanding of these enzymes but also in enabling practical applications. As described in section 5, an understanding of evolutionary processes can also be used for assembly-line PKS engineering. However, many questions about assembly-line PKS evolution remain to be answered. Perhaps foremost on this list is understanding how iterative PKSs gave rise to their assembly line counterparts. Other evolutionary questions are equally relevant. For example, how did *trans*-AT PKSs emerge from their *cis*-AT predecessors through the loss of AT domain? And what can we learn about PKS evolution from their nonuniform distribution among bacteria?

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.chemrev.9b00525>.

Schematic representation of phylogenetic relationships between KS domains from different types of FASs and PKSs; network of 3551 distinct assembly-line PKS clusters, color coded according to the phylum of their host. The catalogue, the dendrogram, and the network file: http://web.stanford.edu/group/orphan_pks/. The code used in this work: https://github.com/aleksnivina/Orphan_PKS_catalog. (PDF)

AUTHOR INFORMATION

Corresponding Author

*Phone: 650-723-3143. E-mail: khosla@stanford.edu.

ORCID

Chaitan Khosla: 0000-0001-6529-495X

Notes

The authors declare no competing financial interest.

Biographies

Aleksandra Nivina is a postdoctoral researcher in Prof. Chaitan Khosla's group at Stanford University. She obtained a M.Sc. in Bioengineering, a PharmD, and a Ph.D. in Interdisciplinary Biology from Université Paris Descartes, where she was part of the Center for

Research and Interdisciplinarity. During her Ph.D., she worked with Prof. Didier Mazel at Institut Pasteur in Paris, France, where she studied the bacterial integron system and its recombination. Her current research aims to understand the mechanisms that drive natural product diversification and their evolutionary role in the arms race between antibiotic production and resistance.

Kai P. Yuet is a postdoctoral researcher in Prof. Chaitan Khosla's group at Stanford University. He obtained a Ph.D. in Chemical Engineering from the California Institute of Technology. During his Ph.D., he worked with Profs. David Tirrell and Paul Sternberg to develop tools for spatiotemporally specific proteomic analysis in *C. elegans*. He earned an S.B. in Chemical Engineering and an S.B. in Biology from the Massachusetts Institute of Technology. He is currently focusing on understanding the biosynthetic repertoire of the NOCAP (NOCardiosis-Associated Polyketide) synthase as well as understanding the synthase's role in the pathogenicity of *Nocardia*.

Jake Hsu is a doctoral student in Prof. Chaitan Khosla's group at Stanford University. He earned a B.Sc. in chemical engineering from the University of Pennsylvania in 2017, and his doctoral work is focused on structural analysis of the 6-deoxyerythronolide B synthase.

Chaitan Khosla is a professor of Chemistry and of Chemical Engineering at Stanford University, where he has taught for nearly three decades. During this period, he has supervised the dissertation research of over 50 doctoral students and the postdoctoral research of more than 50 trainees. He currently also serves as the Baker Family co-Director of Stanford ChEM-H.

ACKNOWLEDGMENTS

We thank Maureen Hillenmeyer for providing the code that was used to generate the catalogue of PKS clusters in 2013¹⁰², as well as Dillon Cogan, Gergana Vandova, and Sherry Li for helpful suggestions. Support from the Stanford Research Computing Center for computational resources is also acknowledged. This research was supported, in part, by NIH R01 GM087934, FRM SPE20170336834 (to A.N.), NIH F32 GM123637 (to K.P.Y.), and NSF DGE 1656518 (to J.H.).

REFERENCES

- (1) Chen, H.; Du, L. Iterative Polyketide Biosynthesis by Modular Polyketide Synthases in Bacteria. *Appl. Microbiol. Biotechnol.* **2016**, *100*, 541–557.
- (2) Khosla, C.; Herschlag, D.; Cane, D. E.; Walsh, C. T. Assembly Line Polyketide Synthases: Mechanistic Insights and Unsolved Problems. *Biochemistry* **2014**, *53*, 2875–2883.
- (3) Dodge, G. J.; Maloney, F. P.; Smith, J. L. Protein-Protein Interactions in “Cis-at” Polyketide Synthases. *Nat. Prod. Rep.* **2018**, *35*, 1082–1096.
- (4) Kosol, S.; Jenner, M.; Lewandowski, J. R.; Challis, G. L. Protein-Protein Interactions in Trans-at Polyketide Synthases. *Nat. Prod. Rep.* **2018**, *35*, 1097–1109.
- (5) Robbins, T.; Liu, Y.-C.; Cane, D. E.; Khosla, C. Structure and Mechanism of Assembly Line Polyketide Synthases. *Curr. Opin. Struct. Biol.* **2016**, *41*, 10–18.
- (6) Walsh, C. T.; Tang, Y. *Natural Product Biosynthesis*; The Royal Society of Chemistry, 2017.
- (7) Keatinge-Clay, A. T. The Uncommon Enzymology of Cis-Acyltransferase Assembly Lines. *Chem. Rev.* **2017**, *117*, 5334–5366.
- (8) Helfrich, E. J. N.; Piel, J. Biosynthesis of Polyketides by Trans-at Polyketide Synthases. *Nat. Prod. Rep.* **2016**, *33*, 231–316.
- (9) Gokhale, R. S.; Tsuji, S. Y.; Cane, D. E.; Khosla, C. Dissecting and Exploiting Intermodular Communication in Polyketide Synthases. *Science* **1999**, *284*, 482–485.
- (10) Broadhurst, R. W.; Nietlispach, D.; Wheatcroft, M. P.; Leadlay, P. F.; Weissman, K. J. The Structure of Docking Domains in Modular Polyketide Synthases. *Chem. Biol.* **2003**, *10*, 723–731.
- (11) Piel, J. Biosynthesis of Polyketides by Trans-at Polyketide Synthases. *Nat. Prod. Rep.* **2010**, *27*, 996–1047.
- (12) Hertweck, C. The Biosynthetic Logic of Polyketide Diversity. *Angew. Chem., Int. Ed.* **2009**, *48*, 4688–4716.
- (13) Fischbach, M. A.; Walsh, C. T. Assembly-Line Enzymology for Polyketide and Nonribosomal Peptide Antibiotics: Logic, Machinery, and Mechanisms. *Chem. Rev.* **2006**, *106*, 3468–3496.
- (14) Sundaram, S.; Hertweck, C. On-Line Enzymatic Tailoring of Polyketides and Peptides in Thio-template Systems. *Curr. Opin. Chem. Biol.* **2016**, *31*, 82–94.
- (15) Fischbach, M. A.; Walsh, C. T.; Clardy, J. The Evolution of Gene Collectives: How Natural Selection Drives Chemical Innovation. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 4601–4608.
- (16) Keatinge-Clay, A. T.; Maltby, D. A.; Medzihradzky, K. F.; Khosla, C.; Stroud, R. M. An Antibiotic Factory Caught in Action. *Nat. Struct. Mol. Biol.* **2004**, *11*, 888–893.
- (17) Lowry, B.; Li, X.; Robbins, T.; Cane, D. E.; Khosla, C. A Turnstile Mechanism for the Controlled Growth of Biosynthetic Intermediates on Assembly Line Polyketide Synthases. *ACS Cent. Sci.* **2016**, *2*, 14–20.
- (18) Robbins, T.; Kapilivsky, J.; Cane, D. E.; Khosla, C. Roles of Conserved Active Site Residues in the Ketosynthase Domain of an Assembly Line Polyketide Synthase. *Biochemistry* **2016**, *55*, 4476–4484.
- (19) Cox, R. J. Polyketides, Proteins and Genes in Fungi: Programmed Nano-Machines Begin to Reveal Their Secrets. *Org. Biomol. Chem.* **2007**, *5*, 2010–2026.
- (20) Jenke-Kodama, H.; Sandmann, A.; Müller, R.; Dittmann, E. Evolutionary Implications of Bacterial Polyketide Synthases. *Mol. Biol. Evol.* **2005**, *22*, 2027–2039.
- (21) Jenke-Kodama, H.; Dittmann, E. Evolution of Metabolic Diversity: Insights From Microbial Polyketide Synthases. *Phytochemistry* **2009**, *70*, 1858–1866.
- (22) Herbst, D. A.; Jakob, R. P.; Zähringer, F.; Maier, T. Mycocerosic Acid Synthase Exemplifies the Architecture of Reducing Polyketide Synthases. *Nature* **2016**, *531*, 533–537.
- (23) Quadri, L. E. N. Biosynthesis of Mycobacterial Lipids by Polyketide Synthases and Beyond. *Crit. Rev. Biochem. Mol. Biol.* **2014**, *49*, 179–211.
- (24) Ahlert, J.; Shepard, E.; Lomovskaya, N.; Zazopoulos, E.; Staffa, A.; Bachmann, B. O.; Huang, K.; Fonstein, L.; Czisny, A.; Whitwam, R. E.; et al. The Calicheamicin Gene Cluster and Its Iterative Type I Eneidyne PKS. *Science* **2002**, *297*, 1173–1176.
- (25) Metz, J. G.; Roessler, P.; Facciotti, D.; Levering, C.; Dittrich, F.; Lassner, M.; Valentine, R.; Lardizabal, K.; Domergue, F.; Yamada, A.; et al. Production of Polyunsaturated Fatty Acids by Polyketide Synthases in Both Prokaryotes and Eukaryotes. *Science* **2001**, *293*, 290–293.
- (26) Campbell, E. L.; Cohen, M. F.; Meeks, J. C. A Polyketide-Synthase-Like Gene Is Involved in the Synthesis of Heterocyst Glycolipids in *Nostoc punctiforme* Strain ATCC 29133. *Arch. Microbiol.* **1997**, *167*, 251–258.
- (27) Herbst, D. A.; Townsend, C. A.; Maier, T. The Architectures of Iterative Type I PKS and FAS. *Nat. Prod. Rep.* **2018**, *35*, 1046–1069.
- (28) Wang, H.; Fewer, D. P.; Holm, L.; Rouhiainen, L.; Sivonen, K. Atlas of Nonribosomal Peptide and Polyketide Biosynthetic Pathways Reveals Common Occurrence of Nonmodular Enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 9259–9264.
- (29) Shen, B.; Chen, M.; Cheng, Y.; Du, L.; Edwards, D. J.; George, N. P.; Huang, Y.; Oh, T.; Sanchez, C.; Tang, G.; et al. Prerequisites for Combinatorial Biosynthesis: Evolution of Hybrid NRPS/PKS Gene Clusters. *Ernst Schering Res. Found. Workshop* **2005**, *51*, 107–126.
- (30) Miyanaga, A.; Kudo, F.; Eguchi, T. Protein-Protein Interactions in Polyketide Synthase-Nonribosomal Peptide Synthetase Hybrid Assembly Lines. *Nat. Prod. Rep.* **2018**, *35*, 1185–1209.

