Natural Product Reports

REVIEW

View Article Online
View Journal

Cite this: DOI: 10.1039/c9np00048h

Evolutionary dynamics of natural product biosynthesis in bacteria

Marc G. Chevrette,[a] Karina Gutiérrez-García,[b] Nelly Selem-Mojica,[b] César Aguilar-Martínez,[b] Alan Yañez-Olvera,[b] Hilda E. Ramos-Aboites,[b] Paul A. Hoskisson[c] and Francisco Barona-Gómez*[b]

Covering: 2008 up to 2019

The forces of biochemical adaptive evolution operate at the level of genes, manifesting in complex phenotypes and the global biodiversity of proteins and metabolites. While evolutionary histories have been deciphered for some other complex traits, the origins of natural product biosynthesis largely remain a mystery. This fundamental knowledge gap is surprising given the many decades of research probing the genetic, chemical, and biophysical mechanisms of bacterial natural product biosynthesis. Recently, evolutionary thinking has begun to permeate this otherwise mechanistically dominated field. Natural products are now sometimes referred to as 'specialized' rather than 'secondary' metabolites, reinforcing the importance of their biological and ecological functions. Here, we review known evolutionary mechanisms underlying the overwhelming chemical diversity of bacterial secondary metabolism, focusing on enzyme promiscuity and the evolution of enzymatic domains that enable metabolic traits. We discuss the mechanisms that drive the assembly of natural product biosynthetic gene clusters and propose formal definitions for 'specialized' and 'secondary' metabolism. We further explore how biosynthetic gene clusters evolve to synthesize related molecular species, and in turn how the biological and ecological roles that emerge from metabolic diversity are acted on by selection. Finally, we reconcile chemical, functional, and genetic data into an evolutionary model, the dynamic chemical matrix evolutionary hypothesis, in which the relationships between chemical distance, biomolecular activity, and relative fitness shape adaptive landscapes.

Received 4th August 2019

DOI: 10.1039/c9np00048h

rsc.li/npr

1. Introduction
1.1. Key concepts on the evolution of natural product biosynthesis
1.1.1. Enzyme promiscuity in bacterial natural product biosynthesis
1.1.2. A Darwinian view of natural product biosynthesis
1.1.3. Evolution of bacterial populations
1.2. Updated views on the evolution of secondary metabolism
2. Evolution of modular natural product biosynthetic enzymes
2.1. Revisiting PKS and NRPS modularity in light of evolution
2.2. Evolution of domain–domain interactions in PKS and NRPS enzymes
3. Evolution of non-modular natural product biosynthetic enzymes
3.1. Evolution of single-domain enzymes in natural product biosynthesis
3.1.1. Evolution of precursor supply enzymes and metabolic integration
3.2. Evolution of enzymes related to resistance and immunity mechanisms
4. Evolutionary dynamics at different scales: from domains to BGCs
4.1. Phylogenomics of BGCs: revealing key biosynthetic genes driving chemical diversity
4.1.1. Main genetic mechanisms driving chemical innovation
4.1.2. Increased chemical diversity by interactions of multiple biosynthetic genes
4.1.3. High metabolic diversity synthesized by conserved BGCs
4.2. Fixation, distribution, and selective sweeps in BGCs

[a]Wisconsin Institute for Discovery, Department of Plant Pathology, University of Wisconsin-Madison, Madison, WI, USA. E-mail: chevrette@wisc.edu

[b]Evolution of Metabolic Diversity Laboratory, Unidad de Genómica Avanzada (Langebio), Cinvestav-IPN, Irapuato, Guanajuato, Mexico. E-mail: francisco.barona@cinvestav.mx

[c]Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, Glasgow, UK. E-mail: paul.hoskisson@strath.ac.uk

5. Unifying concepts to bridge evolutionary theory and the biosynthesis of natural products
5.1. Defining secondary and specialized metabolism
5.2. One pathway, many compounds: how and why?
5.2.1. The dynamic chemical matrix evolutionary hypothesis
6. Concluding remarks and perspectives
7. Definitions
8. Conflicts of interest
9. Acknowledgements
10. References

1. Introduction

The chemical products of metabolism shape the abounding diversity of life. Natural products, also referred to as secondary or specialized metabolites, are a major means through which the microbial world communicates and modulates interactions within and between species. Metabolite-mediated interactions can influence evolutionary fitness landscapes through direct antagonism, niche defense, and signaling. Apart from their ecological functions, bacterial natural products have been used as therapeutics for cancer and other human diseases and are the main source of clinically used antimicrobials to combat infectious diseases. Other uses (*e.g.* as immunosuppressants, food additives, crop protectant agents, *etc.*) further highlight their paramount importance in modern society. Exponentially growing genomic, metabolomic, and other evidence has enabled the characterization of the genes and functions responsible for the assembly, transport, regulation, and ultimately biological role(s) of natural products. Investigations into the evolutionary forces acting on natural products can both offer insight into microbial interactions and guide discovery efforts for new molecules of therapeutic or commercial value.

1.1. Key concepts on the evolution of natural product biosynthesis

Herein, we aim to refine concepts concerning the evolution of enzymes within natural product biosynthesis. Themes from bacterial population genetics are integrated to provide additional evolutionary context. It is not our intent to comprehensively discuss all principles needed to describe the evolution of complex traits. Rather, we explore emerging concepts from the last decade as they directly apply to secondary metabolism and their natural products. Additionally, we refer readers to early reviews on bacterial natural product phylogenetics, by Jenke-Kodama & Dittmann[1] and Schmitt & Barker,[2] as well as a recent update in the role of phylogenetics in natural products research by Ziemert and co-workers.[3] These works provide a foundation for the phylogenetic approaches developed over the last decade to mine genomes for novel natural products and describe their evolutionary histories. Thus, natural product research has progressed beyond mere speculation of biosynthetic origins and moved towards rigorous evolutionary models. Furthermore, we reappraise the seminal reviews of evolution of secondary metabolism in bacteria from a decade ago (Subsection 1.2) to provide a historical account of the rapid progression of evolutionary thinking in natural product biosynthesis since the last revision of this topic. Subsequently, the remaining sections address key recent advances (including both scientific literature and genomic databases) that have further shaped our understanding of secondary metabolism and the evolution of chemical diversity. We go on to propose an evolutionary model to explain how new biosynthetic gene clusters (BGCs) are assembled and disassembled under different evolutionary forces (*e.g.* neutral forces or different types of selection) which we term the Dynamic Chemical Matrix Evolutionary (DCME) hypothesis. We also explicitly define common terms used by the natural product research community (*i.e.* secondary and specialized metabolism) such that they become evolutionarily informative. A comprehensive list of natural product biosynthetic enzymes and hypotheses regarding their specific evolution is beyond the scope of this work and not provided here. Rather, we discuss selected, concept-driven examples in detail.

1.1.1. Enzyme promiscuity in bacterial natural product biosynthesis. A wide range of perspectives concerning the evolution of enzyme function highlight the promise of rational protein engineering towards specific molecules or activities.[4,5] Harnessing enzyme promiscuity (*i.e.* an enzyme's ability to accept more than one substrate) is a common and recurring concept in protein engineering, supported by experimental evidence across many enzymes and biochemical pathways.[6,7] Enzyme promiscuity and functional evolution are closely linked: compared to enzymes with strict substrate specificities, "generalist" enzymes require fewer mutations to attain new activities and thus have a more parsimonious route towards neofunctionalization.[8–11] Alterations in substrate specificity are a common mechanism of functional enzyme evolution, as the overall catalytic mechanism of the reaction is often bound by tight physicochemical and physiological constraints.[12,13] In nature, enzymes rarely evolve high catalytic efficiency (k_{cat}/K_M).[4,14] Even if the diffusion-rate limit is reasonably attainable, the random and stepwise nature of evolution by mutation makes maintained catalytic efficiency improbable and often inaccessible.[4,14] Instead of evolving towards optimal efficiency, most enzymes instead exhibit only sufficient catalytic conversion rates.[4,14]

The stability of protein configuration also influences functional evolution. Mutations affecting protein stability (and, by extension, flexibility) affect the interplay between protein conformation and enzymatic function.[15–17] As a protein fold accommodating the active site of an enzyme becomes increasingly flexible, the likelihood that it may accept more than one substrate increases. If flexibility becomes too great, however, it may then result in an unstable, non-functional enzyme. Thus, the functional evolution of enzymes is only possible within the thresholds of biophysical constraints. These bounds can be easily exceeded if robustness against mutation is precarious.[18] Ancestral sequence reconstructions suggest broader substrate specificities in ancestral enzyme forms,[19] further supporting the hypothesis that enzyme promiscuity allows for innovation of novel enzyme function. Unfortunately, there are currently only a few reports on the ancestral reconstruction of natural product biosynthetic enzymes.[20] The recent work from Waglechner *et al.*, describing a dated reconstruction of the genes and domains

involved in glycopeptide biosynthesis, suggests that ancestral reconstructions may start to be adopted.[21] Future studies on ancestral enzyme and/or pathway states will help us generate and test evolutionary hypotheses, potentially informing the development of biocatalysts and engineering tools.

Enzyme promiscuity is a function of neutral evolutionary processes, implying it is not selected for under adaptive (Darwinian) evolution.[10] However, promiscuity provides a sort of standing variation upon which adaptive forces can act. Experimental evidence provided by the directed evolution of metallo-β-lactamase enzyme families suggests that cryptic genetic variation shapes the adaptive evolutionary potential of enzymes, even when catalytic efficiencies are not efficient by human standards.[22] In primary metabolism, positive selection purges promiscuity, but can also result in generalist enzymes with proficient overall activities.[23–25] However, these observations may be more related to the avoidance of binding undesirable substrates rather than optimizing the primary reaction. Biochemical data profiling the promiscuity of bacterial amino acid biosynthesis further supports the role of genetic cryptic variation and neutral evolution in enzyme evolution.[26,27] Often, it is assumed that if enzymes share a common evolutionary origin, they will also have the same function. This is not necessarily the case. Some initiatives, such as the quest for orthologs consortium, have sought to standardize and benchmark ortholog predictions, yet much work remains to be done. Orthologs and paralogs with high sequence similarity can exhibit extremely different activity profiles.[26,27] Further, subtle changes in sequence can have dramatic impacts on the substrate specificity, even when the structural or catalytic contributions of mutations are not obvious.[23,28,29] It is therefore tempting to speculate that enzyme promiscuity is responsible for the overwhelming metabolic diversity observed in bacteria, the majority of which has yet to be explored. Together, the evolutionary forces acting within central metabolism imply enzymes are more phenotypically plastic than appreciated in the classical 'lock and key' model found in introductory textbooks.

The evolutionary features associated with enzyme neofunctionalization are broadly present in enzymes and pathways for the biosynthesis of natural products (sometimes called secondary metabolites or specialized metabolites). For instance, enzyme promiscuity has been broadly documented in plant specialized metabolism.[30] The promiscuity of enzymes in natural product biosynthesis may be exploited to accept non-natural substrates fed to fermentations, to create 'unnatural natural products' through combining synthetic chemistry and synthetic biology approaches.[31,32] Irrespective of the term used, or the application in which promiscuity is leveraged, *in vivo* metabolomics[33,34] and large-scale data analyses[35] suggest natural product biosynthetic enzymes can accept many more substrates than those we can measure *in vitro* and/or model from enzyme kinetics, and that much chemical novelty remains to be discovered. Moreover, enzymes from secondary metabolism can be promiscuous in multiple ways, sometimes in the substrates they accept and sometimes in the reaction mechanisms they follow.[36] Neither promiscuity is generally observed in enzymes from primary metabolism that are under tighter purifying selection.

The sesquiterpene cyclases and type III polyketide synthases biosynthesize a wide range of natural products. The promiscuous nature of these enzymes to accept a range of substrates, coupled with their intrinsic flexibility, has largely been attributed to the reactivity of the linear transition-state intermediates generated during biosynthesis.[36] The intimate relationships between these enzymes and their respective substrates results in the production of a wide range of (often difficult to predict) products.[37] The overall promiscuity of the enzyme allows biosynthesis to expand beyond substrate diversity and include catalytic promiscuity. This results in the remarkable diversity of natural products that we observe, the majority of which have not yet been placed in an evolutionary context. The large population size of bacteria in the environment may mean that in addition to natural selection the non-adaptive processes of genetic drift, mutation, and recombination may also influence evolution of such pathways.[38] Moreover, diversity driven by enzyme promiscuity in natural product biosynthesis is different to the combinatorial potential of some biosynthetic pathways. This is exemplified by the Ribosomally synthesized and Post-translationally modified Peptides (RiPPs) whose biosynthetic enzymes can exhibit extreme promiscuity: the number of final biosynthesized products increases through diversity of the starting substrate(s), irrespective of the biosynthetic enzymes of the pathways themselves.[39,40]

1.1.2. A Darwinian view of natural product biosynthesis. All cellular life depends on metabolism. It is amongst the most genetically and phenotypically complex traits on Earth. Metabolic responses are the main mechanisms through which sessile organisms cope with environmental change and associated biotic and abiotic stresses. Plasticity in metabolic response has allowed bacteria to dominate life on Earth. In recent years our understanding of the inter- and intra-species exchange of metabolic products has exponentially grown, leading to an appreciation that metabolites mediate a vast array of ecological interactions.[41] This much needed ecological context to metabolism offers insight into the selective pressures operating on bacterial populations which drive the evolution of metabolism and, in particular, the evolution of natural products. The forces that modulate the molecular features that underpin enzyme evolution still require further study. Nevertheless, the 'Darwinian view on metabolism' from Firn[42] and Firn and Jones[43] suggests that selection may act distinctly on different metabolic sub-networks and/or pathways.

Firn and co-workers propose a framework in which the molecular characteristics of metabolic intermediates may influence the evolutionary trajectories of both the pathway itself and the overall network of metabolism.[42–44] Revising ill-defined and uninformative nomenclature of 'primary' and 'secondary' metabolism, these authors suggest alternative terms that instead reflect evolutionary selective forces. For example, the selection acting on amino acid biosynthetic pathway intermediates relates to the properties of the final product of that pathway, such that precursors and enzymes in the pathway will

be subject to selective pressure to avoid congeners. Purifying selection operates throughout the pathway, reducing enzyme promiscuity at every step. Firn and Jones call this example 'integrated' metabolism, which implies the molecular features of pathway intermediates are only important as they are converted into the derived property of that pathway (*e.g.* an amino acid or other bioactive metabolites).

Firn and Jones[43] considered two alternative defining molecular features of metabolism that reflect the intrinsic properties of a pathway. A pathway may be under selection due to the physicochemical properties that dictate its function. For example, ferric iron-chelating siderophores can leverage free electrons from a range of atoms within diverse moieties (*e.g.* N, O, or S). This reflects the diversity in chemical (and therefore biosynthetic) routes to the physicochemical molecular features needed and partially explains the metabolic diversity exhibited by siderophores in nature.[45] These pathways have sometimes been referred to as "supportive" metabolism as they often appear less integrated into the pathways of central metabolism and perhaps neutral evolution can more freely sample the evolutionary and chemical space of their metabolic subnetworks. This may also explain why the so-called supportive metabolism resembles networks of molecular species rather than the discrete pathways of a succession of interconvertible molecules in integrated metabolism.

Molecular evolution is also driven by biomolecular activity, defined by Firn and Jones[43] as how a metabolite has an affinity towards its molecular target (*e.g.* proteins, membranes, nucleic acids *etc.*). Since there is an astounding molecular complexity of potential targets and metabolites it is unlikely that a metabolite alone will have a strong influence on biomolecular activity simply *via* random perturbation. However, this likelihood must be considered in the context of the dynamic fitness landscape and massive timescales in which evolution operates (*i.e.* eco-evolutionary dynamics). Enzyme promiscuity enables the exploration of chemical space, "screening" a range of outcomes against a backdrop of chance. The small likelihood of increasing activity may lead to fixation of the trait. This can be considered analogous to the screening of microorganisms during the 'Golden Age of antibiotic discovery' in the 1950s–70s.[44] Therefore, "speculative" pathways may evolve towards increased chemical diversity, yet to achieve this, promiscuity in pathway enzymes must be tolerated to enable adaptation and for fixation to occur. Such promiscuity likely arises through neutral evolution, relaxed purifying selection, or positive selection towards chemical diversity. Evidence of the latter may come from the occurrence of selective sweeps that have been suggested to fix BGCs during local adaptation.[46]

1.1.3. Evolution of bacterial populations. A major distinction between the evolution of natural product biosynthesis in prokaryotes and eukaryotes lies in the different genetics that govern heritability. In eukaryotes, genetic diversity and innovation is often driven by meiotic recombination and/or regular gene duplication.[47] In contrast, prokaryotes employ alternate strategies to achieve genetic diversity, including conjugation, asexual recombination, and (compared to eukaryotes) relatively higher rates of horizontal gene transfer (HGT). Prokaryotes also have larger effective population sizes.[48] In turn, there may be very different rates and types of gene set expansions between these two groups of organisms despite their enzyme products may behave mechanistically similarly. Here, we highlight concepts related to bacterial populations that have emerged in recent years.

Genetic flux in bacterial populations depends on HGT and the ability of incoming DNA to recombine into the recipient genome.[49] The rates and ability of natural product producing organisms to participate in HGT varies widely. In *Streptomyces* there is limited evidence for natural competence or widespread functional transducing phages, but there is extensive HGT *via* conjugation.[49–53] Evidence for HGT in myxobacteria is extensive although mechanistic studies have so far been limited.[54] Whether rates of HGT are high or low in natural product producers is difficult to discern unequivocally. This is in part due to horizontally acquired DNA being easy to identify in bacterial genomes when recently acquired from phylogenetically different species from genomics (DNA with different GC content and presence of phage sequence, insertion sequences or transposons following HGT). What is less easy to identify is incoming DNA from closely related strains and species, and this is currently an area of active research.