- (31) Piel, J.; Hui, D.; Fusetani, N.; Matsunaga, S. Targeting Modular Polyketide Synthases with Iteratively Acting Acyltransferases From Metagenomes of Uncultured Bacterial Consortia. *Environ. Microbiol.* **2004**, *6*, 921–927.
- (32) Lopez, J. V. Naturally Mosaic Operons for Secondary Metabolite Biosynthesis: Variability and Putative Horizontal Transfer of Discrete Catalytic Domains of the Epothilone Polyketide Synthase Locus. *Mol. Genet. Genomics* **2004**, *270*, 420–431.
- (33) Moffitt, M. C.; Neilan, B. A. Evolutionary Affiliations Within the Superfamily of Ketosynthases Reflect Complex Pathway Associations. *J. Mol. Evol.* **2003**, *56*, 446–457.
- (34) Nguyen, T.; Ishida, K.; Jenke-Kodama, H.; Dittmann, E.; Gurgui, C.; Hochmuth, T.; Taudien, S.; Platzer, M.; Hertweck, C.; Piel, J. Exploiting the Mosaic Structure of Trans-Acyltransferase Polyketide Synthases for Natural Product Discovery and Pathway Dissection. *Nat. Biotechnol.* **2008**, *26*, 225–233.
- (35) Jenner, M.; Frank, S.; Kampa, A.; Kohlhaas, C.; Pöplau, P.; Briggs, G. S.; Piel, J.; Oldham, N. J. Substrate Specificity in Ketosynthase Domains From Trans-at Polyketide Synthases. *Angew. Chem., Int. Ed.* **2013**, *52*, 1143–1147.
- (36) Jenner, M.; Afonso, J. P.; Bailey, H. R.; Frank, S.; Kampa, A.; Piel, J.; Oldham, N. J. Acyl-Chain Elongation Drives Ketosynthase Substrate Selectivity in Trans-Acyltransferase Polyketide Synthases. *Angew. Chem., Int. Ed.* **2015**, *54*, 1817–1821.
- (37) Vander Wood, D. A.; Keatinge-Clay, A. T. The Modules of Trans-Acyltransferase Assembly Lines Redefined with a Central Acyl Carrier Protein. *Proteins: Struct., Funct., Genet.* **2018**, *86*, 664–675.
- (38) Zhang, L.; Hashimoto, T.; Qin, B.; Hashimoto, J.; Kozono, I.; Kawahara, T.; Okada, M.; Awakawa, T.; Ito, T.; Asakawa, Y.; et al. Characterization of Giant Modular PKSs Provides Insight Into Genetic Mechanism for Structural Diversification of Aminopolyol Polyketides. *Angew. Chem., Int. Ed.* **2017**, *56*, 1740–1745.
- (39) Keatinge-Clay, A. T. Polyketide Synthase Modules Redefined. *Angew. Chem., Int. Ed.* **2017**, *56*, 4658–4660.
- (40) Haydock, S. F.; Aparicio, J. F.; Molnár, I.; Schwecke, T.; Khaw, L. E.; König, A.; Marsden, A. F.; Galloway, I. S.; Staunton, J.; Leadlay, P. F. Divergent Sequence Motifs Correlated with the Substrate Specificity of (Methyl)Malonyl-CoA:Acyl Carrier Protein Transacylase Domains in Modular Polyketide Synthases. *FEBS Lett.* **1995**, *374*, 246–248.
- (41) Ridley, C. P.; Lee, H. Y.; Khosla, C. Evolution of Polyketide Synthases in Bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 4595–4600.
- (42) Jensen, K.; Niederkrüger, H.; Zimmermann, K.; Vagstad, A. L.; Moldenhauer, J.; Brendel, N.; Frank, S.; Pöplau, P.; Kohlhaas, C.; Townsend, C. A.; et al. Polyketide Proofreading by an Acyltransferase-Like Enzyme. *Chem. Biol.* **2012**, *19*, 329–339.
- (43) Jenner, M.; Afonso, J. P.; Kohlhaas, C.; Karbaum, P.; Frank, S.; Piel, J.; Oldham, N. J. Acyl Hydrolases From Trans-at Polyketide Synthases Target Acetyl Units on Acyl Carrier Proteins. *Chem. Commun.* **2016**, *52*, S262–S265.
- (44) Caffrey, P. Conserved Amino Acid Residues Correlating with Ketoreductase Stereospecificity in Modular Polyketide Synthases. *ChemBioChem* **2003**, *4*, 654–657.
- (45) Ansari, M. Z.; Sharma, J.; Gokhale, R. S.; Mohanty, D. In Silico Analysis of Methyltransferase Domains Involved in Biosynthesis of Secondary Metabolites. *BMC Bioinf.* **2008**, *9*, 454.
- (46) Hopwood, D. A. Genetic Contributions to Understanding Polyketide Synthases. *Chem. Rev.* **1997**, *97*, 2465–2498.
- (47) Gokhale, R. S.; Khosla, C. Role of Linkers in Communication Between Protein Modules. *Curr. Opin. Chem. Biol.* **2000**, *4*, 22–27.
- (48) Traitcheva, N.; Jenke-Kodama, H.; He, J.; Dittmann, E.; Hertweck, C. Non-Colinear Polyketide Biosynthesis in the Aureothin and Neoareothin Pathways: an Evolutionary Perspective. *Chem-BioChem* **2007**, *8*, 1841–1849.
- (49) Jenke-Kodama, H.; Börner, T.; Dittmann, E. Natural Biocombinatorics in the Polyketide Synthase Genes of the Actinobacterium *Streptomyces Avermitilis*. *PLoS Comput. Biol.* **2006**, *2*, No. e132.
- (50) Ochman, H.; Lawrence, J. G.; Groisman, E. A. Lateral Gene Transfer and the Nature of Bacterial Innovation. *Nature* **2000**, *405*, 299–304.
- (51) Wiener, P.; Egan, S.; Wellington, E. M. H. Evidence for Transfer of Antibiotic-Resistance Genes in Soil Populations of *Streptomyces*. *Mol. Ecol.* **1998**, *7*, 1205–1216.
- (52) Egan, S.; Wiener, P.; Kallifidas, D.; Wellington, E. M. H. Phylogeny of *Streptomyces* Species and Evidence for Horizontal Transfer of Entire and Partial Antibiotic Gene Clusters. *Antonie van Leeuwenhoek* **2001**, *79*, 127–133.
- (53) Flórez, L. V.; Scherlach, K.; Miller, I. J.; Rodrigues, A.; Kwan, J. C.; Hertweck, C.; Kaltenpoth, M. An Antifungal Polyketide Associated with Horizontally Acquired Genes Supports Symbiont-Mediated Defense in *Lagria villosa* Beetles. *Nat. Commun.* **2018**, *9*, 2478.
- (54) Hochmuth, T.; Piel, J. Polyketide Synthases of Bacterial Symbionts in Sponges-Evolution-Based Applications in Natural Products Research. *Phytochemistry* **2009**, *70*, 1841–1849.
- (55) Lawrence, D. P.; Kroken, S.; Pryor, B. M.; Arnold, A. E. Interkingdom Gene Transfer of a Hybrid NPS/PKS From Bacteria to Filamentous Ascomycota. *PLoS One* **2011**, *6*, No. e28231.
- (56) Stonesifer, J.; Matsushima, P.; Baltz, R. H. High Frequency Conjugal Transfer of Tylosin Genes and Amplifiable DNA in *Streptomyces Fradiae*. *Mol. Gen. Genet.* **1986**, *202*, 348–355.
- (57) Stinear, T. P.; Mve-Obiang, A.; Small, P. L. C.; Frigui, W.; Pryor, M. J.; Brosch, R.; Jenkin, G. A.; Johnson, P. D. R.; Davies, J. K.; Lee, R. E.; et al. Giant Plasmid-Encoded Polyketide Synthases Produce the Macrolide Toxin of *Mycobacterium Ulcerans*. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 1345–1349.
- (58) Suwa, M.; Sugino, H.; Sasaoka, A.; Mori, E.; Fujii, S.; Shinkawa, H.; Nimi, O.; Kinashi, H. Identification of Two Polyketide Synthase Gene Clusters on the Linear Plasmid pSLA2-L in *Streptomyces Rochei*. *Gene* **2000**, *246*, 123–131.
- (59) Szczepanowski, R.; Braun, S.; Riedel, V.; Schneider, S.; Krahn, I.; Pühler, A.; Schlüter, A. The 120 592 Bp IncF Plasmid pRSB107 Isolated From a Sewage-Treatment Plant Encodes Nine Different Antibiotic-Resistance Determinants, Two Iron-Acquisition Systems and Other Putative Virulence-Associated Functions. *Microbiology* **2005**, *151*, 1095–1111.
- (60) Carniel, E. The *Yersinia* High-Pathogenicity Island: an Iron-Uptake Island. *Microbes Infect.* **2001**, *3*, 561–569.
- (61) Ginolhac, A.; Jarrin, C.; Robe, P.; Perrière, G.; Vogel, T. M.; Simonet, P.; Nalin, R. Type I Polyketide Synthases May Have Evolved Through Horizontal Gene Transfer. *J. Mol. Evol.* **2005**, *60*, 716–725.
- (62) Chen, J.-M.; Cooper, D. N.; Chuzhanova, N.; Férec, C.; Patrinos, G. P. Gene Conversion: Mechanisms, Evolution and Human Disease. *Nat. Rev. Genet.* **2007**, *8*, 762–775.
- (63) Santoyo, G.; Romero, D. Gene Conversion and Concerted Evolution in Bacterial Genomes. *FEMS Microbiol. Rev.* **2005**, *29*, 169–183.
- (64) Pidot, S. J.; Hong, H.; Seemann, T.; Porter, J. L.; Yip, M. J.; Men, A.; Johnson, M.; Wilson, P.; Davies, J. K.; Leadlay, P. F.; et al. Deciphering the Genetic Basis for Polyketide Variation Among *Mycobacteria* Producing Mycolactones. *BMC Genomics* **2008**, *9*, 462.
- (65) Zucko, J.; Long, P. F.; Hranueli, D.; Cullum, J. Horizontal Gene Transfer and Gene Conversion Drive Evolution of Modular Polyketide Synthases. *J. Ind. Microbiol. Biotechnol.* **2012**, *39*, 1541–1547.
- (66) Thattai, M.; Burak, Y.; Shraiman, B. I. The Origins of Specificity in Polyketide Synthase Protein Interactions. *PLoS Comput. Biol.* **2007**, *3*, e186.
- (67) Cheng, K.; Rong, X.; Huang, Y. Widespread Interspecies Homologous Recombination Reveals Reticulate Evolution Within the Genus *Streptomyces*. *Mol. Phylogenet. Evol.* **2016**, *102*, 246–254.
- (68) Doroghazi, J. R.; Buckley, D. H. Widespread Homologous Recombination Within and Between *Streptomyces* Species. *ISME J.* **2010**, *4*, 1136–1143.
- (69) Gay, D. C.; Gay, G.; Axelrod, A. J.; Jenner, M.; Kohlhaas, C.; Kampa, A.; Oldham, N. J.; Piel, J.; Keatinge-Clay, A. T. A Close Look

at a Ketosynthase From a Trans-Acyltransferase Modular Polyketide Synthase. *Structure* **2014**, *22*, 444–451.

(70) Kapur, S.; Chen, A. Y.; Cane, D. E.; Khosla, C. Molecular Recognition Between Ketosynthase and Acyl Carrier Protein Domains of the 6-Deoxyerythronolide B Synthase. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 22066–22071.

(71) Gay, D. C.; Wagner, D. T.; Meinke, J. L.; Zogzas, C. E.; Gay, G. R.; Keatinge-Clay, A. T. The LINKS Motif Zippers Trans-Acyltransferase Polyketide Synthase Assembly Lines Into a Biosynthetic Megacomplex. *J. Struct. Biol.* **2016**, *193*, 196–205.

(72) Lohman, J. R.; Ma, M.; Osipiuk, J.; Nocek, B.; Kim, Y.; Chang, C.; Cuff, M.; Mack, J.; Bigelow, L.; Li, H.; et al. Structural and Evolutionary Relationships of “at-Less” Type I Polyketide Synthase Ketosynthases. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 12693–12698.

(73) Moss, S. J.; Martin, C. J.; Wilkinson, B. Loss of Co-Linearity by Modular Polyketide Synthases: a Mechanism for the Evolution of Chemical Diversity. *Nat. Prod. Rep.* **2004**, *21*, 575–593.

(74) Gaitatzis, N.; Silakowski, B.; Kunze, B.; Nordsiek, G.; Blöcker, H.; Höfle, G.; Müller, R. The Biosynthesis of the Aromatic Myxobacterial Electron Transport Inhibitor Stigmatellin Is Directed by a Novel Type of Modular Polyketide Synthase. *J. Biol. Chem.* **2002**, *277*, 13082–13090.

(75) Olano, C.; Wilkinson, B.; Sanchez, C.; Moss, S. J.; Sheridan, R.; Math, V.; Weston, A. J.; Brana, A. F.; Martin, C. J.; Oliynyk, M.; Mendez, C.; Leadlay, P. F.; Salas, J. A Biosynthesis of the Angiogenesis Inhibitor Borrelidin by *Streptomyces Parvulus* Tü4055: Cluster Analysis and Assignment of Functions. *Chem. Biol.* **2004**, *11*, 87–97.

(76) He, J.; Hertweck, C. Functional Analysis of the Aureothin Iterative Type I Polyketide Synthase. *ChemBioChem* **2005**, *6*, 908–912.

(77) Busch, B.; Ueberschaar, N.; Behnken, S.; Sugimoto, Y.; Werneburg, M.; Traitcheva, N.; He, J.; Hertweck, C. Multifactorial Control of Iteration Events in a Modular Polyketide Assembly Line. *Angew. Chem., Int. Ed.* **2013**, *52*, 5285–5289.

(78) Sugimoto, Y.; Ishida, K.; Traitcheva, N.; Busch, B.; Dahse, H.-M.; Hertweck, C. Freedom and Constraint in Engineered Non-linear Polyketide Assembly Lines. *Chem. Biol.* **2015**, *22*, 229–240.

(79) Wilkinson, B.; Foster, G.; Rudd, B. A.; Taylor, N. L.; Blackaby, A. P.; Sidebottom, P. J.; Cooper, D. J.; Dawson, M. J.; Buss, A. D.; Gaisser, S.; et al. Novel Octaketide Macrolides Related to 6-Deoxyerythronolide B Provide Evidence for Iterative Operation of the Erythromycin Polyketide Synthase. *Chem. Biol.* **2000**, *7*, 111–117.

(80) Hardt, I. H.; Steinmetz, H.; Gerth, K.; Sasse, F.; Reichenbach, H.; Höfle, G. New Natural Epothilones From *Sorangium Cellulosum*, Strains So Ce90/B2 and So Ce90/D13: Isolation, Structure Elucidation, and SAR Studies. *J. Nat. Prod.* **2001**, *64*, 847–856.

(81) Kapur, S.; Lowry, B.; Yuzawa, S.; Kenthirapalan, S.; Chen, A. Y.; Cane, D. E.; Khosla, C. Reprogramming a Module of the 6-Deoxyerythronolide B Synthase for Iterative Chain Elongation. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 4110–4115.

(82) Medema, M. H.; Cimermancic, P.; Sali, A.; Takano, E.; Fischbach, M. A. A Systematic Computational Analysis of Biosynthetic Gene Cluster Evolution: Lessons for Engineering Biosynthesis. *PLoS Comput. Biol.* **2014**, *10*, No. e1004016.

(83) Shelest, E.; Heimerl, N.; Fichtner, M.; Sasso, S. Multimodular Type I Polyketide Synthases in Algae Evolve by Module Duplications and Displacement of at Domains in Trans. *BMC Genomics* **2015**, *16*, 1015.

(84) Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; et al. Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 539–539.

(85) Cortés, J.; Haydock, S. F.; Roberts, G. A.; Bevitt, D. J.; Leadlay, P. F. An Unusually Large Multifunctional Polypeptide in the Erythromycin-Producing Polyketide Synthase of *Saccharopolyspora Erythraea*. *Nature* **1990**, *348*, 176–178.