Bacterial evolution is rapid. Within metabolic pathways, there has been a tendency to focus on the efficiency and specificity of enzymes rather than the mechanisms by which novel metabolic functions can evolve. The evolution of true biological novelty is rare. Gene expansion events are an important force that provides organisms the opportunity to evolve new functions.[55] Gene expansions fall into two categories – HGT and gene duplications. Regardless of the origin of new genetic material, expansion and divergence are cornerstones of genetic diversification. Transient periods of elevated gene dosage are key to the emergence of new function.[56] Gene expansion events that do not provide a fitness advantage are likely to be rapidly lost from bacterial genomes, and therefore populations, given the strong mutational bias towards deletion in bacteria.[57] This suggests that expansions visible in extant genomes were successful, individually adaptive,[58] or very recent. BGCs can be the cargo of HGT over geologic timescales,[59,60] yet how this may bring about diversification and recombination within clusters is poorly understood. 'Pathway swapping' allows homologous core genes within BGCs to be exchanged or to associate with additional genes and diversify biosynthesis,[61] emphasizing the importance of population dynamics on the evolution of BGCs.

Adaptation shapes the diversity that we see in nature. Environmental and ecological conditions shape the adaptation of bacterial populations. Evolution acts on populations, each ecological niche offering its own specific challenges.[62,63] The large effective population sizes of bacteria suggest that natural selection can overcome random genetic drift, with deleterious variants rapidly purged from genomes by selective sweeps.[48] In a BGC context, horizontal acquisition of a BGC allows the recipient to sample novel ecological niches following dispersal, thus creating new ecological lineages.[64,65] However, such scenarios are reliant on sufficient genetic exchange within the population, which may not always be plausible in populations

with large, diffuse, or rapidly changing boundaries. Gene flow can be especially limited in populations from environments that are spatially structured and diverse (*e.g.* soil).[66] Recent work from Ziemert and colleagues[67] describe that mobilization and functional replacement of BGCs can occur in natural environments. These authors demonstrated that salinichelin siderophores have replaced desferrioxamine siderophores in the genomes of the marine actinobacterium *Salinispora* on at least three independent occasions during the evolution of the genus. This example of pathway replacement is consistent with there being high-selective pressure against retention of functionally redundant gene clusters.

1.2. Updated views on the evolution of secondary metabolism

The explosion in DNA sequencing data, advances in methodology for functional characterization of genomes (*e.g.* transcriptomics, proteomics and metabolomics), and advances in molecular tools have expedited molecular evolutionary analyses of secondary metabolism and its associated enzymes. We will critically analyze and discuss the seminal work in this field from previous decades and update these ideas considering recent advances. We begin with the works of Jenke-Kodama, Dittmann and co-workers on the phylogenetic analysis of arguably the most intensively studied group of enzymes from secondary metabolism, the polyketide synthases (PKSs).[1,68–71] We then revisit the work of Fischbach, Walsh, and Clardy[72,73] which emphasizes chemical innovation and the evolution of gene collectives in bacteria.

Phylogenetic analyses of the diverse and widespread PKS megasynthase family have established an evolutionary framework where PKSs are related to the fatty acid synthases (FASs) of central metabolism.[1,68–71] The proposed relationship between FASs and PKSs is not a new idea. Nevertheless, until the rigorous phylogenetic analysis of Jenke-Kodama and co-workers, arguments were only speculative and circumstantial (though visionary).[74,75] Jenke-Kodama *et al.* provide a broad view of PKS evolution, bringing together the different enzyme classes from across the bacterial phylogeny. Type I and type II PKSs exhibit modular and iterative biosynthetic logic, respectively, in which key acyltransferase (AT) and ketosynthase (KS) domains and the variability present in the dehydratase (DH) and ketoreductase (KR) domains dictate the assembly of metabolic products. Phylogenetic reconstructions from these authors encompassing all FAS and PKS biosynthetic enzymes describe a crucial ancestral bifurcation in which the FabF branch begets FASs and type I PKSs, and the FabH branch begets type II and III PKSs.

One observation that puzzled Jenke-Kodama *et al.*[1] was the inconsistent taxonomic distribution of secondary metabolite biosynthetic clusters, which they assumed was due to restricted horizontal gene transfer (HGT). This is a fundamental observation, born from their studies of PKS enzymes, and may reflect the ancestral origin of these clusters within genomes. This is consistent with the low estimates of HGT[76] and suggestions that there is an underappreciated role of vertical BGC evolution[64] within the streptomycetes and other natural product-producing bacteria. Recently, the ancestral origin of the natural product scytonemin, a so-called 'sunscreen peptide' from cyanobacteria was investigated and found to date back to the Great Oxidation Event of around two billion years ago.[60] Phylogenetic reconstructions have also shown similar results for *Streptomyces* natural products such as the aminoglycoside antibiotic streptomycin,[64] and for the hybrid polyketide/peptide antimycin.[77] These results reinforce the idea that natural product pathways are ancient genetic elements.

Investigating the genetic basis for PKS diversification in a single bacterial genome (*Streptomyces avermitilis*), Jenke-Kodama *et al.*[69] demonstrated that the expansion and divergence of most PKS domains (AT, DH, and KR), but remarkably not the KS domain, resulted in polyketide diversity, including the avermectins, oligomycin, and a polyene macrolide. This implies that PKSs not only evolve widely throughout bacterial lineages, but also within species. This observation has further been supported by recent work that indicates that the evolutionary history of each group of natural product biosynthetic enzymes must be considered if we are to better exploit the combinatorial biosynthesis and to integrate natural products in to synthetic biology strategies.[78,79] The work of Jenke-Kodama and co-workers, was a landmark in our understanding and ability to decipher the evolutionary mechanisms that underpin polyketide diversity that humankind has exploited so well in the clinic and in agriculture.

The idea of 'gene collectives', as a central genetic trait to aid our understanding of the evolution of natural product BGCs in bacteria has been proposed.[72,73] Although these authors did not reconstruct evolutionary histories of natural product biosynthetic genes or pathways, they reviewed the available literature with an emphasis on chemical innovation spanning two major levels: (i) how natural product biosynthetic genes may have evolved and (ii) how BGCs may evolve after recruitment of biosynthetic genes. They provide plausible genetic mechanisms for the emergence of BGCs. Given this was the state-of-the-art at the time, these authors adopted a BGC-centric view of this problem, in which BGCs were considered the main and only final-product of the evolutionary process sustaining metabolic diversity. This bias is related to the recurrent experience in both industrial and academic settings whereby a BGC strongly relates to a measurable biomolecular activity (albeit within a narrow biological framework of assays) which may be attributed to a single chemical structure. The concept of BGCs has proven incredibly valuable and led to advances in the field of natural products, as it has sustained many of the early recombinant technologies and still supports state-of-the-art synthetic biology approaches.[80–83] A BGC-centric view, however, is insufficient when trying to understand the evolutionary origins of the observed chemical diversity.

Given a lack of evolutionary reconstructions and the emphasis on a single final-product, the gene collective concept only partially accounts for the genetic mechanisms and selective pressures that drive metabolic diversity. This is an important distinction which may impact the general conclusions drawn from earlier work – "mutation and natural selection [are]

potent forces driving chemical innovation by creating new molecules with different biological activities from their ancestors".[72] Fischbach *et al.*[72] and Walsh and Fischbach[73] acknowledged that the diversification suggested by their examples posed an evolutionary challenge, as the lifetime of non-functional BGCs would be limited in nature. They argued in favor of intermediary gene clusters that must produce a useful molecule to justify their maintenance within the genome long enough to be selected for prior to assembly into a BGC. However, the discussion they offered was limited to possible benefits related to discrete biological functions (*e.g.* siderophore activity by intermediaries). A major issue with this proposal is that such a benefit must be solved in a case-by-case basis, an unlikely scenario given the large metabolic diversity arising from both convergent and divergent evolution. This apparent conundrum could be solved by the DCME hypothesis (see 5.2.1 for discussion), in which evolutionary mechanisms and selective pressures are integrated. Here, the fitness landscape as it relates to the chemical product(s) of secondary and specialized metabolism (see Section 5.1 for discussion on terminology) need not be a static constraint, as in nature ever-changing environments and selective pressures are likely the norm at the population level. Thus, the benefit of each intermediary within the emerging metabolic matrix is part of a whole process upon which selective pressures can operate.

The examples called upon by these authors to advance the idea of gene collectives are reviewed and discussed in an authoritative manner, describing the pervasive nature of both accelerated and convergent evolution of secondary metabolism. Why is it that convergent evolution takes place more often than expected, for such a costly genetic trait, especially considering the vertical evolution seen in PKSs? This remains a key, yet unanswered, question. In part, the answers may be related to two very interesting predictions made by Fischbach and co-workers: (i) the idea that BGCs evolved from sub-clusters coding for enzymes recruited from central metabolism; and (ii) that enzyme modularity influences the evolvability of natural product biosynthesis as a complex genetic trait.[72,73] Moreover, these authors anticipated that conjugative enzymes may allow for exploration of the chemical space after combinatorial assembly of sub-clusters, in analogy to what happens in PKS and NRPS biosynthetic systems where condensation domains contribute to this role. This possibility, however, could not be thoroughly explored at the time, as the whole gene-cluster sequence phylogenies of today[84,85] were not yet available. For examples of the latter refer to Section 4.1.

2. Evolution of modular natural product biosynthetic enzymes

Most evolutionary studies in natural product biosynthesis have focused on the PKS enzyme family. Phylogenetic analyses focusing on individual domains have reconstructed an evolutionary history of the three main types of PKSs (see Section 1.2). This work underscores the expansion-and-recruitment events that are characteristic of these FAS–PKS systems. More recent studies have reinforced these early ideas and provided new hypotheses for the chemical diversity observed from a single PKS pathway within a given strain. This is exemplified by the work of Hoefler *et al.*[86] that links bacteria physiology and secondary metabolism to describe the colony-localized, antibiotic linearmycins. In addition to gene expansions and duplications, PKSs also evolve through insertions and deletions[87] and by natural hybridizations (*i.e.* natural combinatorial biosynthesis) which has been documented for both PKSs and non-ribosomal peptide synthetases.[79,88] These genetic mechanisms can modulate chemical diversity and have consequently been exploited to produce 'unnatural' natural products, first through combinatorial biosynthesis (largely by deletion experiments) and more recently through evolution-driven synthetic biology (discussed throughout this paper).

PKSs and NRPSs can evolve through concerted evolution where highly similar DNA recombines through homologous recombination, concomitant with gene conversion in homologs. This has profound implications in the evolution of highly homologous stretches of sequence, as in the modular megasynthases. In turn, concerted evolution may serve to drive increased similarity.[78,79,89] Consequently, concerted evolution can confound phylogenetic analyses, masking early evolutionary events. Moreover, and pertinent to the evolution of natural product biosynthetic enzymes, while concerted evolution operates within multi-domain modular enzymes, it likely acts as a homogenizing force more widely in BGCs. This homogeneity increases the likelihood of recombination events within and between PKSs and NRPSs, leading to an efficient mechanism to explore the chemical space. Importantly, this is bounded by both sequence similarity and function: sequences that are too similar may not affect chemistry; sequences that are too divergent may not recombine at all, and those that do may not function within an existing modular assembly line. The modularity of secondary metabolic enzymes is at the core of the remarkable evolvability of secondary metabolism as a whole. Considering these observations, it would be fascinating to investigate the wider role of concerted evolution across different natural product biosynthetic enzyme classes to investigate whether these processes operate in non-modular enzymes as well.

It remains poorly understood how enzyme modularity can influence the complexity of natural product biosynthesis, given that evidence of independent evolution and concerted evolution is seen throughout the phylogenies of natural product biosynthetic enzymes and gene clusters.[79] To this end, a comprehensive analysis of the PKS and NRPS gene families is required, in which the role of co-evolution and co-diversification is examined within and between the enzymatic modules. Such studies will provide a more comprehensive and holistic view of multimodular enzyme evolution. Recently, this has been attempted,[90] revealing unexpected observations with key implications for our fundamental understanding of these modular enzymes. Similarly, what constitutes a 'true' module, at least in certain PKSs, has recently been redefined due to evolutionary analyses. Through the investigation of four closely related, giant modular PKS biosynthetic systems responsible for the synthesis of

aminopolyol polyketides (mediomycin, neomediomycin, ECO-02301, and neotetrafibricin), Abe and co-workers showed that within these systems the processing domains evolutionarily co-migrate with the downstream KS domain rather than the KS upstream, as previously postulated.[90] While the rational engineering of PKSs has been problematic to date, these findings offer hope in applying evolutionary approaches to increase successful engineering outcomes.[91]

In contrast, the natural history of NRPSs has yet to be reconstructed and investigations into the origins of NRPS domains that contribute to chemical diversity have only just begun. Early phylogenetic analyses of adenylation[92,93] (A) and condensation[94] (C) domains suggested a domain-level phylogenetic signal that drives specificity towards their aminoacyl adenylated substrates, yet relatively few evolutionary studies have recently being reported. A notable exception is in the case of the formylation (F) domain, which provides a potential link between central and secondary metabolism.[95] The recent accumulation of high quality structural data for NRPSs[96,97] and their tailoring enzymes[98,99] provides an unprecedented opportunity to embark in combined phylogenetic and functional evolutionary analyses. This structural data can be leveraged by new understanding about NRPS modules organization. Three notable synthetic biology examples are the natural combinatorial biosynthesis of two pyrrole amide NRPS BGCs giving rise to three pyrrole amide antibiotics;[88] dissection of the NRPS sustaining microcystin diversity in cyanobacteria;[100] and the novel nonribosomal peptide engineering by Bode and co-workers.[101,102] These advances have opened the door for *de novo* combinatorial synthesis of unnatural natural peptides, inspired by evolution, driven by synthetic biology.

2.1. Revisiting PKS and NRPS modularity in light of evolution

PKS and NRPS modularity is central to our understanding of the biosynthesis of natural products. Current models to explain the evolution of these enzymes have relied on the modular biosynthetic assembly mechanisms associated with these megasynthases. The proposal that universal genetic mechanisms are at work within these BGCs, including gene expansions (or duplication), recombination, and horizontal gene transfer,[69,72,73,79] may begin to account for the appearance of multidomain enzymes responsible for assembly and diversity within the PKSs and NRPSs. These forces can act on individual BGCs with well-defined genetic boundaries and discrete (and measurable) biomolecular activity linked to a chemical structure. The latter is particularly true for bacteria, but clustered secondary metabolism genes are also becoming increasingly recognized in plants, fungi, and nematodes,[103,104] suggesting that multi-domain, modular PKS and NRPS are the standard form taken by BGCs irrespective of taxonomic lineage. Ebony, one of the most well studied proteins in *Drosophila*, is an NRPS that is linked to cuticle pigmentation and various neuronal activities.[105] Not only is this an example of NRPSs in higher eukaryotes, but it is a major locus for adaptive radiation. Regulatory mutations that affect Ebony drive the adaptive evolution of pigmentation within and between *Drosophila* species.[106–109]

During the last decade, the broad existence of modularity that defines the megasynthases has been challenged from two perspectives that, in part, stem from a proposal to cease PKS classification based on their modularity or their taxonomic position in the absence of large-scale data-driven analyses.[110] The first perspective challenging modularity arose from exploration of large sequence databases representing all domains of life.[111] These analyses demonstrated a universal occurrence of PKS and NRPS domains in a wide range of taxa, beyond those usually expected to produce polyketide or nonribosomal peptide small molecules. Recent work demonstrating widespread hybrid PK/NRP production by nematodes[104] suggests the picture is not as simple as we once thought. Interestingly, the large-scale analyses of Wang *et al.*[111] suggested that approximately half of the PKS and NRPS enzymes are accounted for by non-modular biosynthetic systems. This may indicate that they are part of a BGC or encoded by sub-clusters situated elsewhere in the chromosome, yet belonging to a concerted biosynthetic pathway.[79] In recent years support for the latter hypothesis has come from siderophore biosynthetic pathways,[111–113] but it may yet represent a more general phenomenon in other metabolites with critical ecological functions.

Secondly, as more polyketide and nonribosomal peptide pathways are experimentally characterized, increasing numbers of single-domain enzymes acting *in trans* are confirmed to carry out the enzymatic conversions necessary for the synthesis of the final natural product. This is exemplified by the *trans*-AT polyketide synthases,[71,87,114,115] where single-domain enzymes found within the BGC interact with a 'main' multidomain PKS or NRPS. There are also more extreme cases where enzymes with one or more domains are located elsewhere in the genome yet interact with the core BGC. Interestingly, in part due to chance, and as a consequence of large scale 'omics technologies,[116] evidence is growing for several biosynthetic systems interact across distinct and distant genomic loci to generate secondary metabolites with bioactivity.[117,118] While originally deemed exceptional, these cases may be more widespread than previously appreciated. These phenomena have largely been overlooked, particularly in bacteria, due to the pervasive BGC-centric view of secondary metabolism. Further study is needed to connect non-modularity to the evolution of natural products and the analysis of heterologously expressed BGCs.

The distribution of enzymes contributing to secondary metabolism across the genome is an important concept, especially for those who study bacterial systems. Genomic location of BGCs and enzymes is now reported by some automated, natural product genome-mining tools.[119] It is likely that a locus within a BGC would evolve under different constraints and rates compared to loci in BGCs elsewhere in the genome. Considering the pioneering work of Jenke-Kodama and co-workers,[69] would their phylogenetic reconstructions of PKS domains tell the same story if single-domain enzymes were considered? Work addressing this question has appeared in the last decade or so, largely in the context of bioprospecting.[84,120,121] However, we are still lacking comprehensive phylogenies for both

modular PKSs and NRPSs, that account for the inter-domain interactions known to be core evolutionary force acting on these biosynthetic systems. Inter-domain signaling and/or compatibility as an evolutionary force has been drawn from in-depth structural analyses (see following section), with key support for this arising from studies of *trans*-AT domains of PKSs that exhibit extreme promiscuity in terms of the substrates they can tolerate and the domains with which they can interact.[32,114]

2.2. Evolution of domain–domain interactions in PKS and NRPS enzymes

Evolution of domain–domain interactions in natural product megasynthases has received a lot of recent attention, leading to structural and functional insights that may assist in understanding the complex phylogenies exhibited by PKSs,[114,122–125] FASs,[126,127] and NRPSs.[96,128] These inter-domain interaction studies have also facilitated the development of new engineering tools[129] based on the results of mechanistic studies of PKSs activity.[130] In NRPS biosynthesis, MbtH is now well understood as a critical protein playing a promiscuous functional role in activating adenylation (A) domains despite the lack of catalytic activity.[131–133] The evolutionary origins of MbtH remain a mystery, yet the inter-protein docking domains (linker) of these enzymes have been identified in both NRPSs[134] and PKSs.[87] These linkers have evolved in a context of modularity, providing the correct structural organization of the enzyme. Given the recent development of evolutionary bioinformatic tools,[84,85,135] it is clear that these domains are open to investigation independently or as part of whole BGC phylogenies to reveal further insight into their origins and evolution.

Inter-domain evolution can take many forms. This has largely been embraced by evolution-driven synthetic biologists and protein engineers to develop pipelines for pathway design and bioengineering. Examples include rapamycin[136] and pikromycin[137] biosynthesis in PKS engineering, the use of a range of unnatural and natural peptides for NRPS engineering,[101,102] and for antimycins when engineering hybrid NRPS–PKS biosynthetic systems.[138] While these studies did not specifically aim to investigate the evolutionary principles underlying metabolic diversity in PKS or NRPS systems, they nevertheless provide useful evidence of the following: (i) PKS and NRPS pathways likely evolve *via* recombination of conserved domains that can be homogenized by concerted evolution; (ii) the range of pathways encoded by BGCs and their unique evolutionary histories that can inform synthetic biology approaches to enzyme and pathway engineering; and (iii) the metabolic diversity of modular pathways is pervasive, despite the prevailing view that evolution results in a 'main product' with a particular bioactivity. It is important to keep in mind how both a molecule and its target influence bioactivity. Some natural products require specific structural moieties in order to retain function against their target. In nature, evolutionary changes in the target itself can dynamically impact fitness. In antimicrobials, where selection is strong, evolution at the molecule–target interface can enter into dynamics wherein a molecule and target are reciprocally changing bioactivity landscapes.[64,139] Conversely, disruptive selection may play a role in diversification when new challenges, and thus new fitness optima, are introduced. This can take the form of changing nutrient/precursor availability and has been hypothesized as one explanation of extreme combinatorics within certain RiPP pathways.[39] With these clarifications in mind, observations (i) and (ii) above are directly linked by our ability to apply evolutionary insight to pathway engineering in biotechnology. Observation (iii) is an often-overlooked phenomenon as it relates to the production of "undesirable" by-products that may offer fundamental insight into the evolution of natural product biosynthetic enzymes and pathways. The latter is the essence of our DCME hypothesis (see Section 5.2.1), which builds on the screening hypothesis first proposed by Firn and Jones.[44]

3. Evolution of non-modular natural product biosynthetic enzymes

The proposed evolutionary connection between natural product biosynthetic enzymes and central metabolism is not new. However, not until recently has this idea been explored beyond the modular FAS–PKS system. Most enzyme families within secondary metabolism are non-modular (also known as non-templated). Despite its overwhelming diversity, given that metabolism, as a whole, is a closed system, the only way new pathways may arise is through recycling and repurposing already available proteins at the population level. This applies not only to secondary metabolism, but also any form of peripheral or supportive metabolism with origins in central metabolism. As introduced previously, metabolism evolves through a series of expansion-and-recruitment events, enabling the evolution of enzyme substrate specificities. Only rarely are new chemical conversions invented. For example, cytochrome P450s are promiscuous enzymes in terms of the substrates upon which they act, yet their chemical mechanisms based in free radicals are highly conserved. Single-domain, non-modular enzymes likely have different constraints and genetic mechanisms than modular, multidomain enzymes. This has been discussed in the previous sections for the PKSs, including aspects of its modularity. Though the PKS exhibit extreme chemical combinatorial potential, secondary metabolism is far more diverse. It is not our intention to exhaustively explore these possibilities here. Rather, we highlight a few examples that are instrumental in the biosynthesis, function, evolutionary trajectory of natural products in bacteria and focus on evolutionary principles that are exemplified by these selected biosynthetic enzyme families.

For example, type III PKSs (also called chalcone synthases) are a non-modular enzyme family, originally discovered in plants. Today, they have been described in organisms from across the tree of life, including bacteria and many non-plant eukaryotes. Type III PKSs direct the synthesis of both aromatic and pyrone small polyketides[37,140] and are extremely plastic, resulting in tremendous chemical diversity. In part, this diversity seems to be driven by the reactivity of its own acyl-CoA

substrates after elongation, which produces intrinsically reactive transition-state intermediaries allowing for many forms of cyclization.[36] It has been suggested that the evolutionary origin of this enzyme family was a fatty-acid binding protein with a chalcone-isomerase fold that lacked catalytic activity.[20,141] This evolutionary relationship was deciphered after a combination of phylogenetic analyses, directed evolution, and biochemical and structural biology.

Terpene synthases are responsible for the assembly of isoprene building blocks and represent an extremely diversified group of biosynthetic enzymes found throughout bacteria[142] and across the tree of life.[143–147] Cyclases that act within terpenoid biosynthetic systems are equally diverse, both in their substrate specificities and their taxonomic distribution.[148] The beta-lactam synthetase family is also widely taxonomically distributed and likely is evolutionarily related to the central metabolism enzyme asparagine synthase.[149–151] The beta-lactone secondary metabolites formed by ATP-dependent synthetases are another example of an enzyme family that can be found within the biosynthetic pathways of a diversity of scaffolds.[152] An ever-increasing diversity of BGCs use glycosyltransferases to link sugars together (*e.g.* aminoglycosides) or selectively install them onto chemical scaffolds.[153,154] Within the bacterial genus *Streptomyces*, the single-domain enzymes (including glycosyltransferases) involved in the biosynthesis of the aminoglycoside kanamycin are widely distributed and likely have ancient origins.[64] Further, many other single-domain enzymes involved in bacterial natural product biosynthesis (*e.g.* cytochrome P450s, methyltransferases, halogenases, radical SAMs, *etc.*) have varying levels of promiscuity, and may have evolved from central metabolism.

3.1. Evolution of single-domain enzymes in natural product biosynthesis

The appearance of novel natural product biosynthetic enzymes from functional paralogs of central metabolism has the potential to unlock previously undescribed 'chemical dark matter'.[155] All of these cases are closer to single-domain enzymes compared to modular PKS and NRPS megasynthases. Furthermore, single-domain enzymes may provide the evolutionary raw material for the assembly of megasynthases *via* gene fusions. Although there are many examples of recruitment events of homologs from central to secondary metabolism, as discussed in the previous sections, historical descriptions were mostly limited to anecdotal and/or serendipitous discoveries after laborious and time-consuming BGC biochemical characterizations. These cases were often merely suggestions rather than confirmed cases of evolutionary trajectories from rigorous evolutionary analyses. For example, it is reasonable to believe that the initial phosphonate biosynthetic enzymes, phosphoenolpyruvate (PEP) mutase or PepM and phosphonopyruvate (PnPy) decarboxylase or PpD, may have evolved from their central metabolic counterparts.[156,157] PepM converts PEP to PnPy and could be considered a homolog of phosphoglycerate mutase, responsible for conversion of 3-phosphoglycerate (3-PGA) to 2-phosphoglycerate (2-PGA) in glycolysis. PnPy then undergoes decarboxylation by PpD to form phosphonoacetaldehyde (PnAA) as a thermodynamically favored strategy.

These types of observations and the need for more rigorous (and automated) phylogenomic tools inspired the development of EvoMining (developed in author Barona-Gómez's lab), which identifies secondary metabolic enzymes evolving from central metabolism. EvoMining's core premise is that enzymes in BGCs have evolved from promiscuous enzymes operating elsewhere in metabolism *via* expansion-and-recruitment events.[135,158,159] Therefore, not only does this approach generate unbiased predictions of novel natural product biosynthetic pathways, but it provides general lessons about the evolution of biosynthetic enzymes as a whole. For instance, enzymes related to natural product biosynthesis may have evolved to serve extra-biosynthetic functions, as is the case in physiological adaptations[158,160] or immunity mechanisms such as antibiotic resistance. The former enzymes seem to provide metabolic adaptations related to precursor supply; whereas the latter appeared as resistance mechanisms, later during the evolutionary process leading to BGCs, in order to cope with the toxic activity of some natural products. This scenario is the foundation of a similar approach called the Antibiotic Resistance Target Seeker[161] (ARTS) which relies on resistance mechanisms being embedded within BGCs. To add even more complexity, the resistance can even emerge from noncoding regions.[162] The predictive power of new tools to find novel natural product pathways and decipher the evolutionary trajectories of resistance mechanisms remains an area of interest and is likely to yield interesting results in the future.

The EvoMining approach was successfully applied towards the unbiased discovery of a new pathway with unprecedented enzymatic reactions and chemical scaffolds in the model organisms *Streptomyces coelicolor* and *Streptomyces lividans*.[135] Even after the many decades of genetics and genomics in *S. coelicolor*,[163] evolutionary-inspired approaches are still able to find novel, validated pathways. In this pathway, a divergent AroA homolog, AroA2, allows for metabolic assimilation of arsenate. Remarkably, AroA2 has diverged from its primary metabolic counterpart AroA such that the thermodynamic balance of the reaction has inverted, opposite of how it typically proceeds during the synthesis of the aromatic amino acid intermediary 5-enol-pyruvyl-shikimate-3-phosphate (5-EPS-3P). This remarkable enzymatic innovation enables the incorporation of an arsenate into the metabolism of these *Streptomyces* species and facilitates the synthesis of a C–As covalent bond present in the intermediary arsenoacetaldehyde (AsAA), produced after the decarboxylase activity of a PnPd homolog encoded in the same BGC.[135]

The success of the previous example has been recently expanded upon through a comprehensive analysis of the expansion-and-recruitment profiles of 42 conserved enzyme families from central metabolism. In representative Archaea, Cyanobacteria, Actinobacteria, and *Pseudomonas* genomes, even when enzyme families were conserved in all the seed genomes used to retrieve their homologs, there were always cases of missing genes in one or more of the genomes used for assembly of the genomic databases. This speaks to the large metabolic

diversity that exists even in what could be considered integrated metabolic pathways, typically believed to be highly conserved and not variable. In some lineages, observed expansion events were quite low, whereas in other lineages the same enzyme could be highly expanded. Even when enzyme families were not expanded, they nevertheless could serve as starting points for the evolution of novel secondary metabolic enzymes. These examples were termed shell enzymes[159] and are usually encoded by the pangenome.

Scytonemin biosynthesis involves glutamate dehydrogenase (GDH) and acetolactate synthase (ALS) homologs encoded in central metabolism by the *gdh* and *ilvB* genes, respectively. The evolution of this pathway is an outstanding example of the dynamic behavior of biosynthetic genes. Recent phylogenetic reconstructions have shown that *gdh* and *ilvB* were recruited from central metabolism prior to the Great Oxygenation Event, approximately two billion years ago,[60,159] to construct the scytonemin BGC still seen today. The reactions catalyzed by their central metabolic ancestors are limited to L-glutamate and pyruvate, respectively. In contrast, the scytonemin biosynthetic genes have broadened their substrate specificities: the *gdh* homolog *scyB* accepts L-tryptophan and the *ilvB* homolog *scyA* accepts indole-3-pyruvate (I-3Py; the product of the reaction catalyzed by ScyB) and *p*-hydroxy-phenyl-pyruvic acid (*p*-HPP). *p*-HPP is produced from prephenate, *via* the action of TyrA involved in the synthesis of the aromatic amino acid L-tyrosine. Both L-tryptophan and *p*-HPP precursors are produced from chorismate, the final product of the common aromatic amino acid biosynthetic pathway that includes as key enzymes AroG and AroB (Fig. 1).

Enzyme expansions involving precursor supply are apparent in the scytonemin BGC from phylogenetic reconstructions of each of genes (independently and as a whole using concatenated protein sequences and gene neighborhoods).[60,159] Not all of the genes needed to provide L-tryptophan and *p*-HHP from glycolysis were found to be part of the scytonemin BGC. Instead, the genes recruited for this purpose, including *aroB*, *aroG*, *tryA*, *trpE/G*, *trpD*, *trpC*, *trpA* and *trpB*, likely have an important effect in metabolic flux to modulate the availability of key precursors, L-tryptophan and *p*-HPP (Fig. 1). This observation suggests that biochemical interactions between integrated and speculative metabolism can be under strong evolutionary pressures.[164] It is tempting to speculate that such recruitments occur more often in pathways that have been evolving for long periods. In this particular case, enzyme expansions not only arose from gene duplications, but also horizontal transfer events.[60] This highlights the complexity of evolutionary reconstruction and the challenges faced in deciphering the natural evolutionary histories of secondary metabolism.

3.1.1. Evolution of precursor supply enzymes and metabolic integration. The supply of precursors is critical in secondary metabolism, providing the building blocks for biosynthesis. The physiological and environmental cues for secondary metabolite production often are related to nutrient-limitation where regulatory decisions within the cell may result in metabolic conflict, balancing a declining growth rate and reduced nutrient availability with a need to produce secondary metabolites[164] and initiate morphological development and/or potentially chemical defense. Genomic analysis of so called 'talented' bacterial strains (*i.e.* those who produce many secondary metabolites) reveals gene expansion events outside of BGCs likely to provide organisms with alternative routes to make metabolic intermediates and cope with declining precursor concentrations in the face of potentially co-regulated primary metabolic genes.[165] A recent survey of the primary metabolic enzymes across Actinobacteria[160] confirmed the extensive expansion of gene families associated with primary metabolism (12 enzymes [48 genes] from glycolysis, TCA cycle, and amino acid metabolism). However, most expansions arose *via* HGT with only two enzymes (phosphofructokinase and pyruvate kinase) originating in duplication events. Detailed further study is warranted to elucidate why an organism might expand such genes. Nonetheless, some of these expansion events provide evolutionary robustness to strains to enable survival in periods of nutrient stress and may facilitate secondary metabolite production.[160] Additionally, some genes that might superficially appear as functioning only in primary metabolism actually directly contribute to the production of secondary metabolites.[135] and precursor biosynthetic genes may provide shared material for the enzymes of multiple BGCs.[166–168]

There remains a degree of trial and error in the engineering of precursor supply, even in the presence of high-quality genome-scale metabolic-network models (GSMs). Nevertheless, GSMs of some Actinobacteria have been used to explain evolutionary trajectories of both supportive and integrated metabolism. From these models, it was discovered that *Streptomyces* have two redundant copies of most glycolytic enzymes,[158] and several other metabolic features in early diverging Actinobacteria.[169] The observation of glycolytic redundancy was confirmed with subsequent experimental characterization of the pyruvate kinases (*pyk* genes, encoding Pyk1 and Pyk2) present in *S. coelicolor*. These paralogs have similar biochemical function, yet distinctly modulate secondary metabolite production and fitness.[160] This work showed that relatively high sequence similarity can still manifest in vastly different kinetic parameters, cofactor specificity and allosteric regulation, even when both paralogs are constitutively expressed. Understanding the specific mechanisms of what controls two complex multigenic traits that integrate multiple phenotypic behaviors (*i.e.* natural product biosynthesis, growth and sporulation) remains to be fully elucidated. Nevertheless, this highlights a significant problem with GSMs in genomes with large numbers of genetic expansions: it is difficult to tease out the nuances of each isoenzyme from modelling alone[66] and there is still a need to undertake detailed kinetic studies of enzyme function. When considering precursor supply, neither GSMs nor evolution-inspired engineering is sufficient to successfully implement rational synthetic biology approaches for natural product biosynthesis. Even in relatively simple systems (*e.g.* the overexpression of short amino acid biosynthetic pathways to increase precursor supply), outcomes are often unpredictable.[170] An example is the *proABC* genes needed for the synthesis of the prodiginines precursor proline. *proC*

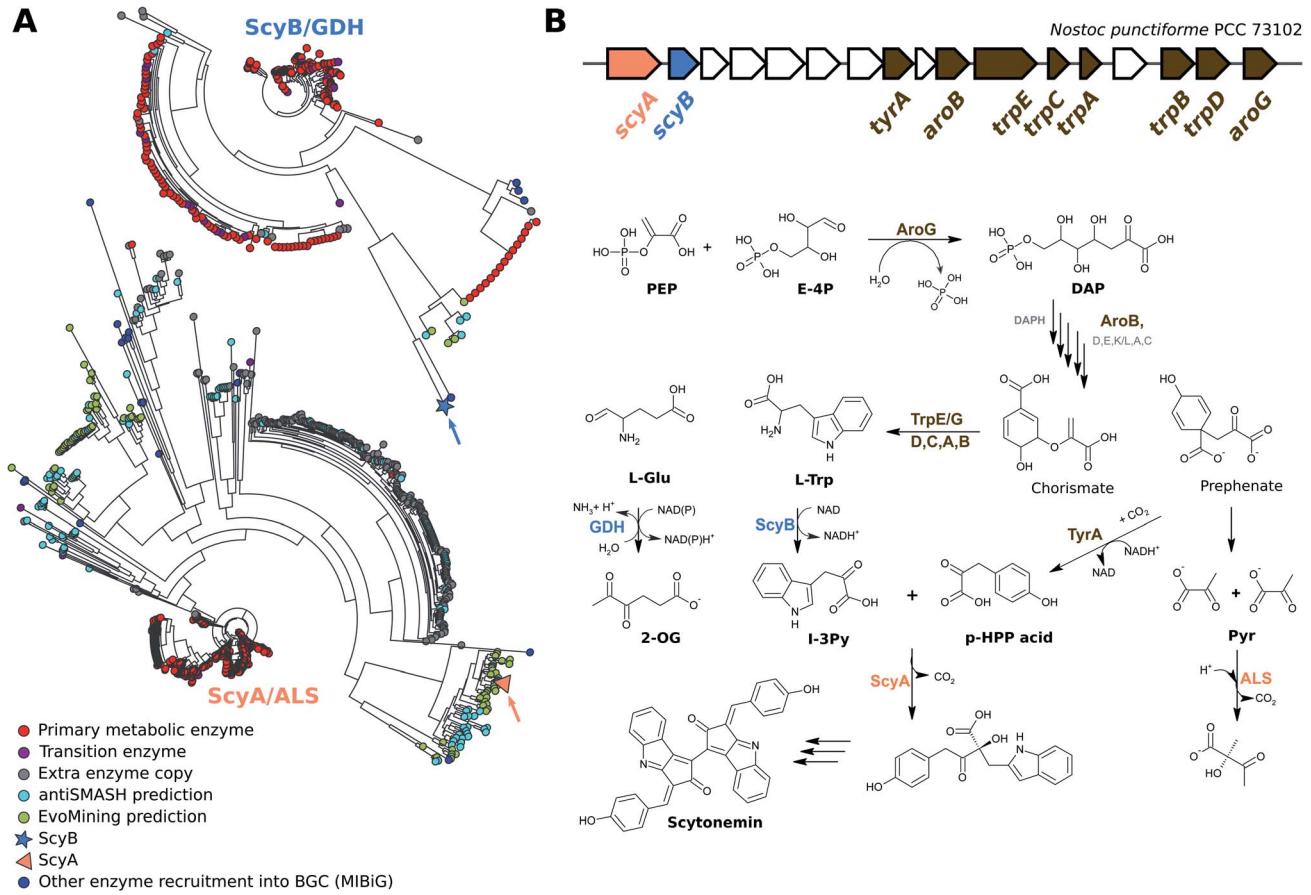

Fig. 1 Evolution of key enzymes of the scytonemin pathway. (A) EvoMining analysis. EvoMining trees for GDH/ScyB and ALS/ScyA enzymes is shown, revealing a natural history mediated by expansion-and-recruitment events, from central or integrated metabolism into specialized or speculative biosynthesis. Color code for the different evolutionary key enzyme types is provided within the figure. (B) Biosynthetic pathway. Central metabolic enzymes are shown in grey. Enzymes shown to have undergone expansion-and-recruitment events are depicted with the same color: brown, AroG and AroB for enzymes recruited to provide chorismate (Ch); TyrA for synthesis of prephenate (Pph); and TrpE/G,D,C,A,B for synthesis of L-tryptophan. Blue, glutamate dehydrogenase (GDH) and ScyB, which generates 2-oxoglutarate (2-OG) and indole-3-pyruvate (I-3Py) from L-glutamate (L-Glu) and L-tryptophan (L-Trp), respectively. Orange, acetolactate synthase (ALS) and ScyA, which generates the scytonemin dipeptide precursor from I-3Py and *p*-hydroxy-phenyl-pyruvic acid (*p*-HPP). In contrast, ALS uses two molecules of pyruvate (Pyr) to produce acetolactate.