(86) Donadio, S.; Staver, M. J.; McAlpine, J. B.; Swanson, S. J.; Katz, L. Modular Organization of Genes Required for Complex Polyketide Biosynthesis. *Science* **1991**, *252*, 675–679.

(87) Klaus, M.; Ostrowski, M. P.; Austerjost, J.; Robbins, T.; Lowry, B.; Cane, D. E.; Khosla, C. Protein-Protein Interactions, Not Substrate Recognition, Dominate the Turnover of Chimeric Assembly Line Polyketide Synthases. *J. Biol. Chem.* **2016**, *291*, 16404–16415.

(88) Watanabe, K.; Wang, C. C. C.; Boddy, C. N.; Cane, D. E.; Khosla, C. Understanding Substrate Specificity of Polyketide Synthase Modules by Generating Hybrid Multimodular Synthases. *J. Biol. Chem.* **2003**, *278*, 42020–42026.

(89) Klaus, M.; D’Souza, A. D.; Nivina, A.; Khosla, C.; Gringer, M. Engineering of Chimeric Polyketide Synthases Using SYNZIP Docking Domains. *ACS Chem. Biol.* **2019**, *14*, 426–433.

(90) Peng, H.; Ishida, K.; Sugimoto, Y.; Jenke-Kodama, H.; Hertweck, C. Emulating Evolutionary Processes to Morph Aureothin-Type Modular Polyketide Synthases and Associated Oxygenases. *Nat. Commun.* **2019**, *10*, 3918.

(91) Walsh, C. *Antibiotics*; ASM Press: Washington, DC, 2003.

(92) Romero, D.; Traxler, M. F.; López, D.; Kolter, R. Antibiotics as Signal Molecules. *Chem. Rev.* **2011**, *111*, 5492–5505.

(93) Adusumilli, S.; Mve-Obiang, A.; Sparer, T.; Meyers, W.; Hayman, J.; Small, P. L. C. Mycobacterium Ulcerans Toxic Macrolide, Mycolactone Modulates the Host Immune Response and Cellular Location of M. Ulcerans in Vitro and in Vivo. *Cell. Microbiol.* **2005**, *7*, 1295–1304.

(94) Trindade-Silva, A. E.; Lim-Fong, G. E.; Sharp, K. H.; Haygood, M. G. Bryostatins: Biological Context and Biotechnological Prospects. *Curr. Opin. Biotechnol.* **2010**, *21*, 834–842.

(95) Perry, R. D.; Balbo, P. B.; Jones, H. A.; Fetherston, J. D.; DeMoll, E. Yersiniabactin From *Yersinia Pestis*: Biochemical Characterization of the Siderophore and Its Role in Iron Transport and Regulation. *Microbiology* **1999**, *145*, 1181–1190.

(96) Haeder, S.; Wirth, R.; Herz, H.; Spitter, D. Candididin-Producing *Streptomyces* Support Leaf-Cutting Ants to Protect Their Fungus Garden Against the Pathogenic Fungus *Escovopsis*. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 4742–4746.

(97) Wilson, M. C.; Mori, T.; Rückert, C.; Uria, A. R.; Helf, M. J.; Takada, K.; Gernert, C.; Steffens, U. A. E.; Heycke, N.; Schmitt, S.; et al. An Environmental Bacterial Taxon with a Large and Distinct Metabolic Repertoire. *Nature* **2014**, *506*, 58–62.

(98) Blunt, J. W.; Copp, B. R.; Munro, M. H. G.; Northcote, P. T.; Prinsep, M. R. Marine Natural Products. *Nat. Prod. Rep.* **2011**, *28*, 196–268.

(99) Traxler, M. F.; Kolter, R. Natural Products in Soil Microbe Interactions and Evolution. *Nat. Prod. Rep.* **2015**, *32*, 956–970.

(100) Ehrenreich, I. M.; Waterbury, J. B.; Webb, E. A. Distribution and Diversity of Natural Product Genes in Marine and Freshwater Cyanobacterial Cultures and Genomes. *Appl. Environ. Microbiol.* **2005**, *71*, 7401–7413.

(101) Jenke-Kodama, H.; Müller, R.; Dittmann, E. Evolutionary Mechanisms Underlying Secondary Metabolite Diversity. *Prog. Drug Res.* **2008**, *65*, 119–140.

(102) O’Brien, R. V.; Davis, R. W.; Khosla, C.; Hillenmeyer, M. E. Computational Identification and Analysis of Orphan Assembly-Line Polyketide Synthases. *J. Antibiot.* **2014**, *67*, 89–97.

(103) Kinashi, H. Giant Linear Plasmids in *Streptomyces*: a Treasure Trove of Antibiotic Biosynthetic Clusters. *J. Antibiot.* **2011**, *64*, 19–25.

(104) Bordeleau, E.; Ghinet, M. G.; Burrus, V. Diversity of Integrating Conjugative Elements in Actinobacteria: Coexistence of Two Mechanistically Different DNA-Translocation Systems. *Mob. Genet. Elements* **2012**, *2*, 119–124.

(105) Penn, K.; Jenkins, C.; Nett, M.; Udway, D. W.; Gontang, E. A.; McGlinchey, R. P.; Foster, B.; Lapidus, A.; Podell, S.; Allen, E. E.; et al. Genomic Islands Link Secondary Metabolism to Functional Adaptation in Marine Actinobacteria. *ISME J.* **2009**, *3*, 1193–1203.

(106) Sit, C. S.; Ruzzini, A. C.; Van Arnam, E. B.; Ramadhar, T. R.; Currie, C. R.; Clardy, J. Variable Genetic Architectures Produce

Virtually Identical Molecules in Bacterial Symbionts of Fungus-Growing Ants. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 13150–13154.

(107) Donadio, S.; Monciardini, P.; Sosio, M. Polyketide Synthases and Nonribosomal Peptide Synthetases: the Emerging View From Bacterial Genomics. *Nat. Prod. Rep.* **2007**, *24*, 1073.

(108) Zhou, Z.; Gu, J.; Li, Y.-Q.; Wang, Y. Genome Plasticity and Systems Evolution in *Streptomyces*. *BMC Bioinf.* **2012**, *13*, S8.

(109) Hoff, G.; Bertrand, C.; Piotrowski, E.; Thibessard, A.; Leblond, P. Genome Plasticity Is Governed by Double Strand Break DNA Repair in *Streptomyces*. *Sci. Rep.* **2018**, *8*, 5272.

(110) Marsolier-Kergoat, M.-C.; Yeramian, E. GC Content and Recombination: Reassessing the Causal Effects for the *Saccharomyces Cerevisiae* Genome. *Genetics* **2009**, *183*, 31–38.

(111) Dittmann, E.; Gugger, M.; Sivonen, K.; Fewer, D. P. Natural Product Biosynthetic Diversity and Comparative Genomics of the Cyanobacteria. *Trends Microbiol.* **2015**, *23*, 642–652.

(112) Ishida, K.; Welker, M.; Christiansen, G.; Cadel-Six, S.; Bouchier, C.; Dittmann, E.; Hertweck, C.; Tandeau de Marsac, N. Plasticity and Evolution of Aeruginosin Biosynthesis in Cyanobacteria. *Appl. Environ. Microbiol.* **2009**, *75*, 2017–2026.

(113) Williams, D. H.; Stone, M. J.; Hauck, P. R.; Rahman, S. K. Why Are Secondary Metabolites (Natural Products) Biosynthesized? *J. Nat. Prod.* **1989**, *52*, 1189–1208.

(114) Firm, R. D.; Jones, C. G. The Evolution of Secondary Metabolism - a Unifying Model. *Mol. Microbiol.* **2000**, *37*, 989–994.

(115) Firm, R. D.; Jones, C. G. Natural Products – a Simple Model to Explain Chemical Diversity. *Nat. Prod. Rep.* **2003**, *20*, 382–391.

(116) Starcevic, A.; Zucko, J.; Simunkovic, J.; Long, P. F.; Cullum, J.; Hranueli, D. ClustScan: an Integrated Program Package for the Semi-Automatic Annotation of Modular Biosynthetic Gene Clusters and in Silico Prediction of Novel Chemical Structures. *Nucleic Acids Res.* **2008**, *36*, 6882–6892.

(117) Conway, K. R.; Boddy, C. N. ClusterMine360: a Database of Microbial PKS/NRPS Biosynthesis. *Nucleic Acids Res.* **2012**, *41*, D402–D407.

(118) Khater, S.; Gupta, M.; Agrawal, P.; Sain, N.; Prava, J.; Gupta, P.; Grover, M.; Kumar, N.; Mohanty, D. SBSPKSv2: Structure-Based Sequence Analysis of Polyketide Synthases and Non-Ribosomal Peptide Synthetases. *Nucleic Acids Res.* **2017**, *45*, W72–W79.

(119) Ichikawa, N.; Sasagawa, M.; Yamamoto, M.; Komaki, H.; Yoshida, Y.; Yamazaki, S.; Fujita, N. DoBISCUIT: a Database of Secondary Metabolite Biosynthetic Gene Clusters. *Nucleic Acids Res.* **2012**, *41*, D408–D414.