elicits production of prodiginines while its deletion does not yield a proline auxotroph.[170] This is further complicated in that the product of *proC*, pyrroline-5-carboxylate reductase, can be compensated by naturally promiscuous IlvC isomeroreductases, allowing cross-talk between proline and branched-chain amino acid biosynthesis.[27]

Given that many BGCs include genes devoted to the synthesis of pathway intermediates, precursors for the synthesis of their cognate natural product are often overlooked when attempting to engineering designer products. As mentioned, there is a reasonable expectation that these enzymes should be metabolically integrated with the biosynthetic enzymes that direct the synthesis of the natural product. Yet, it is often the case that the additional genes involved in precursor-supply are not sufficient to reconstitute a complete metabolic pathway coexpressed with the main BGC. Similarly to the scytonemins BGC (introduced in previous section), which partially includes precursor supply genes, analysis and modeling of the Calcium-Dependent Antibiotic (CDA) BGC in *S. coelicolor* has been problematic.[171] The approaches we have discussed here could be exploited to develop an evolution-driven exploration of the genes that are specifically relevant for cross-talk between central and secondary metabolism. One could predict that an integrated sub-network of metabolism may become apparent in a partitioned primary and 'secondary' metabolism. Building on GSM approaches for modeling metabolic flux with an emphasis only on the enzymes (and/or metabolic nodes) that may have been recruited during evolution to increase flux of key precursor metabolites. These approaches are not possible with current GSMs, but with increased integration of omics technologies and cell biology, such approaches may become possible soon.

3.2. Evolution of enzymes related to resistance and immunity mechanisms

Since the days of Waksman, we have relied on bioactivity to guide us to novel natural products. Even in light of

pharmacologic-, target-, and chemistry-based diversification of natural products, discovery efforts have seen diminishing returns.[172] Genome-mining approaches to natural product discovery have identified potential drug leads, however these approaches rely on 'knowing what a biosynthetic cluster looks like' and often lead us to rediscover biosynthesis that has been previously described. For BGCs that produce a toxic compound, an immunity mechanism must also either evolve along-side the BGC or arrive late in the evolutionary assembly. This concept was further developed by Wright and coworkers[173] as a means of identifying novel glycopeptide-producing strains by screening for vancomycin resistance. Comprehensive phylogenetic reconstructions are beginning to shed light on the relationships between the gene content of specific BGCs and the evolution of resistance, exemplified by a recent reconstruction from Waglechner *et al.* describing the gene gain and loss dynamics across the glycopeptide evolutionary history.[21] In some cases these immunity mechanisms are encoded within the BGC and, as many of these clusters are transferred horizontally between strains, this leads to the acquisition of their cognate resistance mechanisms.[59,174,175] Using a manual bioinformatic approach to search for novel BGCs, Tang *et al.*[176] screened the genomes of natural product producing strains for duplicated genes with 'housekeeping' functions. In this work Tang *et al.* identified an ortholog of a fatty acid synthase that was in proximity to a BGC, suggesting it would be likely a duplication of target resistance mechanism. Resistance by this mechanism has been shown previously for platencin, platensimycin, novobiocin and griselimycin.[177–179] Remarkably, duplicate genes from central metabolism likely indicate the target for a novel metabolite and highlight the importance of adaptive exploration of sequence space to evolve antibiotic resistance. A good example is rifampicin resistance, where RNA polymerase mutations evolve rapidly in response to exposure.[180] In producing organisms, however, duplications may be the favored route to resistance.[177–179,181]

Approaches to identify novel BGCs that exploit the evolutionary pressure to possess an associated immunity mechanism highlight the need for HGT events to transfer both the BGC itself and the immunity mechanism. Perhaps the evolution of resistance could represent acquisitions where resistance evolves in response to the production of a functional molecule, such as duplication of a house-keeping gene to provide resistance. Initially, such a duplication may not be located in the BGC, as in elfamycin/kirromycin and duplicated EF-Tu.[181] Similarly, rifampicin resistance is conferred by the duplication of the RNA polymerase beta-subunit RpoB in the pathogen *Nocardia farcinica*[182] and glycopeptide producer *Nonmuraea* (formerly *Actinomadura*) A40926.[180] How resistance mechanisms subsequently co-localize may in part be due to minimizing stochastic growth effects from unequal expression of BGC genes if the resistance mechanism is not tightly co-regulated. Such effects are seen in metabolic genes where co-localization into the same pathway minimizes stochastic metabolic stalling.[183]

4. Evolutionary dynamics at different scales: from domains to BGCs

Knowing the evolutionary trajectories of domains, modules, sub-clusters, and BGCs is a fundamental prerequisite for a better understanding of the evolution of natural product biosynthesis. This was early on acknowledged,[72,73] but it had not been explicitly investigated until more recent studies.[79,135,159,184–186] Deciphering these evolutionary dynamics of BGCs will be important to inform synthetic biology approaches, but when approaching pathway engineering it will be important to appreciate that each BGC and its component enzymes have their own evolutionary histories, even if they belong to a common biosynthetic or chemical family.[79] There has been success in swapping modules and domains within biosynthetic pathways, yet these efforts were largely completed after identification of loci amenable to genetic engineering. Notably, loci that have successfully been engineered have experienced their own evolutionary trajectories, and successful engineering examples designed *a priori* have been informed by evolutionary patterns (*i.e.* using sequence analysis) and are discussed throughout this work (see Section 2.1). However, while the critical role of domain–domain interactions during evolution of natural product biosynthetic enzymes is not a new concept,[187] it was not until recently that the increased clarity of their evolutionary dynamics has resulted in successful biosynthetic engineering (see Section 2.2). To our knowledge, however, there are limited examples of pathway engineering of the non-modular, or single domain biosynthetic enzymes, however recent advances in synthetic biology methodology are now moving these enzymes in to the realms of engineering possibilities (see Section 3). There are two important evolutionary observations, we believe which may help to explain the slow pace at which PKS and NRPS pathway engineering has proceeded and both of which relate to the relatively low genetic diversity within their biosynthetic systems.

First, researchers do not always investigate homologous biosynthetic systems once a BGC directing the synthesis of a bioactive metabolite has been discovered. This is an important shortcoming as it is only through the study of closely related and divergent BGCs that we will be able to highlight evolutionary trajectories.[77,188–190] This, however is beginning to change, with biotechnology companies, interested in natural products and novel enzymes building large (meta)genomic databases to explore chemical novelty and congeners. Whilst, these are not generally available for evolutionary analysis by academia, there is a need to develop open efforts, which until now have been limited to relatively few closely related strains.[191] Secondly, with the role of concerted evolution increasingly appreciated as a significant driver of the evolution of secondary metabolism, the investigation of closely related systems can be highly informative of recombination-based approaches. This requires the *a priori* identification of domains and module boundaries with sufficient diversity to develop insight and belonging to sub-clusters and BGCs. As discussed previously, it is these sub-clusters that can conjugate to generate chemical

diversity and assemble gene collectives. Sub-clusters from different genomic loci can act together to assemble their products.[192,193] On the other extreme, sub-clusters from the same genomic loci can form discrete products, both through independent biosynthesis (*e.g.* zwittermicin and kanosamine[194]) or from shared biosynthetic machinery with promiscuous protein–protein docking (*e.g.* the vatiamides[195]). Further, certain glycosyl transferases have been shown to serve as activators of other enzymes belonging to sub-clusters[153,154,196] and tailoring enzymes have been observed to act in multiple sub-clusters within the same genome.[197] Through harnessing the chemical diversity and therefore the array of biological syntheses that can be achieved through diverse compositions of sub-clusters and BGCs[73] we can gain insight in to how these BGCs have evolved. Recent phylogenomic analyses support this approach[46,59,85,113,135,198,199] where the natural variations in gene presence or absence have been used to construct evolutionary histories. Although detailed examples are relatively few in number, our ability to reconstruct these evolutionary histories is becoming a reality through whole-gene cluster phylogenies.[21,85] Currently, these approaches provide a means to prioritize BGCs for exploitation and below we review some examples that emphasize this kind of approach.

4.1. Phylogenomics of BGCs: revealing key biosynthetic genes driving chemical diversity

In recent years, phylogenomic approaches have been adopted by some groups in the natural products community. Although the field has widely embraced genomics, even to the point of spawning the sub-discipline of genome-mining, phylogenetic reconstructions have come about at a slower pace. Perhaps this is due to the reliance of traditional phylogenetics on common and conserved gene sequences under neutral evolution to derive natural histories (*e.g.* in Actinobacteria[200]). In contrast, natural product biosynthetic genes are extremely heterogeneous,[201] and thus population genomics and phylogenomic methods pose very specific challenges not traditionally faced by molecular evolutionary biologists. In recent years, these hurdles have been overcome with relative success and *ad hoc* evolutionary bioinformatics tools have become available.[84] A particularly useful tool that has emerged to perform phylogenomic analyses of BGCs is the CORe Analysis of Syntenic Orthologs to prioritize Natural products biosynthetic gene clusters (CORASON) algorithm, developed in author Barona-Gómez's lab.[85] CORASON extracts a shared protein core from sub-clusters and BGCs to be used for construction of concatenated sequence phylogenies. CORASON further provides a visualization of the gene neighborhood in which these proteins are encoded, facilitating analysis of the evolutionary dynamics of biosynthetic genes. The displayed figures use a color-code for each gene that is at least partially conserved throughout the tree (as revealed by the frequency bars) and uses a gradient of color intensity to denote the sequence similarity to the query sequence (*e.g.* dark: high similarity).

Examples of evolution-driven genome mining discoveries are discussed elsewhere, both in this review and beyond. Selected cases are revisited in the following three sub-sections, each accompanied by up to date CORASON phylogenomic reconstructions to provide complementary analyses to those reported in the primary literature. These reconstructions provide non-obvious insights that are subsequently discussed. First, the case of rapamycin and gene cluster family members is used to highlight the main genetic mechanisms driving chemical diversity, namely, genetic rearrangements of modular enzymes and acquisition of new genes by BGCs; second, the evolution of clavulanic acid chemical diversity, involving multiple paralogous sub-clusters evolving independently or during co-evolution of this BGC with that of cephamycin C, emphasizes the neglected role of multiple loci involved in speculative metabolism; and finally, the rich chemical diversity produced by highly conserved and integrated pathways, using as an example the biosynthesis of hydroxamate siderophores (*e.g.* desferrioxamines), helps to realize how chemical diversity goes beyond the notion of one BGC corresponding to one main final-product. In addition, the desferrioxamine example describes how new modular enzymes may appear during evolution and how these novel enzymes may diverge to produce related molecules in distantly related organisms (*e.g.* putrebactin). All together, these examples give rise to a new view on the evolution of secondary metabolism. These paradigms are incorporated in a unifying theory, the DCME hypothesis, presented in Section 5.

4.1.1. Main genetic mechanisms driving chemical innovation. As anticipated by Fischbach and co-workers a decade ago (and supported by evidence presented and discussed herein), it is safe to state that the two main genetic mechanisms driving the evolution of natural product chemical diversity are (i) gene expansion-and-recruitment events; and (ii) genetic insertions and deletions within recombination-prone PKS and NRPS modular enzymes. Recently, the evolutionary history of the hybrid peptides detoxins and rimosamides was deciphered and exploited to unveil novel chemical diversity.[202] These molecules act as "*anti*-antibiotics" and have a common origin yet they are produced by independent BGCs (different only by one NRPS module). Through CORASON phylogenomic analyses, clade-specific BGCs were identified that direct the synthesis of these natural products. As a proof-of-concept, novel chemical diversity surrounding the rimosamides and detoxins was experimentally confirmed after screening of a large strain collection of Actinobacteria known to produce these hybrid metabolites and/or encode the newly described key biosynthetic genes.[85] A homolog of TauD was identified by EvoMining as the key enzyme allowing for accurate phylogenetic reconstructions, presumably evolved from the *tauABCD* operon form central metabolism involved in assimilation of sulphite by oxygenolytic release from the amino acid taurine. New variants identified following this evolution-driven approach could be placed within three different sub-clades of the CORASON tree and had pathway-conserved genes coding for P450 enzymes and an enoyl-CoA hydratase/isomerase, yielding the fatty acid amide detoxin S1 and hydroxylations in detoxins P1–P3, respectively. The spectinomycin BGC was found to be located immediately downstream of the predicted detoxin BGC, suggesting

a possible link between the encoded pathways at this locus in corresponding *Streptomyces* species.

As for genetic rearrangements, take for example the Rapamycin/FK class of natural products which includes rapamycin, FK506, FK520, the nocardiopsins, meridamycin, and others. Both rapamycin and FK506 have been used extensively as immunosuppressive drugs that have enabled countless transplant surgeries. Within the cell, conserved chemical moieties within these molecules bind the prolyl isomerase FKBP12 and then, in complex with FKBP12, bind their ultimate cellular target through a variable chemical region (in mammalian cells, rapamycin–FKBP12 targets mTOR and FK506–FKBP12 targets calcineurin).[203,204] Within the BGCs of this class of molecules, the key biosynthetic components that give rise to the conserved, FKBP-binding region of the molecules are a single-module NRPS that installs pipecolic acid,[205] a lysine cyclodeaminase that makes the pipecolic acid,[206] and a chorismatase responsible for the biosynthesis of the aromatic PKS starter unit.[207] Within the PKSs genes, there is evidence of both inter- and intra-domain recombination (Fig. 2), congruent with recent efforts to accelerate the chemical diversification of the rapamycin pathway through engineering approaches.[136] In nature, rapamycin/FK PKS assembly line diversification appears within the aforementioned 'variable' region, as the FKBP-binding 'conserved' region is evolutionarily constrained. Likely, these changes are tied closely to both function and ecology, which may be quite different given the taxonomic distribution of family members shown in Fig. 2. Type I PKS loci with distinct selective pressures and type I PKS recombination within the same pathways have also been described elsewhere[69,208] and may occur naturally in response to dynamic fitness landscapes and/or chemical and functional constraints. The rapamycin/FK family also exhibits evidence of shuffling the content of single domain enzymes, including cytochrome P450s and methyltransferases.

4.1.2. Increased chemical diversity by interactions of multiple biosynthetic genes. Given that genome mining requires only a draft genome to begin analyses, placing BGCs into broader genetic and metabolic context has become possible at large scales. Comparative analyses have begun to show that biosynthetic genomic elements can co-occur and interact at multiple scales, from single genes to sub-clusters, to complete BGCs. For instance, two independent tRNA-utilizing cyclodipeptide synthases, NozA and NcdA, were identified and characterized in alkaloid diketopiperazine nocardioazine biosynthesis in a marine *Nocardiopsis* species.[117,209] Despite the

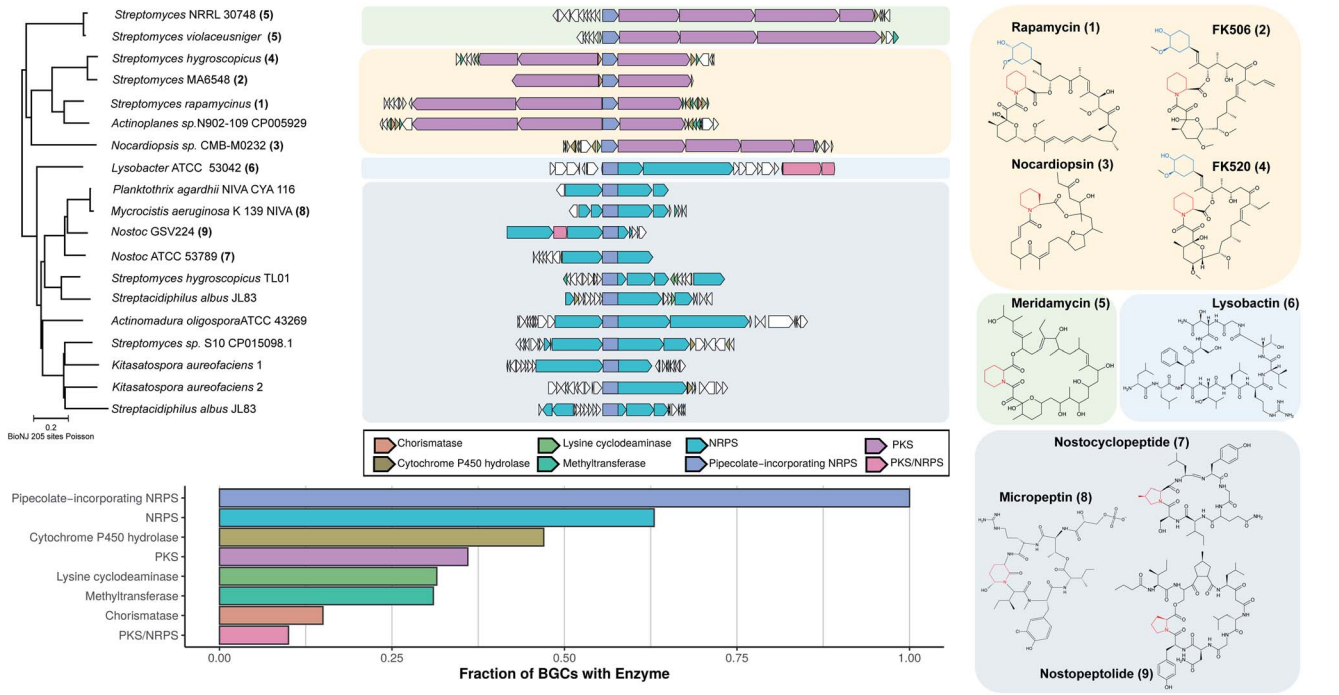

Fig. 2 Evolutionary dynamics of the Rapamycin/FK BGCs. Rapamycin phylogenomics using CORASON identified the pipecolate-incorporating NRPS gene (red) as the key and sole conserved enzyme throughout related BGCs. Gene frequency bars, an output of CORASON, is shown in the bottom left-hand side of the panel. For the sake of clarity, default color gradient provided by CORASON as an output, was discarded. Four main groups of BGCs could be identified: two groups located in the upper main branch of the tree (meridamycin and FK/rapamycin/nocardiopsin) are characterized by the presence of PKSs with different lengths (yellow) and a P450 hydroxylase (dark grey). In addition, the rapamycin/FK class includes genes for a lysine cyclodeaminase (orange) and chorismatase (blue), which are dedicated to synthesizing the conserved FKBP-binding region of these metabolites. At the lower main branch, BGCs corresponding to the other two discrete groups (lysobactin and cyanopeptin/micropeptin/nostocyclopeptide) are characterized by having NRPS/PKS hybrid (gray) and NRPS (light blue) genes, respectively. Unlike the upper branch, these BGCs do not contain a chorismatase enzyme gene, which is reflected by the absence of aromatic moieties from their structures (colored blue). Rearrangements including methyltransferases (green), P450 hydroxylases, lysine cyclodeaminases, NRPS and PKS domains around the pipecolate incorporating NRPS are visualized along the evolutionary natural history unveiled by the tree.

genes coding for these enzymes being physically separated in different genomic loci, they exhibited high tryptophanyl-tRNA activity and were involved in the synthesis of nocardioazines. The selective pressures on these different gene copies and the effects of their enzyme products on chemical diversity are currently unresolved. However, observations such as these are likely common and may be supported once research efforts go beyond the 'one BGC produces one main product' bias. Fortunately, reports beyond this assumption are starting to appear, adding to anecdotal cases where scientists end up characterizing 'redundant' systems by chance. For characterization to occur more comprehensively and efficiently, BGC databases, (*e.g.* MIBiG[210]) must be linked to high-quality genome sequences (similar to the large, yet private, sequence collections of industry), such that academic efforts can leverage genomic information beyond the BGC to describe trends in secondary metabolism.

Two recent cases in the well-studied species *Xenorhabdus* and *Pseudomonas* support the need for secondary metabolism research to extend beyond a BGC-centric view. The phenazine biosynthetic machinery of the entomopathogenic nematode symbiont *Xenorhabdus szentirmaii* is housed in one BGC, yet it can be used by another, separate BGC resulting in diversification and novel activities of the resulting natural products,[61] both potentially under specific evolutionary constraints due to the bacterium's unique ecological niche. Phenazine biosynthesis is broadly distributed throughout many bacterial taxa and is found in BGCs that code for diverse chemical structures and bioactivities.[211] In *X. szentirmaii*, phenazine biosynthesis appears partially duplicated and the resulting copies consisted of different versions of the same BGC with different gene content. The relaxed selection likely generated by a full subcluster duplication allowed for the appearance of a novel phenazine with a better antibiotic activity and perhaps redundant genes were lost over time. A second case involves two unrelated BGCs that nevertheless interact at the metabolic and regulatory levels. The phloroglucinol pathway, leading to 2,4-diacetylphloroglucinol (DAPG, see following section) in the rhizosphere bacterium *Pseudomonas protegens*, accumulates an aromatic polyketide intermediary that can be fluorinated by an enzyme encoded in the pyoluteorin BGC, increasing the chemical diversity surrounding DAPG from a co-regulated biosynthetic pathway.[212] Moreover, this mechanism provides *P. putida* with a biological mechanism to mitigate plant pathogens, such as *Erwinia*. Subsequent studies indicate *P. protegens* is prone global regulatory mutations that completely switch off secondary metabolism when pyoluteorin biosynthesis is expressed.[213] Presence of pyoluteorin itself was not sufficient to explain this phenomenon.[213] The authors of this report discussed the metabolic burden associated with expression of this pathway as a plausible explanation for these results, but the potential chemical diversity beyond its 'main' final-product was not considered.

Despite obvious relevance for clinical antibiotics and the fact that multiple paralogy was realized relatively early,[214] the clavulanic acid and cephamycin C have been poorly investigated as an example of multiple and interacting BGCs. These molecules can be produced by the same *Streptomyces* strains, or independently.[215] Cephamycin C is a clavam antibiotic for which resistance has been broadly documented, mainly through the activity metallo-β-lactamases capable of opening the clavam ring.[216] In nature, certain *Streptomyces* have evolved an antidote to inhibit this resistance mechanism in the form of clavulanic acid and related congeners.[217] *Streptomyces clavuligerus* strain ATCC 27064 is the main producer of these two interacting and clavam-containing bioactive metabolites. Based in the clavam-related gene repertoire of this organism, which includes paralogous subclusters present in its genome (chromosome and plasmid), we conducted a phylogenomic analysis using CORASON (Fig. 3A). Given that the clavam ring can be produced by at least four different convergent biosynthetic pathways,[218] this analysis provides only a partial evolutionary reconstruction of clavam-related metabolites. Still, it provides enough insights to make the case for multiple BGCs increasing chemical diversity as a genetic mechanism to "speculate" over chemical space during the evolution of natural products. Genome sequences of the clavulanic acid producing *Streptomyces jumonjinensis* and *Streptomyces katsurahamanus* were recently obtained and are included here.[219]

Based on our CORASON analysis, only *S. clavuligerus* ATCC 27064, *S. jumonjinensis*, and *S. katsurahamanus* harbor most of the genes involved in the synthesis of cephamycin C upstream of the BGC of clavulanic acid (blue genes). Together, these genes are approximately 60 kilobases (Kb), as previously reported.[215] Two discrete clades emerge. The first consists of seven organisms: *S. fulvoviridis* JNXH01, *S. olivaceus* JNWM01, *S. flavovirens* JOAB01, *Streptomyces* sp. NRRL B-24051, *Streptomyces* sp. NRRL S-325 JOIW01, *S. pratensis* ATCC 33331 and *Streptomyces* sp. PAMC26508. The second is formed by only two organisms: *S. viridis* JRZE01 and *S. viridis* DSM 43017. These clades exhibit different genomic organization of the clavulanic acid BGC (Fig. 3A; black asterisks), containing all the early genes (purple genes) necessary for the synthesis of clavaminic acid, shown central and highlighted in purple in Fig. 3B. Moreover, the evolutionary dynamics revealed by our CORASON analysis include additional genes involved in the late steps for the synthesis of clavulanic acid (black genes), as previously described,[215,217,220] suggesting the possible biosynthesis of novel 5*S* clavams compounds, such as *N*-acetyl-glycyl-clavaminic acid and *N*-acetyl-clavaminic acid (Fig. 3B; black box) identified in the *dcl8* mutant of *S. clavuligerus*.[221] Interestingly, the *ccaR* gene was found within these BGCs. CcaR is an essential transcriptional activator for the biosynthesis of cephamycin C, clavulanic acid, and 5*S* clavams.[222]

All the *Streptomyces* strains analyzed possess the biosynthetic machinery leading to biosynthesis of clavaminic acid, the most advanced intermediate common to both the clavulanic acid and the 5*S* clavam biosynthetic pathways,[223] from which valclavam and 2-hydroxyethylclavam in *S. antibioticus* Tü1718 or 2-carboxymethylideneclavam, 2-formyloxymethyclavam, 2-hydroxymethylclavam, clavam-2-carboxylic acid, 8-hydroxyalanylclavam, and alanylclavam in *S. clavuligerus* ATCC 27064, are produced.[215,217] According to our phylogenomic analysis, the orthologous genes *akr* (*cvm1*) and *lig*, which belong

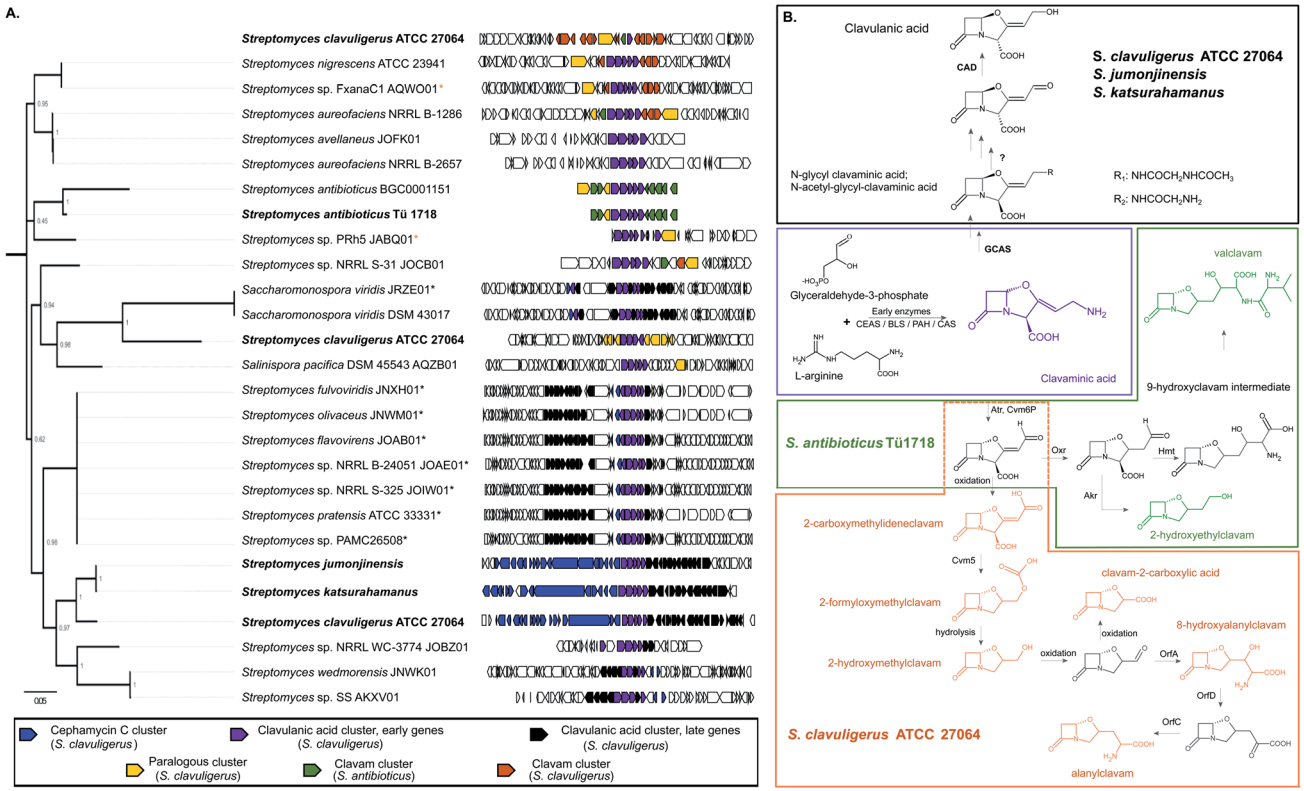

Fig. 3 Evolutionary dynamics and biosynthetic diversity of clavam BGCs. (A) Phylogenomics. The tree was reconstructed with three conserved proteins (CEAS, carboxyethylarginine; BLS, beta-lactam synthetase; and PAH, proclavaminic acid amidinohydrolase), as revealed by CORASON. For the sake of clarity, default color gradient provided by CORASON as an output, was discarded. These conserved genes represent early enzymes involved in the synthesis of clavaminic acid (purple genes). Orthologous genes were colored according to the reference BGCs of *S. clavuligerus* ATCC 27064 and *S. antibioticus* Tü 1718 (marked in bold). Two novel clavulanic acid and Cephamycin C producers, *S. jumonjinensis* and *S. katsurahamanus* were included. White genes include the forty genes upstream and downstream of all clavam compounds BGCs that were analyzed and found to be unrelated to these pathways, due to lack of universal conservation. Possible novel clavams-producers could follow specific biosynthetic pathways to produce clavam metabolites (marked with an asterisk). (B) Biosynthetic pathways. A purple box highlights early enzymes involved in the biosynthesis of clavaminic acid as the first step towards biosynthesis of clavam metabolites. Three proposed biosynthetic pathways leading to 5*R* clavulanic acid in *S. clavuligerus* ATCC 27064 (black box) and 5*S* clavams in *S. clavuligerus* ATCC 27064 and *S. antibioticus* Tü 1718 are shown (orange-red and green boxes, respectively). Some enzymes belonging to the clavam cluster, as well as the paralogous plasmid-borne cluster from *S. clavuligerus*, are involved in the biosynthesis of 5*S* clavam compounds. Selected cofactors and co-substrates are shown. Putative intermediates metabolites are in grey, and names and abbreviations are provided within the panel.

to the clavam cluster of *S. antibioticus* Tü1718, are present in *S. aureofaciens* NRRL B-1286 JNWR01 and *Streptomyces* sp. NRRL S-31 JOCB01, respectively. The involvement of these genes in the biosynthesis of 5*S* clavams is not yet clear, however, mutation of *cvm1* abolishes production of clavam-2-carboxylate, 2-hydroxymethylclavam, and alanylclavam compounds.[224] Besides these strains, *S. nigrescens* ATCC 23941, *Streptomyces* sp. FxanaC1, and *Streptomyces* sp. PRh5 contain the *cvm7p* ortholog gene that encodes a two-domain transcriptional regulator (orange asterisk), which specifically regulates the last steps during biosynthesis of 5*S* clavam compounds.[225,226] In some organisms, the presence of *orfA* orthologs were also found: this gene has been associated with the biosynthesis of the alanylclavam compound.[227] Together, these genomic insights suggest that other Actinobacteria species could synthesize similar compounds to those previously reported or proposed (compounds in gray). More importantly, evolutionary patterns together with functional insights suggest that chemical diversity is the result of a complex and dynamic interplay between several subclusters and/or BGCs that can be functionally linked, as is observed in *S. clavuligerus*.

4.1.3. High metabolic diversity synthesized by conserved BGCs. The notion of one pathway generating one main product is prominent in highly conserved BGCs. At first glance, this may seem reasonable, yet experimental results are often paradoxical, as will be discussed. Indeed, real observations tell a completely different story, which nevertheless is typically overlooked. The demand for a 'clean' chromatogram seems to present an irresistible temptation for the natural product research community. There is likely not a single case where the one pathway to one product paradigm rings true. Take for example the highly conserved BGC directing 'mainly' the synthesis of 2,4-diacetylphloroglucinol (DAPG) in *Pseudomonas* species, including *P. fluorescens* and other plant-associated *Pseudomonas* species.[228] DAPG has been found to exert an array of biomolecular activities,[199,228,229] some that can be rationalized as relevant in the

biological and ecological contexts in which DAPG-producing organisms live.[230,231] It is often overlooked that this relatively simple pathway consisting of five enzymes also produces a series of congeners,[232] some of whom exhibit strong biomolecular activities.[199] This may not come as a surprise given that the phloroglucinol pathway includes a type III PKS. Nevertheless, other enzymes are required to produce DAPG and congeners. Examples like this abound in the literature, from metabolites with unknown biomolecular activities to broadly investigated natural products such as the siderophores, which exert their known activities thanks to their physicochemical features that allow them to chelate iron and other metal ions.

Desferrioxamines (dFOs) are a broadly occurring class of hydroxamate natural products that act as *bona fide* siderophores in response to iron limitation (*i.e.* they are iron-chelating metabolites that facilitate assimilation of insoluble, ferric iron).[45] dFOs have begun to catch the attention of microbial ecologists, as they may provide a mechanism through which bacteria evolve and interact within communities.[67,113] Structurally, they consist of hydroxamate monomers derived from lysine or ornithine. Given their demonstrated biological functions, they are often referred to as specialized metabolites, even when they are a clear example of supportive metabolism having a central role in the growth and development of the producing organisms. Beyond this, there exists an ever-growing diversity in dFO chemical variants despite being synthesized by a relatively simple BGC consisting of four conserved, non-modular, single-domain biosynthetic genes: *desABCD*.[233,234] A fifth gene, *desG*, has been recently identified. *desG* can either be part of the *des* BGC or adopt a different position in the genome. It has been suggested to be a tailoring enzyme (annotated as a penicillin amidase) responsible for incorporating an aromatic ring into the hydroxamate polymer, giving rise to aryl-capped dFOs.[113] DesG activity might be an ecologically relevant modification that affects the solubility of dFOs and siderophore metal scavenging.[235] The dFOs have been intensively investigated mainly in model bacterial strains. Genomic studies in these model organisms have helped to decipher the biosynthetic, regulatory, and transport mechanisms that sustain siderophore-mediated iron-acquisition,[113,233,234,236–240] allowing for integration of evolutionary analyses.[235]

Circumstantial observations suggest that dFOs first appeared in *Streptomyces* species and then were horizontally transferred to related Actinobacteria[113,241–243] and other bacterial phyla, such as the Gammaproteobacteria[244] and Alphaproteobacteria,[245] where occurrence of this BGC is relatively limited. The natural history and evolutionary dynamics of key *des* genes can be resolved by the CORASON tree shown in Fig. 4A. Together with a previously published species topology[200] these results show highly conserved BGCs without protein fusions in the *Streptomyces*, as well as putative inter- and intra-phylum HGT events: between the Actinobacteria and Proteobacteria, as well as horizontal acquisition of the BGC by certain organisms, such as *Corynebacterium* species. These analyses also revealed the occurrence of protein fusions involving different Des enzymes, *e.g.* DesC–D, in the Gammaproteobacteria *Erwinia* and *Pantoea*, plus Actinobacteria from the genus *Clavibacter* and several other paraphyletic Micrococcales species;[246] as well as DesB–C fusions in bacteria from the genus *Kocuria* and *Microbacterium*. A DesB–C–D fusion is found in *Cryobacterium roopkundense*, a rare Actinobacterium from the family Microbacteriaceae, situated in the tree close to *Arthrobacter* species. These evolutionary observations are relevant as precursor feeding and synthetic biology approaches have evidenced the promiscuous nature of the Des biosynthetic enzymes,[247–249] leading to congeners of the acyl-capped dFO-B, used clinically to treat patients with secondary iron overload disease or acute iron poisoning. Moreover, the phylogenomics of the *des* cluster revealed fusions between DesC and Irp2, an iron-related protein first identified in pathogenic *Yersinia* species[250] and shown to be part of the yersiniabactin BGC[251] (in *Nocardia*, *Gordonia araii*, and *Streptomyces carneus*). This fusion correlate with a different gene organization than what is typically seen in *Streptomyces*: *desABCD versus desC-irp2DAB*.