(120) Medema, M. H.; Kottmann, R.; Yilmaz, P.; Cummings, M.; Biggins, J. B.; Blin, K.; de Bruijn, I.; Chooi, Y. H.; Claesen, J.; Coates, R. C.; et al. Minimum Information About a Biosynthetic Gene Cluster. *Nat. Chem. Biol.* **2015**, *11*, 625–631.

(121) Medema, M. H.; Blin, K.; Cimermancic, P.; de Jager, V.; Zakrzewski, P.; Fischbach, M. A.; Weber, T.; Takano, E.; Breitling, R. antiSMASH: Rapid Identification, Annotation and Analysis of Secondary Metabolite Biosynthesis Gene Clusters in Bacterial and Fungal Genome Sequences. *Nucleic Acids Res.* **2011**, *39*, W339–W346.

(122) Cimermancic, P.; Medema, M. H.; Claesen, J.; Kurita, K.; Wieland Brown, L. C.; Mavrommatis, K.; Pati, A.; Godfrey, P. A.; Koehrsen, M.; Clardy, J.; et al. Insights Into Secondary Metabolism From a Global Analysis of Prokaryotic Biosynthetic Gene Clusters. *Cell* **2014**, *158*, 412–421.

(123) Skinnider, M. A.; Dejong, C. A.; Rees, P. N.; Johnston, C. W.; Li, H.; Webster, A. L. H.; Wyatt, M. A.; Magarvey, N. A. Genomes to Natural Products Prediction Informatics for Secondary Metabolomes (PRISM). *Nucleic Acids Res.* **2015**, *43*, 9645–9662.

(124) Adamek, M.; Spohn, M.; Stegmann, E.; Ziemert, N. Mining Bacterial Genomes for Secondary Metabolite Gene Clusters. *Methods Mol. Biol.* **2017**, *1520*, 23–47.

(125) Tran, P. N.; Yen, M.-R.; Chiang, C.-Y.; Lin, H.-C.; Chen, P.-Y. Detecting and Prioritizing Biosynthetic Gene Clusters for Bioactive Compounds in Bacteria and Fungi. *Appl. Microbiol. Biotechnol.* **2019**, *103*, 3277–3287.

(126) Ziemert, N.; Podell, S.; Penn, K.; Badger, J. H.; Allen, E.; Jensen, P. R. The Natural Product Domain Seeker NaPDoS: a Phylogeny Based Bioinformatic Tool to Classify Secondary Metabolite Gene Diversity. *PLoS One* **2012**, *7*, No. e34064.

(127) Helfrich, E. J. N.; Ueoka, R.; Dolev, A.; Rust, M.; Meoded, R. A.; Bhushan, A.; Califano, G.; Costa, R.; Gugger, M.; Steinbeck, C.; et al. Automated Structure Prediction of Trans-Acyltransferase Polyketide Synthase Products. *Nat. Chem. Biol.* **2019**, *15*, 813–821.

(128) Blin, K.; Shaw, S.; Steinke, K.; Villebro, R.; Ziemert, N.; Lee, S. Y.; Medema, M. H.; Weber, T. antiSMASH 5.0: Updates to the Secondary Metabolite Genome Mining Pipeline. *Nucleic Acids Res.* **2019**, *47*, W81–W87.

(129) Hadjithomas, M.; Chen, I.-M. A.; Chu, K.; Ratner, A.; Palaniappan, K.; Szeto, E.; Huang, J.; Reddy, T. B. K.; Cimermancic, P.; Fischbach, M. A.; Ivanova, N. N.; Markowitz, V. M.; Kyrpidis, N. C.; Pati, A.; et al. IMG-ABC: a Knowledge Base to Fuel Discovery of Biosynthetic Gene Clusters and Novel Secondary Metabolites. *mBio* **2015**, *6*, No. 00932.

(130) Blin, K.; Pascal Andreu, V.; de Los Santos, E. L. C.; Del Carratore, F.; Lee, S. Y.; Medema, M. H.; Weber, T. The antiSMASH Database Version 2: a Comprehensive Resource on Secondary Metabolite Biosynthetic Gene Clusters. *Nucleic Acids Res.* **2019**, *47*, D625–D630.

(131) Blin, K.; Wolf, T.; Chevrette, M. G.; Lu, X.; Schwalen, C. J.; Kautsar, S. A.; Suarez Duran, H. G.; de Los Santos, E. L. C.; Kim, H. U.; Nave, M.; et al. antiSMASH 4.0-Improvements in Chemistry Prediction and Gene Cluster Boundary Identification. *Nucleic Acids Res.* **2017**, *45*, W36–W41.

(132) Atkinson, H. J.; Morris, J. H.; Ferrin, T. E.; Babbitt, P. C. Using Sequence Similarity Networks for Visualization of Relationships Across Diverse Protein Superfamilies. *PLoS One* **2009**, *4*, No. e4345.

(133) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: a Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13*, 2498–2504.

(134) Nougayrède, J.-P.; Homburg, S.; Taieb, F.; Boury, M.; Brzuszkiewicz, E.; Gottschalk, G.; Buchrieser, C.; Hacker, J.; Dobrindt, U.; Oswald, E. *Escherichia Coli* Induces DNA Double-Strand Breaks in Eukaryotic Cells. *Science* **2006**, *313*, 848–851.

(135) Partida-Martinez, L. P.; Hertweck, C. A Gene Cluster Encoding Rhizoxin Biosynthesis in “*Burkholderia Rhizoxina*,” the Bacterial Endosymbiont of the Fungus *Rhizopus Microsporus*. *ChemBioChem* **2007**, *8*, 41–45.

(136) Zhang, F.; He, H.-Y.; Tang, M.-C.; Tang, Y.-M.; Zhou, Q.; Tang, G.-L. Cloning and Elucidation of the FR901464 Gene Cluster Revealing a Complex Acyltransferase-Less Polyketide Synthase Using Glycerate as Starter Units. *J. Am. Chem. Soc.* **2011**, *133*, 2452–2462.

(137) Shou, Q.; Feng, L.; Long, Y.; Han, J.; Nunnery, J. K.; Powell, D. H.; Butcher, R. A. A Hybrid Polyketide-Nonribosomal Peptide in Nematodes That Promotes Larval Survival. *Nat. Chem. Biol.* **2016**, *12*, 770–772.

(138) Eichinger, L.; Pachebat, J. A.; Glöckner, G.; Rajandream, M.-A.; Sugrang, R.; Berriman, M.; Song, J.; Olsen, R.; Szafranski, K.; Xu, Q.; et al. The Genome of the Social Amoeba *Dictyostelium Discoideum*. *Nature* **2005**, *435*, 43–57.

(139) Austin, M. B.; Saito, T.; Bowman, M. E.; Haydock, S.; Kato, A.; Moore, B. S.; Kay, R. R.; Noel, J. P. Biosynthesis of *Dictyostelium Discoideum* Differentiation-Inducing Factor by a Hybrid Type I Fatty Acid-Type III Polyketide Synthase. *Nat. Chem. Biol.* **2006**, *2*, 494–502.

(140) Ghosh, R.; Chhabra, A.; Phatale, P. A.; Samrat, S. K.; Sharma, J.; Gosain, A.; Mohanty, D.; Saran, S.; Gokhale, R. S. Dissecting the Functional Role of Polyketide Synthases in *Dictyostelium Discoideum*: Biosynthesis of the Differentiation Regulating Factor 4-Methyl-5-Pentylbenzene-1,3-Diol. *J. Biol. Chem.* **2008**, *283*, 11348–11354.

(141) Narita, T. B.; Chen, Z.-H.; Schaap, P.; Saito, T. The Hybrid Type Polyketide Synthase SteelyA Is Required for cAMP Signalling in Early *Dictyostelium* Development. *PLoS One* **2014**, *9*, No. e106634.