The appearance of multi-domain enzymes from single-domain biosynthetic proteins during the evolution of the dFOs biosynthetic pathway (Fig. 4A) is a very interesting observation, as these genes contribute to chemical diversity. Historically, the main dFOs are those produced by *Streptomyces*: the cyclic dFO-D and the acyl-capped linear form, dFO-B.[233,234,252] In these cases, protein fusions are not seen. As mentioned, novel aryl-capped versions have been reported recently,[113,253] as well as dFO glycoconjugates.[254] This diversity has been further expanded by detection of a broad range of novel linear and cyclic dFOs after subtly modifying the medium in which *Streptomyces chartreusis* was grown.[255] Moreover, at least in the laboratory using model bacterial strains, the potential role of dFOs in mediating interactions has unveiled further chemical diversity with possible ecological implications.[33,252,256–258] Based on microbial challenges using model organisms, these data describe dFOs structural diversification.[33,41] Overall, these studies show how dFOs production and chemical diversity are prompted by abiotic and biotic factors. Mechanistically speaking, an explanation for these observations may lie in the promiscuity of the Des enzymes. Evolutionarily speaking, these observations are important as such overwhelming chemical diversity arises from a conserved, small BGC, in contrast with the one BGC, one main product paradigm. An interesting case along these lines is the production of dimeric hydroxamate cyclic siderophores (*e.g.* avaroferrin, putrebactin, and bisucaberin) by the marine Proteobacterium *Shewanella algae*, where chemical diversification is mediated by the cellular substrate pool, enzyme promiscuity, and crosstalk of enzymes closely related to DesABCD.[259,260] (Fig. 4B). It is interesting to note that dFOs not only have important physiological roles but also morphological implications in *Streptomyces* and other myceliated Actinobacteria.[256,261] Additionally, mutation of the main dFO siderophore-binding receptor in *S. coelicolor*, *desE*, gives rise to a bald phenotype,[236,262] suggesting that chemical diversity even in simple pathways may arise under different adaptive pressures.

4.2. Fixation, distribution, and selective sweeps in BGCs

Evolutionary trajectories that result in the fixation of a complex trait (*e.g.* natural product biosynthesis) can be reconciled under

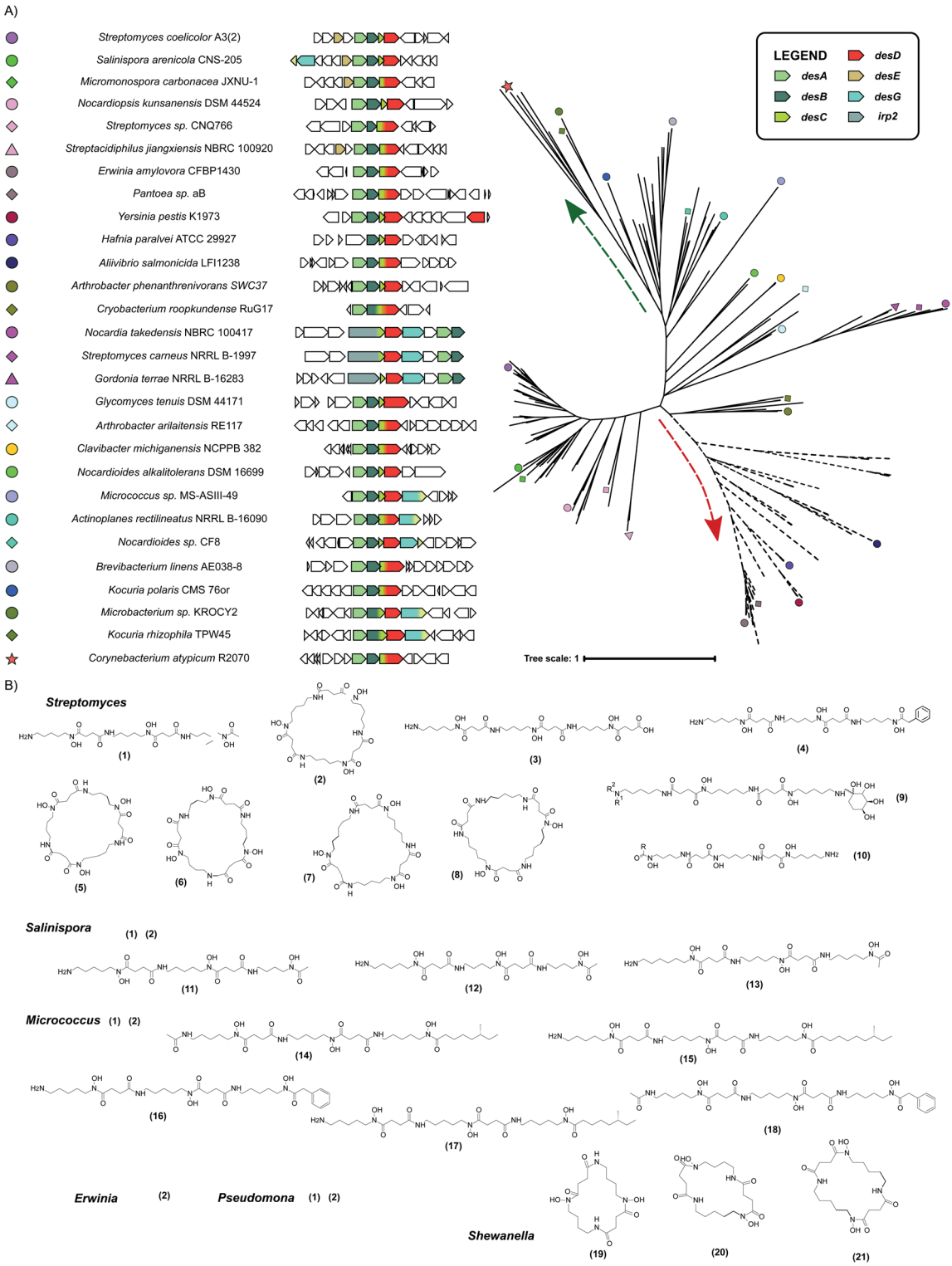

Fig. 4 Evolutionary dynamics and chemical diversity of the dFO BGC. (A) Phylogenomics. CORASON was used to analyse the dFO BGC in organisms found to contain this genotypic trait, namely, Gammaproteobacteria and Actinobacteria. The phylogenetic tree was constructed using the conserved sequences of DesC and DesD, and left unrooted. An Actinobacteria species tree previously reported elsewhere[179] was used to analyse the evolutionary dynamics of this BGC. Selected BGCs that emphasize the evolutionary dynamics discussed in the text are shown. The position of the BGCs within the tree are indicated as colored circles at the tips of each branch with a common origin, next to the corresponding BGCs. Direction of proposed horizontal gene transfer events mentioned in the text are indicated with red dashed arrows. Orthologous genes were colored according to reference BGCs from *Streptomyces coelicolor* A3(2) and *Gordonia terrae* NRRL B-16283 (in bold). For the sake of clarity, default color gradient provided by CORASON as an output was discarded, and protein fusions were manually colored following the gene nomenclature shown in the left-hand side of the panel. (B) Chemical diversity. Chemical structures of dFOs that have been isolated and characterized are shown. The panel is organized following the CORASON tree, starting by the genus *Streptomyces*, which is proposed to be the

the DCME hypothesis (see Section 5.2.1). Considering the impact of selection at the population level, the evolution of promiscuous enzymes and subsequent fixation of pathways may result in the emergence of natural product biosynthetic pathways in populations, perhaps facilitated and accelerated by HGT. Further, selective sweeps through a population can refine the products of BGCs. Such frequency-dependent selection (FDS) has been known for a long time and is able to maintain population diversity (stabilizing FDS) or to purge variation driving the emergence of stable clonal populations (disruptive FDS).[263] It is likely that disruptive FDS is operating when novel bioactive metabolites emerge in a population,[263] yet this model is based on the production of a single bioactive metabolite, instead of metabolic diversity as a whole. In accordance with the One Strain-Many Compounds (OSMAC) approach, Bode and coworkers have very recently observed the occurrence and relevance of population-level metabolic heterogeneity.[201] The dynamic fitness landscapes that shape the metabolism in natural product-producing organisms likely reflect the diversity of the organisms themselves. Yet, some[264] have invoked the Baas Becking hypothesis[265] to explain the distribution of secondary metabolism, after embracing incorrect assumptions of bacterial gene flow, dispersal, and population structure of "everything is everywhere, but the environment selects". If everything were truly everywhere, then the environment alone should determine what is where. Why then do we repeatedly observe so much variation in BGC content across fine-scale geographies?[266,267] Why does horizontal transfer appear to happen so infrequently?[76] Taken to its extreme, if bacteria (and their secondary metabolism) are globally dispersed and their genes undergo rampant horizontal gene transfer then any two bacteria should exhibit species-like, steady gene flow. Yet, there is a huge genetic investment in barriers to HGT such as restriction-modification systems, CRISPR-Cas, and bacteria phage resistance systems.[268,269] Microbial ecologists, evolutionary biologists, and microbiologists hoping to heterologously introduce pathways have long known this is not the case.[270,271] These misconceptions, in part, may be due to careless terminology found in some chemistry-focused literature: ecology, evolution, and environment, while connected, are not the same, nor should these terms be used interchangeably.

Soil is a dynamic and heterogeneous environment composed of a multitude of micro-niches in which organisms reside. The diversity we observe in soil microbes is huge. For example, *Streptomyces* are extremely metabolically diversified,[64] in accordance with the diversity of soil and host-associated habitats in which they are found.[165,272–275] This genus is ancient[76,273] and is an important contributor to global nutrient cycling. Streptomycete bacteria represent >600 validly described species with a remarkable capability to produce a range of clinically useful drug molecules including antibiotics and anticancer agents. Traditionally soil has been a rich hunting ground for *Streptomyces* and other metabolically 'gifted' strains[276] and is the main niche in which plants thrive. A recent study of 830 actinomycete genomes found >11 000 BGCs composed of 4122 families, indicating that there is a vast diversity of strains and chemistry in the environment.[277] Similar scenarios have also been demonstrated for aquatic cyanobacteria inhabiting marine environments.[100]

Recent work in marine actinomycetes has suggested that when closely related organisms are found in the same geographical location, although they share many of the same BGC families, the genes in these BGCs can be under diversifying selection and result in defined populations.[46] It has been hypothesized that the BGCs act adaptively on strains for nutrient and niche defense. However, there has been no systematic testing of this hypothesis nor consideration of its implications on strain diversification and chemical diversity. Environmental niches with similar selective pressures are likely to lead to comparable lifestyles in the organisms that live within them. Therefore, strains occupying equivalent niches might also share functional aspects of their secondary metabolisms. Nevertheless, the ability to acquire novel BGCs, *via* HGT or other novelty-generating evolutionary processes would drive diversification and speciation. Strong evidence of adaptation has been suggested for secondary metabolites as they play a role in competition for resources, communication and signaling. Moreover, it is clear that these metabolites facilitate interaction with a wide range of soil organisms.[278] Single strains are capable of producing >40 molecules that affect gene expression in other taxa,[279] and within closely related taxa such as *Salinispora*, where 75 closely related strains possessed 124 BGCs.[59] Nevertheless, commonality is observed across phylogenetic scales, where different strains produce molecules that target a range of taxa that they are likely to encounter in the soil (antibacterial, antifungal, *etc.*).[280] Within fungus-growing ant symbionts, selective sweeps within shared BGCs (rather than BGC content alone) stratify populations according to geography and "target" organism ecology.[139] Functional equivalence has emerged in soil streptomycetes (see above), given their interactions with other organisms to both protect resources and coordinate signaling. However, it is likely that specialized production and resistance (analogous to kin selection strategies) may drive strain diversity at the microscale within soil. This has been demonstrated for some natural products (*e.g.* siderophores[281]) and may be more widespread.

Phenotypic heterogeneity promotes adaptive evolution: genetic variation provides the raw material upon which natural selection can act.[282] The acquisition of a BGC *via* horizontal gene transfer (HGT) provides this variation and an opportunity

origin of this BGC. Chemical diversity shown is the result of taxonomic distance, biotic and/or abiotic stress, as discussed in the text. Structures were extracted from the literature, as follows: *Streptomyces*,[33,113,233,234,252] *Salinispora*,[241,243] *Micrococcus*[113,253] (Actinobacteria); and *Erwinia*,[223] *Pseudomonas*,[222] *Shewanella*[237,238] (Proteobacteria). Compounds (**1**) to (**18**) are defined as desferrioxamines, as they consist of trimeric hydroxamates. Compounds (**19**), (**20**) and (**21**) from *Shewanella*, *i.e.* avaroferrin, putrebactin and bisucaberin, respectively, differ from *bona fide* dFOs as they dimeric, rather than trimeric, hydroxamates.

to test the effects of a BGC's encoded molecules on fitness in a given environment. This model suggests that if an advantageous trait is gained, then it should spread rapidly through a population. However, this assumes that heterogeneity is purged from the population through selective sweeps (periodic selection). This is not consistent with recent modeling and population studies across a range of bacterial populations.[283] The maintenance of large inventories of BGCs by ecologically cohesive groups of organisms means that, by accounting for gene flow and selection, phenotype and ecological function can inform on how ecological units function at the population level. Applying this to the vast array of BGCs that we observe in the genomes of natural-product producers suggests that the plasticity we observe in expression is most effective when selective pressures are varied.[46] Population-level secondary metabolomes provide an effective evolutionary strategy when spores are dispersed widely in to new environments and when resistance may evolve in a heterogeneous nature in the target species.[46] Moreover, acquisition of BGCs through HGT offers the opportunity to improve on existing traits through the mechanisms discussed above, which will then sweep through the population. In *Salinispora*, evidence for selective sweeps acting on genomes is evident from studies of the widely distributed *S. arenicola versus* the more localized species *S. pacifica*, where the former shows a more stable and less diverse complement of BGCs.[174]

We can also consider these processes in a broader theoretical framework. In evolutionary terms selection acts upon the function of a trait – that is, phenotype is preserved by purifying selection. Interestingly, if a phenotype is the consequence of a multipart trait (such as a BGC) then the modification of this multipart trait can result in functional differences. Imagine a situation where biological activity of natural product could be augmented by a methylation step; if a methyltransferase with the appropriate specificity is acquired in a genome (by duplication or HGT), then modification of the natural product could occur, altering the potential for that molecule offer a competitive advantage in a given environment. Likewise, mutations in an existing enzyme in a pathway may alter the chemical nature of the final product, again altering selective processes associated with production of a natural product. Finally, biodiversity in any natural environment is shaped by the evolutionary processes at work (speciation, diversification and extinction) coupled with ecological processes (competition and dispersal). The eco-evolutionary dynamics of these processes have a profound effect on survival but also driving single phenotypic traits such as natural product production.

Convergent evolution is an elegant solution to problem solving in biology. Importantly, it is a key indicator of a fundamental issue in biology – how repeatable is adaptive evolution? Given that evolution is the result of stochastic events within a specific ecological context, it is unlikely that the same phenotype would evolve in multiple populations without selection. While this is an area that is relatively unexplored in bacteria, it is surprisingly common in plant secondary metabolite evolution with essentially two main categories present: (1) where independent lineages evolve the synthesis of natural products that have the same or very similar functionality, yet arise from different biochemical routes, and (2) where phylogenetically distinct lineages evolve the ability to produce identical molecules.[284] There are often a multitude of ways to arriving at the same chemical solutions to maximize fitness.[197] Genetic barriers, including restricted gene flow and low standing diversity, place limits on these possibilities within individual niches. Convergent evolution can give context to niche boundaries and function. Further, it may well explain why multiple chemical scaffolds have evolved to bind the same biological target proteins. Examples of clinically validated targets for bioactive molecules include the four chemical classes of antibiotics that inhibit Lipid II activity[285] and the multiple molecules that target peptidoglycan biosynthesis.[286] Identical molecules (*e.g.* fosfomycin[287]) or very similar chemistry (*e.g.* dentigerumycin and gerumycin[288]) may also be produced from evolutionarily distinct BGCs. More commonly, however, closely related compounds are produced by different species (*e.g.* salinosporamide A & K produced by different *Salinispora* species[46]). Though bacterial species boundaries can be difficult to define,[63,270,283,289] this may suggest a mechanism to accelerate sympatric speciation in these organisms. The ever-increasing numbers of bacterial genome sequences coupled to high-throughput metabolomics indicates that this will be an area of increased focus in the future.