- (142) Zhu, G.; Marchewka, M. J.; Woods, K. M.; Upton, S. J.; Keithly, J. S. Molecular Analysis of a Type I Fatty Acid Synthase in *Cryptosporidium Parvum*. *Mol. Biochem. Parasitol.* **2000**, *105*, 253–260.
- (143) Xu, P.; Widmer, G.; Wang, Y.; Ozaki, L. S.; Alves, J. M.; Serrano, M. G.; Puiu, D.; Manque, P.; Akiyoshi, D.; Mackey, A. J.; et al. The Genome of *Cryptosporidium Hominis*. *Nature* **2004**, *431*, 1107–1112.
- (144) Bushkin, G. G.; Motari, E.; Carpentieri, A.; Dubey, J. P.; Costello, C. E.; Robbins, P. W.; Samuelson, J. Evidence for a Structural Role for Acid-Fast Lipids in Oocyst Walls of *Cryptosporidium*, *Toxoplasma*, and *Eimeria*. *mBio* **2013**, *4*, 00387.
- (145) Zhu, G.; LaGier, M. J.; Stejskal, F.; Millership, J. J.; Cai, X.; Keithly, J. S. *Cryptosporidium Parvum*: the First Protist Known to Encode a Putative Polyketide Synthase. *Gene* **2002**, *298*, 79–89.
- (146) Monroe, E. A.; Van Dolah, F. M. The Toxic Dinoflagellate *Karenia Brevis* Encodes Novel Type I-Like Polyketide Synthases Containing Discrete Catalytic Domains. *Protist* **2008**, *159*, 471–482.
- (147) Beedessee, G.; Hisata, K.; Roy, M. C.; Satoh, N.; Shoguchi, E. Multifunctional Polyketide Synthase Genes Identified by Genomic Survey of the Symbiotic Dinoflagellate, *Symbiodinium Minutum*. *BMC Genomics* **2015**, *16*, 941.
- (148) Kohli, G. S.; John, U.; Figueroa, R. I.; Rhodes, L. L.; Harwood, D. T.; Groth, M.; Bolch, C. J. S.; Murray, S. A. Polyketide Synthesis Genes Associated with Toxin Production in Two Species of *Gambierdiscus* (Dinophyceae). *BMC Genomics* **2015**, *16*, 410.
- (149) Loureiro, C.; Medema, M. H.; van der Oost, J.; Sipkema, D. Exploration and Exploitation of the Environment for Novel Specialized Metabolites. *Curr. Opin. Biotechnol.* **2018**, *50*, 206–213.
- (150) Harvey, A. L.; Edrada-Ebel, R.; Quinn, R. J. The Re-Emergence of Natural Products for Drug Discovery in the Genomics Era. *Nat. Rev. Drug Discovery* **2015**, *14*, 111–129.
- (151) Katz, M.; Hover, B. M.; Brady, S. F. Culture-Independent Discovery of Natural Products From Soil Metagenomes. *J. Ind. Microbiol. Biotechnol.* **2016**, *43*, 129–141.
- (152) Dejong, C. A.; Chen, G. M.; Li, H.; Johnston, C. W.; Edwards, M. R.; Rees, P. N.; Skinnider, M. A.; Webster, A. L. H.; Magarvey, N. A. Polyketide and Nonribosomal Peptide Retro-Biosynthesis and Global Gene Cluster Matching. *Nat. Chem. Biol.* **2016**, *12*, 1007–1014.
- (153) Cruz-Morales, P.; Kopp, J. F.; Martínez-Guerrero, C.; Yáñez-Guerra, L. A.; Selem-Mojica, N.; Ramos-Aboites, H.; Feldmann, J.; Barona-Gómez, F. Phylogenomic Analysis of Natural Products Biosynthetic Gene Clusters Allows Discovery of Arseno-Organic Metabolites in Model Streptomycetes. *Genome Biol. Evol.* **2016**, *8*, 1906–1916.
- (154) Selem-Mojica, N.; Aguilar, C.; Gutiérrez-García, K.; Martínez-Guerrero, C. E.; Barona-Gómez, F. EvoMining Reveals the Origin and Fate of Natural Product Biosynthetic Enzymes. *Microb. Genom.* **2019**, *2019*, 260.
- (155) Navarro-Munoz, J. C.; Selem-Mojica, N.; Mullowney, M. W.; Kautsar, S. A.; Tryon, J. H.; Parkinson, E. I.; De Los Santos, E. L. C.; Yeong, M.; Cruz-Morales, P.; Abubucker, S.; Roeters, A.; Lokhorst, W.; Fernandez-Guerra, A.; Cappellini, L. T. D.; Goering, A. W.; Thomson, R. J.; Metcalf, W. W.; Kelleher, N. L.; Barona-Gomez, F.; Medema, M. H. A Computational Framework for Systematic Exploration of Biosynthetic Diversity From Large-Scale Genomic Data. *Nat. Chem. Biol.* **2019**. DOI: 10.1038/s41589-019-0400-9
- (156) Alanjary, M.; Kronmiller, B.; Adamek, M.; Blin, K.; Weber, T.; Huson, D.; Philmus, B.; Ziemert, N. The Antibiotic Resistant Target Seeker (ARTS), an Exploration Engine for Antibiotic Cluster Prioritization and Novel Drug Target Discovery. *Nucleic Acids Res.* **2017**, *45*, W42–W48.
- (157) Kuo, J.; Lynch, S. R.; Liu, C. W.; Xiao, X.; Khosla, C. Partial in Vitro Reconstitution of an Orphan Polyketide Synthase Associated with Clinical Cases of Nocardiosis. *ACS Chem. Biol.* **2016**, *11*, 2636–2641.
- (158) Murphy, K. C. Use of Bacteriophage Lambda Recombination Functions to Promote Gene Replacement in *Escherichia Coli*. *J. Bacteriol.* **1998**, *180*, 2063–2071.
- (159) Zhang, Y.; Buchholz, F.; Muylers, J. P.; Stewart, A. F. A New Logic for DNA Engineering Using Recombination in *Escherichia Coli*. *Nat. Genet.* **1998**, *20*, 123–128.
- (160) Fu, J.; Bian, X.; Hu, S.; Wang, H.; Huang, F.; Seibert, P. M.; Plaza, A.; Xia, L.; Muller, R.; Stewart, A. F.; Zhang, Y. Full-Length RecE Enhances Linear-Linear Homologous Recombination and Facilitates Direct Cloning for Bioprospecting. *Nat. Biotechnol.* **2012**, *30*, 440–446.
- (161) Tu, Q.; Herrmann, J.; Hu, S.; Raju, R.; Bian, X.; Zhang, Y.; Müller, R. Genetic Engineering and Heterologous Expression of the Disorazol Biosynthetic Gene Cluster via Red/ET Recombineering. *Sci. Rep.* **2016**, *6*, 21066.
- (162) Yin, J.; Hoffmann, M.; Bian, X.; Tu, Q.; Yan, F.; Xia, L.; Ding, X.; Stewart, A. F.; Müller, R.; Fu, J.; Zhang, Y. Direct Cloning and Heterologous Expression of the Salinomycin Biosynthetic Gene Cluster From *Streptomyces Albus* DSM41398 in *Streptomyces Coelicolor* A3(2). *Sci. Rep.* **2015**, *5*, 15081.
- (163) Tang, Y.; Frewert, S.; Harmrolfs, K.; Herrmann, J.; Karmann, L.; Kazmaier, U.; Xia, L.; Zhang, Y.; Müller, R. Heterologous Expression of an Orphan NRPS Gene Cluster From *Paenibacillus Larvae* in *Escherichia Coli* Revealed Production of Sevadacin. *J. Biotechnol.* **2015**, *194*, 112–114.
- (164) Larionov, V.; Kouprina, N.; Solomon, G.; Barrett, J. C.; Resnick, M. A. Direct Isolation of Human BRCA2 Gene by Transformation-Associated Recombination in Yeast. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 7384–7387.
- (165) Kouprina, N.; Annab, L.; Graves, J.; Afshari, C.; Barrett, J. C.; Resnick, M. A.; Larionov, V. Functional Copies of a Human Gene Can Be Directly Isolated by Transformation-Associated Recombination Cloning with a Small 3' End Target Sequence. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 4469–4474.
- (166) Kouprina, N.; Larionov, V. TAR Cloning: Insights Into Gene Function, Long-Range Haplotypes and Genome Structure and Evolution. *Nat. Rev. Genet.* **2006**, *7*, 805–812.
- (167) Wenzel, S. C.; Gross, F.; Zhang, Y.; Fu, J.; Stewart, A. F.; Müller, R. Heterologous Expression of a Myxobacterial Natural Products Assembly Line in Pseudomonads via Red/ET Recombineering. *Chem. Biol.* **2005**, *12*, 349–356.
- (168) Yamanaka, K.; Ryan, K. S.; Gulder, T. A. M.; Hughes, C. C.; Moore, B. S. Flavoenzyme-Catalyzed Atropo-Selective N,C-Bipyrrole Homocoupling in Marinopyrrole Biosynthesis. *J. Am. Chem. Soc.* **2012**, *134*, 12434–12437.
- (169) Kim, J. H.; Feng, Z.; Bauer, J. D.; Kallifidas, D.; Calle, P. Y.; Brady, S. F. Cloning Large Natural Product Gene Clusters From the Environment: Piecing Environmental DNA Gene Clusters Back Together with TAR. *Biopolymers* **2010**, *93*, 833–844.
- (170) Yamanaka, K.; Reynolds, K. A.; Kersten, R. D.; Ryan, K. S.; Gonzalez, D. J.; Nizet, V.; Dorrestein, P. C.; Moore, B. S. Direct Cloning and Refactoring of a Silent Lipopeptide Biosynthetic Gene Cluster Yields the Antibiotic Taromycin a. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 1957–1962.
- (171) Bierman, M.; Logan, R.; O'Brien, K.; Seno, E. T.; Rao, R. N.; Schoner, B. E. Plasmid Cloning Vectors for the Conjugal Transfer of DNA From *Escherichia Coli* to *Streptomyces* Spp. *Gene* **1992**, *116*, 43–49.
- (172) Tang, X.; Li, J.; Millán-Aguíñaga, N.; Zhang, J. J.; O'Neill, E. C.; Ugalde, J. A.; Jensen, P. R.; Mantovani, S. M.; Moore, B. S. Identification of Thiotetronic Acid Antibiotic Biosynthetic Pathways by Target-Directed Genome Mining. *ACS Chem. Biol.* **2015**, *10*, 2841–2849.
- (173) Li, Y.; Li, Z.; Yamanaka, K.; Xu, Y.; Zhang, W.; Vlamakis, H.; Kolter, R.; Moore, B. S.; Qian, P.-Y. Directed Natural Product Biosynthesis Gene Cluster Capture and Expression in the Model Bacterium *Bacillus Subtilis*. *Sci. Rep.* **2015**, *5*, 9383.
- (174) Li, Z.-R.; Li, J.; Gu, J.-P.; Lai, J. Y. H.; Duggan, B. M.; Zhang, W.-P.; Li, Z.-L.; Li, Y.-X.; Tong, R.-B.; Xu, Y.; et al. Divergent

Biosynthesis Yields a Cytotoxic Aminomalonate-Containing Precolobactin. *Nat. Chem. Biol.* **2016**, *12*, 773–775.

(175) Jiang, W.; Zhao, X.; Gabrieli, T.; Lou, C.; Ebenstein, Y.; Zhu, T. F. Cas9-Assisted Targeting of Chromosome Segments CATCH Enables One-Step Targeted Cloning of Large Gene Clusters. *Nat. Commun.* **2015**, *6*, 8101.

(176) Jiang, W.; Zhu, T. F. Targeted Isolation and Cloning of 100-Kb Microbial Genomic Sequences by Cas9-Assisted Targeting of Chromosome Segments. *Nat. Protoc.* **2016**, *11*, 960–975.

(177) Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O. Enzymatic Assembly of DNA Molecules Up to Several Hundred Kilobases. *Nat. Methods* **2009**, *6*, 343–345.

(178) Hu, S.; Liu, Z.; Zhang, X.; Zhang, G.; Xie, Y.; Ding, X.; Mo, X.; Stewart, A. F.; Fu, J.; Zhang, Y.; Xia, L. Cre/loxP Plus BAC⁺: a Strategy for Direct Cloning of Large DNA Fragment and Its Applications in *Photobacterium luminescens* and *Agrobacterium tumefaciens*. *Sci. Rep.* **2016**, *6*, 29087.

(179) Dai, R.; Zhang, B.; Zhao, G.; Ding, X. Site-Specific Recombination for Cloning of Large DNA Fragments in Vitro. *Eng. Life Sci.* **2015**, *15*, 655–659.

(180) Baltz, R. H. Gifted Microbes for Genome Mining and Natural Product Discovery. *J. Ind. Microbiol. Biotechnol.* **2017**, *44*, 573–588.

(181) Liu, G.; Chater, K. F.; Chandra, G.; Niu, G.; Tan, H. Molecular Regulation of Antibiotic Biosynthesis in *Streptomyces*. *Microbiol. Mol. Biol. Rev.* **2013**, *77*, 112–143.

(182) Thanapipatsiri, A.; Gomez-Escribano, J. P.; Song, L.; Bibb, M. J.; Al-Bassam, M.; Chandra, G.; Thamchaipenet, A.; Challis, G. L.; Bibb, M. J. Discovery of Unusual Biaryl Polyketides by Activation of a Silent *Streptomyces venezuelae* Biosynthetic Gene Cluster. *ChemBioChem* **2016**, *17*, 2189–2198.

(183) Xie, C.; Deng, J.-J.; Wang, H.-X. Identification of AstG1, a LAL Family Regulator That Positively Controls Ansatrienin Production in *Streptomyces* Sp. XZQH13. *Curr. Microbiol.* **2015**, *70*, 859–864.

(184) Li, S.; Li, Y.; Lu, C.; Zhang, J.; Zhu, J.; Wang, H.; Shen, Y. Activating a Cryptic Ansamycin Biosynthetic Gene Cluster to Produce Three New Naphthalenic Octaketide Ansamycins with N-Pentyl and N-Butyl Side Chains. *Org. Lett.* **2015**, *17*, 3706–3709.

(185) Cobb, R. E.; Wang, Y.; Zhao, H. High-Efficiency Multiplex Genome Editing of *Streptomyces* Species Using an Engineered CRISPR/Cas System. *ACS Synth. Biol.* **2015**, *4*, 723–728.

(186) Zhang, M. M.; Wong, F. T.; Wang, Y.; Luo, S.; Lim, Y. H.; Heng, E.; Yeo, W. L.; Cobb, R. E.; Enghiad, B.; Ang, E. L.; et al. CRISPR-Cas9 Strategy for Activation of Silent *Streptomyces* Biosynthetic Gene Clusters. *Nat. Chem. Biol.* **2017**, *13*, 607–609.

(187) Tong, Y.; Charusanti, P.; Zhang, L.; Weber, T.; Lee, S. Y. CRISPR-Cas9 Based Engineering of Actinomycetal Genomes. *ACS Synth. Biol.* **2015**, *4*, 1020–1029.

(188) Bode, H. B.; Bethe, B.; Höfs, R.; Zeeck, A. Big Effects From Small Changes: Possible Ways to Explore Nature's Chemical Diversity. *ChemBioChem* **2002**, *3*, 619–627.

(189) Ochi, K.; Tanaka, Y.; Tojo, S. Activating the Expression of Bacterial Cryptic Genes by rpoB Mutations in RNA Polymerase or by Rare Earth Elements. *J. Ind. Microbiol. Biotechnol.* **2014**, *41*, 403–414.

(190) Okada, B. K.; Seyedsayamdost, M. R. Antibiotic Dialogues: Induction of Silent Biosynthetic Gene Clusters by Exogenous Small Molecules. *FEMS Microbiol. Rev.* **2017**, *41*, 19–33.

(191) Seyedsayamdost, M. R. High-Throughput Platform for the Discovery of Elicitors of Silent Bacterial Gene Clusters. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 7266–7271.

(192) Netzker, T.; Fischer, J.; Weber, J.; Mattern, D. J.; König, C. C.; Valiante, V.; Schroeckh, V.; Brakhage, A. A. Microbial Communication Leading to the Activation of Silent Fungal Secondary Metabolite Gene Clusters. *Front. Microbiol.* **2015**, *6*, 299.

(193) McDaniel, R.; Ebert-Khosla, S.; Hopwood, D. A.; Khosla, C. Rational Design of Aromatic Polyketide Natural Products by Recombinant Assembly of Enzymatic Subunits. *Nature* **1995**, *375*, 549–554.

(194) Oliynyk, M.; Brown, M. J.; Cortés, J.; Staunton, J.; Leadlay, P. F. A Hybrid Modular Polyketide Synthase Obtained by Domain Swapping. *Chem. Biol.* **1996**, *3*, 833–839.

(195) McDaniel, R.; Thamchaipenet, A.; Gustafsson, C.; Fu, H.; Betlach, M.; Betlach, M.; Ashley, G. Multiple Genetic Modifications of the Erythromycin Polyketide Synthase to Produce a Library of Novel “Unnatural” Natural Products. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 1846–1851.

(196) Weissman, K. J.; Leadlay, P. F. Combinatorial Biosynthesis of Reduced Polyketides. *Nat. Rev. Microbiol.* **2005**, *3*, 925–936.

(197) Menzella, H. G.; Reid, R.; Carney, J. R.; Chandran, S. S.; Reisinger, S. J.; Patel, K. G.; Hopwood, D. A.; Santi, D. V. Combinatorial Polyketide Biosynthesis by De Novo Design and Rearrangement of Modular Polyketide Synthase Genes. *Nat. Biotechnol.* **2005**, *23*, 1171–1176.

(198) Menzella, H. G.; Carney, J. R.; Santi, D. V. Rational Design and Assembly of Synthetic Trimodular Polyketide Synthases. *Chem. Biol.* **2007**, *14*, 143–151.

(199) Eng, C. H.; Backman, T. W. H.; Bailey, C. B.; Magnan, C.; García Martín, H.; Katz, L.; Baldi, P.; Keasling, J. D. ClusterCAD: a Computational Platform for Type I Modular Polyketide Synthase Design. *Nucleic Acids Res.* **2018**, *46*, D509–D515.

(200) Barajas, J. F.; Blake-Hedges, J. M.; Bailey, C. B.; Curran, S.; Keasling, J. D. Engineered polyketides: Synergy between protein and host level engineering. *Synth. Syst. Biotechnol.* **2017**, *2*, 147–166.

(201) Klaus, M.; Grininger, M. Engineering Strategies for Rational Polyketide Synthase Design. *Nat. Prod. Rep.* **2018**, *35*, 1070–1081.

(202) Kalkreuter, E.; Williams, G. J. Engineering Enzymatic Assembly Lines for the Production of New Antimicrobials. *Curr. Opin. Microbiol.* **2018**, *45*, 140–148.

(203) Khosla, C.; Gokhale, R. S.; Jacobsen, J. R.; Cane, D. E. Tolerance and Specificity of Polyketide Synthases. *Annu. Rev. Biochem.* **1999**, *68*, 219–253.