5. Unifying concepts to bridge evolutionary theory and the biosynthesis of natural products

As Firn and Jones originally introduced, assumptions on the evolutionary origins of secondary metabolism were often based on circumstantial, speculative evidence.[42,43] In contrast, these authors provided an evolutionary framework with sound theoretical foundations to the study of secondary metabolism as a complex trait. Building on this concept many authors have developed their own intuitive ideas with the aim of exploiting evolutionary lessons (see Section 4). By combining phylogenetics (including ancestral sequence reconstructions) with biochemical and structural characterization, it has been possible to decipher the natural history of genes and their proteins, even in complex systems.[290] The latter approach has been referred to as the functional synthesis. However, analogous efforts devoted specifically to the evolution of secondary metabolic enzymes are scarce, and speculative debate about the *raison d'etre* of secondary metabolism continues with little more than circumstantial observations, such as the shunt metabolism hypothesis[164] or the parvome.[291] Why can bacteria produce so many related small molecules? This question is often completely overlooked. Here, we aim to reconcile what is known about natural product biosynthesis in bacteria from different reductionist-focused disciplines (chemistry, biophysics, genetics, microbiology, *etc.*), and to provide an evolutionary framework consistent with the mechanisms underlying the evolution of dynamic chemical diversity. To do so we first provide bounded meanings to the terms secondary and

specialized metabolism, which may lead to a better understanding of the evolution of natural product biosynthesis.

5.1. Defining secondary and specialized metabolism

We believe that there is a need to redefine and standardize key terms to advance our understanding on the evolution of natural product biosynthesis as a research community. We argue that specialized metabolism is intimately associated with niche adaptation and/or fine-scale physiology. In this framework, specialized metabolism is a specific case of secondary metabolism, while secondary metabolism can refer to broader and/or labile interactions. As these interactions are difficult to elucidate, secondary metabolism is a more appropriate classification in cases where ecological functions remain unknown. With these distinctions in mind, secondary and specialized should not be used interchangeably. 'Secondary' was first proposed to provide the classification of "not primary" when referring to metabolism, specifically to distinguish metabolites produced late or after growth phases in laboratory conditions.[292] Today, usage of the term secondary metabolism occurs in diverse contexts, from highly conserved pathways to divergent and strain-specific metabolic pathways. Further, the promiscuity found in secondary metabolism can be thought of as a sort of "underground" trait,[8,9] maintained in genomes due to its underlying, albeit difficult to classify, functions that are only required in certain contexts. We do not wish to suggest (nor do we condone) the use of the term "underground metabolism" to replace "secondary metabolism"; rather we wish only to draw an analogy to highlight promiscuity. Primary or central is shared within a lineage and required for autonomous growth, sometimes supported by promiscuous (underground) enzymes and/or redundant or alternative pathways.[26,27] In contrast, secondary metabolism is often dynamic. The distinction between secondary and specialized is relevant to natural product biosynthesis in real ecological settings (*i.e.* specialized metabolism). This remains a major challenge, especially since environmental conditions are not always well modeled in the laboratory.

In what regard are metabolites specialized? Specialized to a lineage? A niche? Expression from an environmental cue? Expression under a media condition? In bacteria, Davies championed the idea that "it is considered appropriate to refer to these molecules as specialized metabolites to emphasize their importance in microbial ecology", where "these molecules" referred to all secondary metabolites.[292] Indeed, some have used "specialized" as a blanket replacement for "secondary". But can you infer their importance from a gene here or a mass to charge ratio there? What about their ecology? How are they distributed? What natural functions do they serve? Are these functions static or dynamic? What is the producer's growth rate in nature? What is their interaction network? Is your sampling both appropriate and sufficient? The above-mentioned questions are all difficult, perhaps impossible, to address, yet their answers are implied when one calls a metabolite "specialized". It may help to remember that "specialized" arose from the plant metabolism field, where specialized metabolites are defined as those either produced specifically in a single tissue or in response to specific ecological interactions and related stresses. The field of bacterial natural product biosynthesis has much to learn in evolutionary definitions from its plant counterparts.[293] Nevertheless, it must be acknowledged that niche structure in bacterial populations is harder to ascertain than in eukaryotic ecosystems.[46] It is interesting to note, however, that for more than a decade plant biochemists have distinguished between different types of metabolic scenarios in relation to evolutionary and ecological processes, and that mycologists differentiate between BGCs and metabolic gene clusters to describe population-level diversity.[294] Following on these definitions, natural auxotrophy in certain bacterial strains, particularly those that have undergone genome reduction, blurs the line between what is primary and what is secondary.[169] Likewise, how does one define if a contemporary metabolite only conserved in particular extant lineages, yet ancestral lineages were likely secondary (*e.g.* scytonemin, streptomycin)? Recently, secondary metabolism, as we define it, has been predicted within genome-scale metabolic models[295] and rewiring of metabolic networks within a heterologous expression host has seen some industrial success.[296] In nature, population-level, periodic sampling of the components of secondary metabolic pathways allows for the appearance of "more specialized" metabolism with strong physicochemical (*e.g.* siderophores) and biomolecular (*e.g.* antibiotics) activities as the fitness landscape changes, sustaining all sorts of relevant biological and ecological activities. Further, this can occur simply under neutrality due to biophysical constraints, leading to high levels of enzyme and metabolic heterogeneity at the population level.[201,297] Thus, there is no need to rule out one concept or definition over another, as they are interrelated within a dynamic eco-evolutionary process that occurs broadly within bacterial populations, from secondary metabolism to specialized metabolism and (very importantly but until now, neglected) *vice versa* (see next section).

We propose these three definitions:

(i) Primary metabolism refers to pathways of wide taxonomic distribution, which can nevertheless be lineage specific, and whose products are necessary for autonomous viability under most conditions tested. Often, purifying selection will occur throughout the pathway to reduce enzyme promiscuity and intermediates are only important, as they are a path to the final product. This is what Firn and Jones call "integrated" metabolism.

(ii) Secondary metabolism then refers to pathways that are less integrated and less taxonomically distributed than primary metabolism. As these pathways are often the result of enzyme promiscuity and recruitment, these pathways may be either under diversifying or purifying selection depending on the position they adopt during evolutionary dynamics. The products of most BGCs will fall into this category, but it is not limited to only BGCs.

(iii) Specialized metabolism is a special case of secondary metabolism in which the metabolite is experimentally described to contribute to niche-specific functions. The vast

majority of BGC products are unlikely to have experimentally validated niche-specific roles.

It is often only when natural product scientists detect strong biomolecular activities, usually using axenic monocultures, that attention is given to the cognate BGC and biosynthetic pathway responsible for production of the related metabolite. This has prompted circumstantial and presumptuous conclusions in relation to the evolution of natural product biosynthesis, adding further confusion about the use of the terms secondary and specialized metabolism. Although a growing number of studies appreciate the role of natural products in microbial community dynamics and non-human host–microbe interactions,[272,278,298–302] the characterization of metabolites and their pathways that lack obvious use in human welfare (usually produced at low levels as part of secondary metabolism) have been largely overlooked. Together, secondary and specialized metabolites represent the main source of novel drugs and other biotechnologically useful compounds. In nature, they are also a complex phenotype upon which evolution can act. Therefore, understanding the evolutionary forces contributing to "natural" natural product biosynthesis and diversification is of paramount societal importance.

5.2. One pathway, many compounds: how and why?

A common empirical observation of researchers working on natural product biosynthesis is encountering that one bacterial strain produces far too many metabolites than predicted by the number of pathways encoded in its genome. Furthermore, metabolites may be closely related yet distinct from those produced by their close relatives. The chemical diversity encountered during microbial screenings can be overwhelming, often presenting technical and experimental challenges related to purification and reproducibility. Many approaches to address these issues have been developed in recent years. From an evolutionary perspective, these observations do deliver a consistent message: each bacterial strain is capable of producing many metabolites of the same chemical class, usually in a condition or media-dependent manner.[303] This is further complicated by the production of different molecules by closely related organisms grown under exactly the same conditions. The differences could be both qualitative or quantitative, usually addressed by industrial microbiologists during the optimization and scale-up of biotechnological processes.

The One Strain-Many Compounds (OSMAC) approach formalizes the above mentioned empirical observations in a useful method for the discovery of novel natural products in microbes.[303,304] This approach has enabled the discovery of novel variants from the same chemical class, as well as completely novel pathways and their metabolites.[304,305] Indeed, more than fifteen years of research based in the OSMAC approach cannot be ignored: chemical diversity itself must be under selection, within biosynthetic pathways and across them, as otherwise OSMAC phenomena should not occur. Alternatively, neutral evolution of secondary and/or specialized metabolism is pervasive, as in any other evolutionary trait recorded to date. But what are the evolutionary mechanisms underlying the maintenance of chemical diversity as a phenotypic trait? How do these traits relate to other molecular features, such as enzyme promiscuity, enzyme modularity, and the occurrence of BGCs? These are fundamental questions that have yet to be adequately addressed. Here, we aim to provide an evolutionary conceptual framework to begin bridging this knowledge gap.

5.2.1. The dynamic chemical matrix evolutionary hypothesis. We propose the Dynamic Chemical Matrix Evolutionary (DCME) hypothesis to help explain the mechanisms underlying the evolution of secondary and specialized metabolism, building from the screening hypothesis and the data revisited in this paper. Together, the DCME and the screening hypotheses provide a model in which secondary metabolites within a chemical matrix are acted upon by evolutionary forces over a fitness landscape (Fig. 5). Here, the final-point, end-product secondary metabolites encoded in sub-clusters or BGCs are approached as complex biological traits. Genes coding for metabolic enzymes represent a main diversifying force contributing to chemical diversity, especially due to their dynamic expansion-and-recruitment events and promiscuous activities. Selection operating upon these genes drives their clustering, increasing the likelihood of co-transcription. Taken to its extreme, this clustering can include protein fusions. Expression and translation within reasonably similar time frames are required if screening of possible chemical conversions for physicochemical and biomolecular activities is to occur. It should be noted, however, that the hypothesized process of gene clustering (and thus metabolic integration) promotes chemical diversity and not specialization of the emerging pathways, which is only implied by the DCME hypothesis. This distinction is supported by the dependence on media type and/or biotic interactions of the multiple products generated by nearly all BGCs characterized to date. Indeed, when a single main final-product is observed, one should question whether the BGC truly can only produce one molecule, or instead this is a function of external influences such as media optimization, as suggested by the OSMAC approach[305,306] or biotic interactions.[33,41,279]

In this framework, secondary metabolic enzymes operating close together, both in space (cellular organization) and time (temporal regulation) lead to multiple chemical species, each with their own molecular affinities and evolutionary forces acting upon them. Moreover, once a product with strong activity is obtained, the structural and functional space surrounding this product is not necessarily explored in response to adaptive conflict. It is equally plausible that shifting fitness landscapes will result in negative (purifying) selection, rather than positive selection (as is often assumed by the natural product community), a possibility only accounted for by the DCME hypothesis. The constraints of negative selection can be overcome by the relatively easy, one-step, single-gene acquisition of immunity mechanisms. Interestingly, many BGCs include as main resistance mechanisms ABC transport systems or efflux pumps, rather than additional specific natural product resistance genes or mutations (*e.g.* ribosome mutations or the modifying enzymes embraced by ARTS). Therefore, it is reasonable to

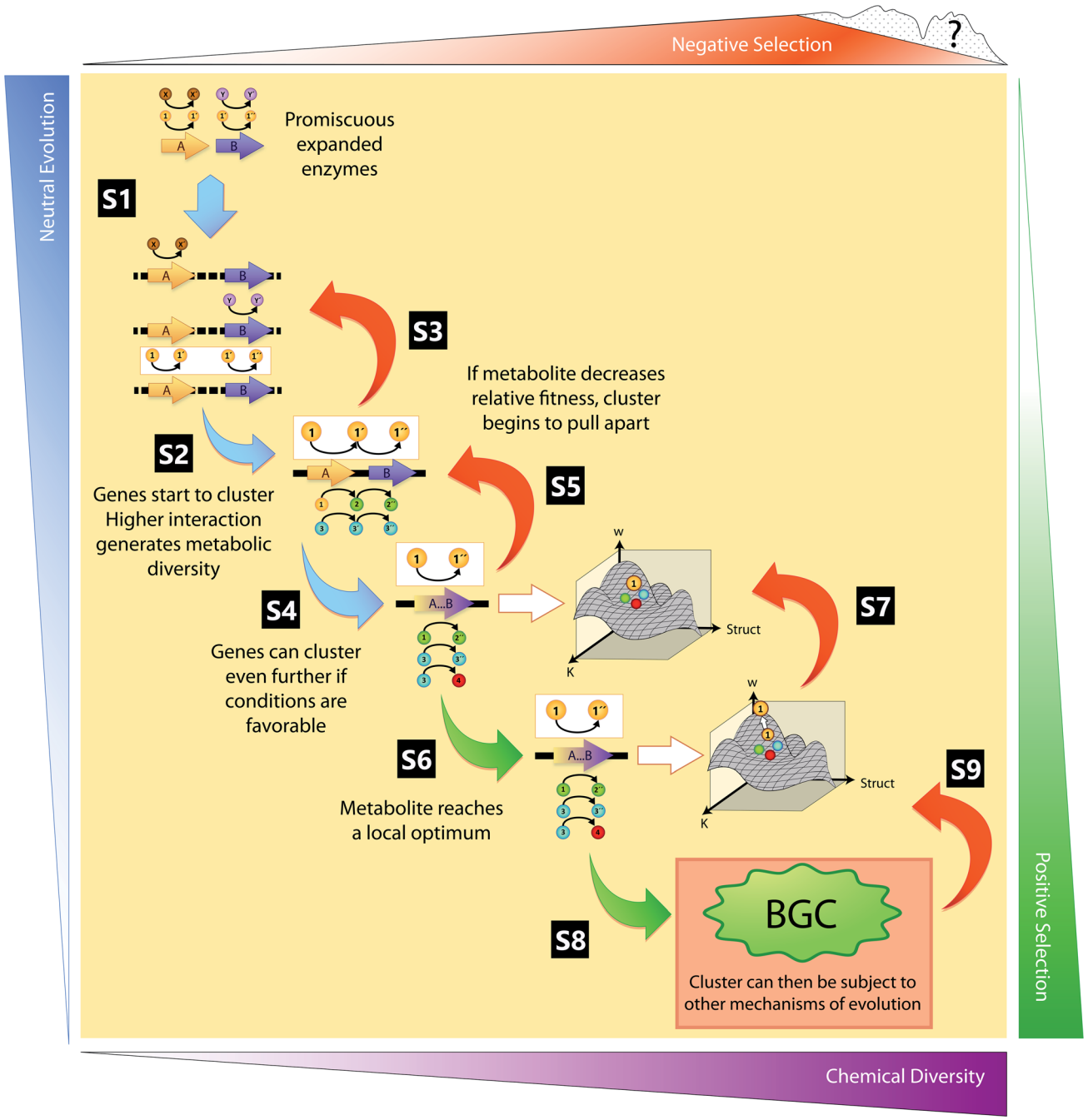

Fig. 5 The Dynamic Chemical Matrix Evolutionary (DCME) hypothesis. Here we propose a number of scenarios (S*x*) that might explain evolution of secondary and specialized metabolism and give rise to BGCs. This process can initially be triggered by expansion-and-recruitment events of enzyme genes with promiscuous activity, under the different evolutionary forces (neutral, negative, and positive). Among other substrates, these promiscuous enzymes might be able to convert the products of enzymes undergoing the same process, thus producing new metabolites (S1). Genes coding for these promiscuous enzymes might come closer, promoting increased interactions among promiscuous enzymes and increasing the probability of formation of new metabolites (S2, S4, S6 and S8). However, during this process, due to undesirable biomolecular activities, genes may be lost, lose effectiveness, and/or drift apart (S3, S5, S7 and S9). On favorable conditions, genes cluster further and increase production of a slightly advantageous metabolite; all while new products of enzyme promiscuity are still produced at low levels resulting in more chemical diversity (S2, S4, S6 and S8). All metabolites navigate through a fitness landscape (see Fig. 6) that determines whether biosynthetic gene clusters and metabolite production are conserved or purged. If new metabolites are toxic, decreasing fitness, gene inactivation and/or de-clustering can eliminate production of metabolite. In case one of the new metabolites provide a fitness advantage, selection drives cluster formation and production to a local optimum within a relative fitness landscape, but without eliminating chemical diversity. Further positive selection and consequential clustering produces a BGC, which can be subject to other mechanisms of evolution such as HGT and genetic sweeps. Blue, green and red arrows indicate the events in which neutral, positive and negative selection are involved, meanwhile axis gradients indicate the strength of its involvement. At late stages, negative selection can be eliminated following innovation peaks related to acquisition of resistance and/or regulatory elements with the potential to effectively mitigate the deleterious effects of emerging metabolites with strong undesirable biomolecular activities. A purple arrow is used to indicate progression of chemical diversity.