(204) Musiol-Kroll, E. M.; Wohlleben, W. Acyltransferases as Tools for Polyketide Synthase Engineering. *Antibiotics* **2018**, *7*, 62.

(205) Yuzawa, S.; Deng, K.; Wang, G.; Baidoo, E. E. K.; Northen, T. R.; Adams, P. D.; Katz, L.; Keasling, J. D. Comprehensive in Vitro Analysis of Acyltransferase Domain Exchanges in Modular Polyketide Synthases and Its Application for Short-Chain Ketone Production. *ACS Synth. Biol.* **2017**, *6*, 139–147.

(206) Sugimoto, Y.; Ding, L.; Ishida, K.; Hertweck, C. Rational Design of Modular Polyketide Synthases: Morphing the Aureothin Pathway Into a Luteoreticuliculin Assembly Line. *Angew. Chem., Int. Ed.* **2014**, *53*, 1560–1564.

(207) Awakawa, T.; Fujioka, T.; Zhang, L.; Hoshino, S.; Hu, Z.; Hashimoto, J.; Kozono, I.; Ikeda, H.; Shin-ya, K.; Liu, W.; Abe, I. Reprogramming of the Antimycin NRPS-PKS Assembly Lines Inspired by Gene Evolution. *Nat. Commun.* **2018**, *9*, 3534.

(208) McDaniel, R.; Kao, C. M.; Hwang, S. J.; Khosla, C. Engineered Intermodular and Intramodular Polyketide Synthase Fusions. *Chem. Biol.* **1997**, *4*, 667–674.

(209) Yoon, Y. J.; Beck, B. J.; Kim, B. S.; Kang, H. Y.; Reynolds, K. A.; Sherman, D. H. Generation of Multiple Bioactive Macrolides by Hybrid Modular Polyketide Synthases in *Streptomyces venezuelae*. *Chem. Biol.* **2002**, *9*, 203–214.

(210) Kellenberger, L.; Galloway, I. S.; Sauter, G.; Böhm, G.; Hanefeld, U.; Cortés, J.; Staunton, J.; Leadlay, P. F. A Polylinker Approach to Reductive Loop Swaps in Modular Polyketide Synthases. *ChemBioChem* **2008**, *9*, 2740–2749.

(211) Carmody, M.; Byrne, B.; Murphy, B.; Breen, C.; Lynch, S.; Flood, E.; Finnan, S.; Caffrey, P. Analysis and Manipulation of Amphotericin Biosynthetic Genes by Means of Modified Phase KC515 Transduction Techniques. *Gene* **2004**, *343*, 107–115.

(212) Crüsemann, M.; Kohlhaas, C.; Piel, J. Evolution-Guided Engineering of Nonribosomal Peptide Synthetase Adenylation Domains. *Chem. Sci.* **2013**, *4*, 1041–1045.

(213) Bozhüyük, K. A. J.; Fleischhacker, F.; Linck, A.; Wesche, F.; Tietze, A.; Niesert, C.-P.; Bode, H. B. De Novo Design and

Engineering of Non-Ribosomal Peptide Synthetases. *Nat. Chem.* **2018**, *10*, 275–281.

(214) Starcevic, A.; Diminic, J.; Zucko, J.; Elbekali, M.; Schlosser, T.; Lisfi, M.; Vukelic, A.; Long, P. F.; Hranueli, D.; Cullum, J. A Novel Docking Domain Interface Model Predicting Recombination Between Homoeologous Modular Biosynthetic Gene Clusters. *J. Ind. Microbiol. Biotechnol.* **2011**, *38*, 1295–1304.

(215) Starcevic, A.; Wolf, K.; Diminic, J.; Zucko, J.; Ruzic, I. T.; Long, P. F.; Hranueli, D.; Cullum, J. Recombinatorial Biosynthesis of Polyketides. *J. Ind. Microbiol. Biotechnol.* **2012**, *39*, 503–511.

(216) Chemler, J. A.; Tripathi, A.; Hansen, D. A.; O'Neil-Johnson, M.; Williams, R. B.; Starks, C.; Park, S. R.; Sherman, D. H. Evolution of Efficient Modular Polyketide Synthases by Homologous Recombination. *J. Am. Chem. Soc.* **2015**, *137*, 10603–10609.

(217) Wlodek, A.; Kendrew, S. G.; Coates, N. J.; Hold, A.; Pogwizd, J.; Rudder, S.; Sheehan, L. S.; Higginbotham, S. J.; Stanley-Smith, A. E.; Warneck, T.; Nur-E-Alam, M.; Radzom, M.; Martin, C. J.; Overvoorde, L.; Samborsky, M.; Alt, S.; Heine, D.; Carter, G. T.; Graziani, E. I.; Koehn, F. E.; McDonald, L.; Alanine, A.; Rodriguez Sarmiento, R. M.; Chao, S. K.; Ratni, H.; Steward, L.; Norville, I. H.; Sarkar-Tyson, M.; Moss, S. J.; Leadlay, P. F.; Wilkinson, B.; Gregory, M. A. Diversity Oriented Biosynthesis via Accelerated Evolution of Modular Gene Clusters. *Nat. Commun.* **2017**, *8*, 1206.

(218) Martinez, J. L. Antibiotics and Antibiotic Resistance Genes in Natural Environments. *Science* **2008**, *321*, 365–367.

(219) Fajardo, A.; Martínez, J. L. Antibiotics as Signals That Trigger Specific Bacterial Responses. *Curr. Opin. Microbiol.* **2008**, *11*, 161–167.

(220) D'Costa, V. M.; Griffiths, E.; Wright, G. D. Expanding the Soil Antibiotic Resistome: Exploring Environmental Diversity. *Curr. Opin. Microbiol.* **2007**, *10*, 481–489.

(221) Wright, G. D. The Antibiotic Resistome: the Nexus of Chemical and Genetic Diversity. *Nat. Rev. Microbiol.* **2007**, *5*, 175–186.

(222) Davies, J.; Davies, D. Origins and Evolution of Antibiotic Resistance. *Microbiol. Mol. Biol. Rev.* **2010**, *74*, 417–433.

(223) D'Costa, V. M.; King, C. E.; Kalan, L.; Morar, M.; Sung, W. W. L.; Schwarz, C.; Froese, D.; Zazula, G.; Calmels, F.; Debruyne, R.; et al. Antibiotic Resistance Is Ancient. *Nature* **2011**, *477*, 457–461.

(224) Palmer, A. C.; Kishony, R. Understanding, Predicting and Manipulating the Genotypic Evolution of Antibiotic Resistance. *Nat. Rev. Genet.* **2013**, *14*, 243–248.

(225) Pan, G.; Xu, Z.; Guo, Z.; Hindra, Ma, M.; Yang, D.; Zhou, H.; Gansemans, Y.; Zhu, X.; Huang, Y.; et al. Discovery of the Leinamycin Family of Natural Products by Mining Actinobacterial Genomes. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, E11131–E11140.

(226) Wang, Y.; Marrero, M. C.; Medema, M. H.; van Dijk, A. D. J. Coevolution-Based Prediction of Protein-Protein Interactions in Polyketide Biosynthetic Assembly Lines. *bioRxiv* **2019**. DOI: 10.1101/669291

(227) Ueoka, R.; Uria, A. R.; Reiter, S.; Mori, T.; Karbaum, P.; Peters, E. E.; Helfrich, E. J. N.; Morinaka, B. I.; Gugger, M.; Takeyama, H.; et al. Metabolic and Evolutionary Origin of Actin-Binding Polyketides From Diverse Organisms. *Nat. Chem. Biol.* **2015**, *11*, 705–712.

(228) Lackner, G.; Peters, E. E.; Helfrich, E. J. N.; Piel, J. Insights Into the Lifestyle of Uncultured Bacterial Natural Product Factories Associated with Marine Sponges. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, E347–E356.

(229) Libis, V.; Antonovsky, N.; Zhang, M.; Shang, Z.; Montiel, D.; Maniko, J.; Ternei, M. A.; Calle, P. Y.; Lemetre, C.; Owen, J. G.; Brady, S. F. Uncovering the Biosynthetic Potential of Rare Metagenomic DNA Using Co-Occurrence Network Analysis of Targeted Sequences. *Nat. Commun.* **2019**, *10*, 3848–3849.

(230) Xue, M.; Kim, C. S.; Healy, A. R.; Wernke, K. M.; Wang, Z.; Frischling, M. C.; Shine, E. E.; Wang, W.; Herzon, S. B.; Crawford, J. M. Structure Elucidation of Colibactin and Its DNA Cross-Links. *Science* **2019**, *365*, No. eaax2685.

(231) Li, Z.-R.; Li, J.; Cai, W.; Lai, J. Y. H.; McKinnie, S. M. K.; Zhang, W.-P.; Moore, B. S.; Zhang, W.; Qian, P.-Y. Macrocyclic Colibactin Induces DNA Double-Strand Breaks via Copper-Mediated Oxidative Cleavage. *Nat. Chem.* **2019**, *11*, 880–889.