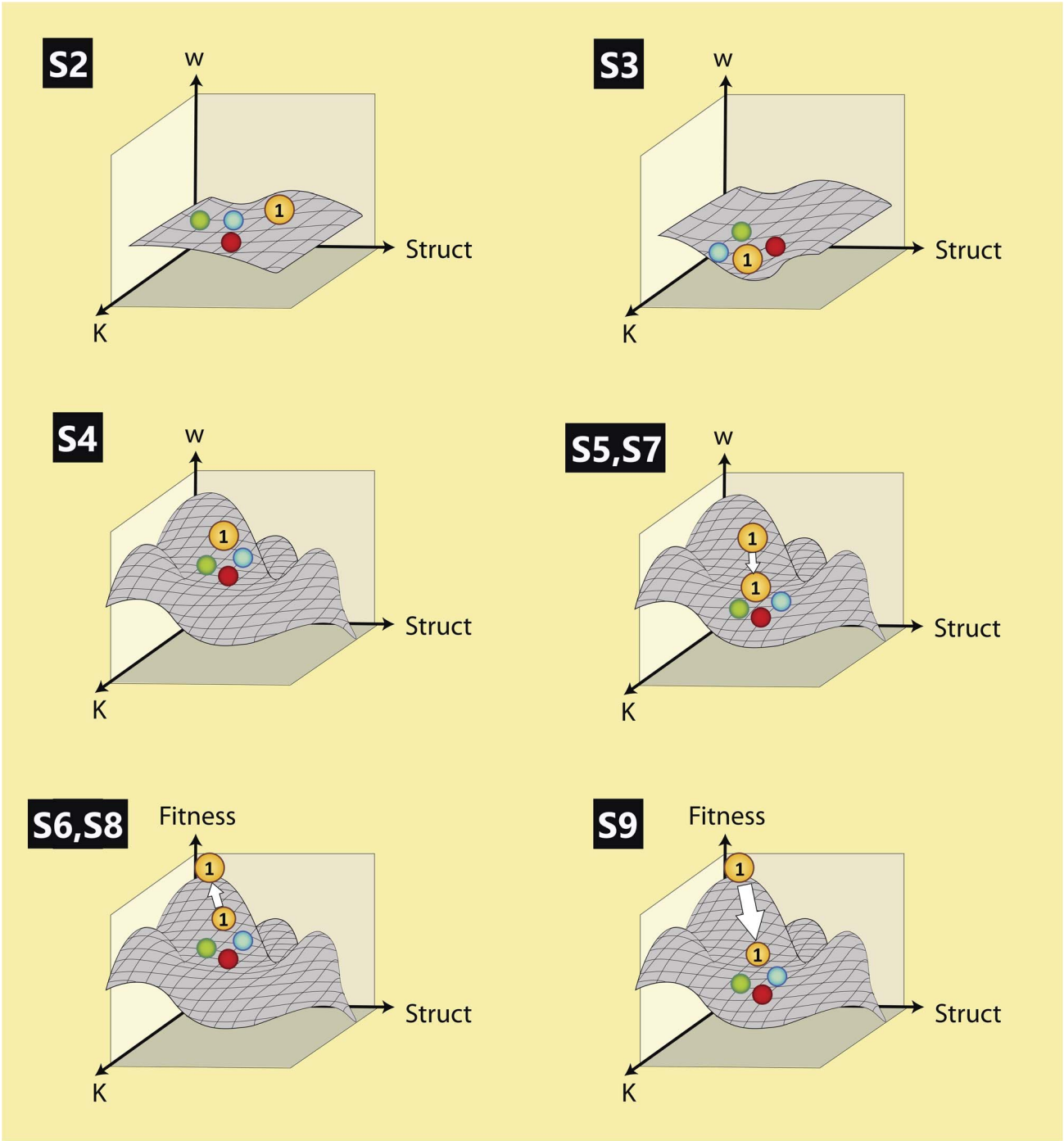

Fig. 6 Possible scenarios within the fitness landscape governing the DCME hypothesis. Clustering and disintegration of BGCs is mainly driven by the interplay of three properties: chemical structure distances (*S*, *e.g.* Tanimoto index), affinity constants related to biomolecular activities (*K*) and relative fitness (*w*). A dynamic fitness landscape describing these properties dictates the types of selection necessary for evolution of secondary and specialized metabolism. During the scenarios described in Fig. 5, the fitness landscape shifts as follows: after genetic expansions, metabolic products of promiscuous enzymes explore the mostly neutral landscape reaching slightly advantageous peaks leading to clustering (S2) or slightly deleterious valleys causing the genes to fall apart (S3). Given the appropriate conditions, a metabolite with a biomolecular activity arises and the fitness landscape changes, more prominent peaks and valleys develop, and selective forces can begin to act upon the genes encoding for the promiscuous enzymes (S4). These selective forces can be either of negative nature (S5 and S7) favoring disintegration of gene cluster, or positive selection leading metabolite production to a local optimum, clustering of genes and eventually formation of a BGC (S6 and S8). Once formed, BGCs can also fall apart if a sufficient strong negative selection acts upon it (S9).

hypothesize that these mechanisms evolved relatively late with regards to the evolution of the biosynthetic machinery as resistance mechanisms are typically found outside of BGCs.[173,307] Likewise, fine-tuning of gene regulation may also come later within this model.

This framework can be formalized through the relationships between chemical structural distances (*S*, *e.g.* Tanimoto index), the physicochemical function or biomolecular activity (*F*, *e.g.* affinity constants) and relative fitness (*w*) (Fig. 6). Describing how fitness is a function of these key properties is akin to the fitness landscapes often used to describe other organismal macro-biological systems.[308] Further, this framework adds classical evolutionary and mechanistic dimensions to the complex trait that natural product biosynthesis represents. In such a landscape, both negative and positive selection can select metabolic products through acting at the level of genes. Therefore, diversification is reflected *via* a promiscuity-driven dynamic chemical matrix far more complex than that perceived by the biocombinatorial nature of BGCs (that typically only account for a single final-product). This chemical matrix should therefore be viewed as a dynamic entity rather than a final-point product, with negative selection occurring as selective sweeps to ensure removal of deleterious combinations of sub-clusters (*e.g.* promiscuous enzyme combinations) throughout the different stages of the proposed evolutionary process. Arguably, and perhaps somewhat counter intuitively, negative selection increases with molecular activity, until resistance and regulatory genes can be recruited. How positive selection may govern this process once resistance and/or regulatory genes are recruited remains an open question, but the DCME hypothesis suggests that this could be related to dynamic assembly and disassembly of BGCs. Further, this may allow organisms escape from adaptive conflict[309,310] related to undesirable physicochemical and biomolecular activities of natural products while exploring the chemical space for beneficial natural products.

Together, the DCME hypothesis postulates that in the early stages of pathway evolution in secondary metabolism neutral forces are dominant (in contrast to directional selection), such that metabolic diversity can be explored, increasing the likelihood of retaining (small) benefits under dynamic ecological conditions. Subsequently, positive forces can select for advantageous activities within tight structure–function relationships. Once a BGC is formed, it may be subject to the different genetic mechanisms that have been previously suggested for the evolution of BGCs, namely, HGT[67] and genetic sweeps.[46] Thus, one may question whether a pathway emerging from these promiscuity-driven explorations of chemical matrices can be called a secondary metabolic pathway. In essence, it is a central metabolic pathway, even if only existing in a monoclonal population. The final product is the trait under selection, not the metabolic diversity surrounding its origin. As an analogy, L-tryptophan auxotrophic mutants (*i.e.* those lacking the L-Trp biosynthetic pathway) can grow in L-Trp-supplemented media. It is only when the environment (albeit a lab-constructed one) changes to L-Trp-free media do they die from lack of an "essential" primary metabolic pathway. This illustrates the situational nature of metabolic effects on phenotype and highlights that essentiality is an artificial, human construct. This extends to the designation of primary *versus* secondary metabolism. Metabolic products exist as complex traits that sustain the survival and fitness of that particular organism in a particular niche, even if the cognate BGC is obtained after HGT (*e.g.* pathogenicity islands). A case emphasizing this situation relates to the ancestral evolution of scytonemin BGC in *Nostoc* cyanobacteria.[60]

The screening hypothesis hints towards the possibility of chemical diversity. However, it lacks possible mechanistic insights into how and why this may occur. The DCME model suggests that mechanistically (in evolutionary terms) this can only take place if BGCs are seen as dynamic entities under the influence of positive, negative, and neutral evolutionary forces. Negative selection is the main difference with the screening hypothesis, fundamentally providing a hypothesis for the evolution of metabolic diversity that goes beyond the idea of biomolecular activity (which necessarily should be addressed in a one-by-one case). In sum, the DCME model provides a general mechanism based on evolutionary theory of dynamic fitness landscapes.

6. Concluding remarks and perspectives

Evolution-inspired synthetic biology has benefited enormously from the availability and accelerated development of *ad hoc* computational and chemical tools[155,184,311–313] which have enabled large-scale analysis of genomic and metabolomic datasets towards engineering predictions. Moreover, general bioinformatics tools have identified a growing biological parts list for pathway engineers, including promoter sequences and terminators correct boundaries of domains and modules, new enzyme functions and biosynthetic logic,[314] and hosts providing suitable metabolic chassis for heterologous expression.[315] Molecular networking analyses of mass spectrometry data have similarly deepened our understanding of metabolite diversity and production.[316–318] Despite these advances, one might predict that the large amount of biological and chemical data should have led to more successful synthetic biology schemes for natural products. Whether this anomaly is hopeful or disparaging remains debated. We remain hopeful. Patterns in sequence and spectra are the result of many millions of years of evolution, shaping the biological components of natural product biosynthesis. Evolution-inspired synthetic biology is still in its infancy. In the years ahead, understanding biosynthetic evolutionary mechanisms should be a core research goal, as these insights promise more efficient design of biological systems for the synthesis of designer metabolites and their precursors.

Three specific topics are important to address in the short term: (i) as technology develops and databases grow, new chemical classes and biosynthetic logic will continue to emerge. For example, there is little doubt that new RiPP chemical classes will emerge at a faster pace than originally conceived.[319,320]

Further, RiPPs often break traditional biosynthetic paradigms and can exhibit unprecedented enzymology. However, we know remarkably little as to the why's and how's of this fascinating new metabolic world. It will be important to accompany discovery-oriented studies with an evolutionary perspective if we are to fully harness chemical diversity. (ii) In order to better exploit structural and biochemical data, comprehensive phylogenies in the form of open, online, and real-time platforms are much needed. As the Dali server[321] allows a structural biologist to investigate the evolution of their protein fold of interest, natural product chemists and microbiologists would enormously benefit from similar *ad hoc* resources to decipher the evolutionary origins of their enzyme of interest in a consolidated and systematic way. (iii) While extensive historical attention has been paid to modular megasynthases, there is an obvious need for further research into non-modular secondary metabolic enzymes and their origins. To do so, our BGC-centric view must be broadened to include alternative and more complex metabolic possibilities, even if this demands the development of new computational resources (*e.g.* pathway databases supplementing full genomes and/or metagenomes in which they occur) and experimental tools to uncover enzymatic crosstalk. Addressing of these three points will not only inspire the development of new approaches but will advance our understanding of how natural products are shaped by evolutionary processes.

In the last decade, our understanding of natural product biosynthesis has benefitted from unprecedented advances in chemistry, genetics, and structural biology. Genomics, metabolomics, and chemical biology have opened a proverbial Pandora's box of metabolic possibilities. In turn, recent efforts have largely focused on harnessing the overwhelming chemical and genetic diversity stemming from this 'Pandoric' metabolism. Researchers in secondary metabolism must keep in mind that evolution is both how and why this natural diversity exists. We believe that only in light of evolution will the complexity of natural product biosynthesis become more accessible and easily manipulated. Here, we have provided an overview of the evolutionary concepts and theory related to natural product biosynthesis through the lens of our own understanding of this complex biological trait. Given the relatively higher success rates of those adopting an evolutionary mindset during drug discovery and synthetic biology efforts, we envisage a promising future for the study of the natural product biosynthesis. In the years to come, we urge the field to move from final product dominated concepts towards a more holistic view that reflects the evolutionary processes that shape secondary metabolites and the genes that assemble them.

7. Definitions

Primary metabolism: pathways of wide distribution within a lineage whose products are necessary for autonomous viability under most conditions tested. Often, purifying selection will occur throughout the pathway to reduce enzyme promiscuity and intermediates are only important as they are a path to the final product.

Secondary metabolism: pathways that are less integrated and less taxonomically distributed than primary metabolism. As these pathways are often the result of enzyme promiscuity and recruitment, these pathways may be either under diversifying or purifying selection depending on if they are early or late in the evolutionary process, respectively.

Specialized metabolism: a special case of secondary metabolism in which the metabolite is experimentally described to contribute to niche-specific functions.

Chemical matrix: structure, biomolecular or physicochemical activity, and relative fitness derived from all possible promiscuous and non-promiscuous enzyme activities of the metabolic network, including all their possible combinations, at a given point in time.

Dynamic Chemical Matrix Evolutionary (DCME) hypothesis: a graphical, virtual representation of how the chemical matrix changes over time in response to evolutionary forces, integrating structure, biomolecular activity, and relative fitness. It incorporates negative selection, in addition to positive selection and neutral evolution, as an often-overlooked evolutionary force for exploration of the chemical diversity throughout evolutionary dynamics of BGCs and their biosynthetic sub-clusters.

8. Conflicts of interest

There are no conflicts to declare.

9. Acknowledgements

We would like to thank Angélica Cibrián-Jaramillo, Pablo Cruz-Morales, Cameron Currie, Jo Handelsman, David A. Hodgson, Marnix Medema, and Kapil Thalan for fruitful and insightful discussions about the evolution of natural product biosynthesis throughout the years. We thank Erika Cruz-Bonilla for help during preparation of figures related to the DCME hypothesis, and most importantly, ACJ for directing us towards fitness landscapes. FBG acknowledges funding from the Royal Society - Newton Advanced Fellowship (NAF\R2\180631); and grants from Conacyt (No. 285746) and Cinvestav-IPN (No. 103), Mexico. PAH acknowledges funding from BBSRC (BB/N023544/1), NERC (NE/M001415/1), BBSRC/NPRONET (NPRONET POC045) and the Microbiology Society, United Kingdom.

10. References

1 H. Jenke-Kodama, R. Müller and E. Dittmann, *Prog. Drug Res.*, 2008, **65**(119), 121–140.
2 I. Schmitt and F. K. Barker, *Nat. Prod. Rep.*, 2009, **26**, 1585.
3 M. Adamek, M. Alanjary and N. Ziemert, *Nat. Prod. Rep.*, 2019, **36**, 1295–1312.
4 D. Davidi, L. M. Longo, J. Jabłońska, R. Milo and D. S. Tawfik, *Chem. Rev.*, 2018, **118**, 8786–8797.
5 D. L. Trudeau and D. S. Tawfik, *Curr. Opin. Biotechnol.*, 2019, **60**, 46–52.
6 O. Khersonsky and D. S. Tawfik, *Annu. Rev. Biochem.*, 2010, **79**, 471–505.
7 A. Peracchi, *Trends Biochem. Sci.*, 2018, **43**, 984–996.

8 S. D. Copley, *Trends Biochem. Sci.*, 2015, **40**, 72–78.
9 S. D. Copley, *Curr. Opin. Struct. Biol.*, 2017, **47**, 167–175.
10 A. Aharoni, L. Gaidukov, O. Khersonsky, S. M. Gould, C. Roodveldt and D. S. Tawfik, *Nat. Genet.*, 2005, **37**, 73–76.
11 R. A. Jensen, *Annu. Rev. Microbiol.*, 1976, **30**, 409–425.
12 S. Martínez Cuesta, S. A. Rahman, N. Furnham and J. M. Thornton, *Biophys. J.*, 2015, **109**, 1082–1086.
13 S. D. Brown and P. C. Babbitt, *J. Biol. Chem.*, 2014, **289**, 30221–30228.
14 M. S. Newton, V. L. Arcus, M. L. Gerth and W. M. Patrick, *Curr. Opin. Struct. Biol.*, 2018, **48**, 110–116.
15 Á. Tóth-Petróczy and D. S. Tawfik, *Curr. Opin. Struct. Biol.*, 2014, **26**, 131–138.
16 J. D. Bloom and F. H. Arnold, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 9995–10000.
17 N. Tokuriki and D. S. Tawfik, *Science*, 2009, **324**, 203–207.
18 M. A. DePristo, D. M. Weinreich and D. L. Hartl, *Nat. Rev. Genet.*, 2005, **6**, 678–687.
19 G. K. A. Hochberg and J. W. Thornton, *Annu. Rev. Biophys.*, 2017, **46**, 247–269.
20 M. Kaltenbach, J. R. Burke, M. Dindo, A. Pabis, F. S. Munsberg, A. Rabin, S. C. L. Kamerlin, J. P. Noel and D. S. Tawfik, *Nat. Chem. Biol.*, 2018, **14**, 548–555.
21 N. Waglechner, A. G. McArthur and G. D. Wright, *Nat. Microbiol.*, 2019, **4**(11), 1862–1871.
22 F. Baier, N. Hong, G. Yang, A. Pabis, C. M. Miton, A. Barrozo, P. D. Carr, S. C. Kamerlin, C. J. Jackson and N. Tokuriki, *eLife*, 2019, **8**, 232793.
23 C. M. Miton, S. Jonas, G. Fischer, F. Duarte, M. F. Mohamed, B. van Loo, B. Kintses, S. C. L. Kamerlin, N. Tokuriki, M. Hyvönen and F. Hollfelder, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, E7293–E7302.
24 L. Noda-García and F. Barona-Gómez, *Mobile Genet. Elem.*, 2013, **3**, e26439.
25 L. Noda-García, A. R. Camacho-Zarco, S. Medina-Ruíz, P. Gaytán, M. Carrillo-Tripp, V. Fülöp and F. Barona-Gómez, *Mol. Biol. Evol.*, 2013, **30**, 2024–2034.
26 A. Khanal, S. Yu McLoughlin, J. P. Kershner and S. D. Copley, *Mol. Biol. Evol.*, 2015, **32**, 100–108.
27 K. Verdel-Aranda, S. T. López-Cortina, D. A. Hodgson and F. Barona-Gómez, *Microb. Biotechnol.*, 2015, **8**, 239–252.
28 E. A. Verduzco-Castro, K. Michalska, M. Endres, A. L. Juarez-vazquez, L. Noda-garcia, C. Chang, C. S. Henry, G. Babnigg, A. Joachimiak and F. Barona-Gomez, *Biochem. J.*, 2016, 1–14.
29 M. G. Chevrette, P. A. Hoskisson and F. Barona-Gómez, in *Comprehensive Natural Products III. Reference Module in Chemistry, Molecular Sciences and Chemical Engineering*, Elsevier, 2019.
30 W. Kreis and J. Munkert, *J. Exp. Bot.*, 2019, **70**, 1435–1445.
31 C. C. Ladner and G. J. Williams, *J. Ind. Microbiol. Biotechnol.*, 2016, **43**, 371–387.
32 S. M. Carpenter and G. J. Williams, *ACS Chem. Biol.*, 2018, **13**, 3361–3373.
33 M. F. Traxler, J. D. Watrous, T. Alexandrov, P. C. Dorrestein and R. Kolter, *mBio*, 2013, **4**, 1–12.
34 K. Bin Kang, M. Ernst, J. J. J. Hooft, R. R. Silva, J. Park, M. H. Medema, S. H. Sung and P. C. Dorrestein, *Plant J.*, 2019, 14292.
35 C. R. Pye, M. J. Bertin, R. S. Lokey, W. H. Gerwick and R. G. Linington, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **114**, 5601–5606.
36 M. B. Austin, P. E. O'Maille and J. P. Noel, *Nat. Chem. Biol.*, 2008, **4**, 217–222.
37 Y. Lim, M. Go and W. Yew, *Molecules*, 2016, **21**, 806.
38 M. Lynch, *Nat. Rev. Genet.*, 2007, **8**, 803–813.
39 D. Sardar and E. W. Schmidt, *Curr. Opin. Chem. Biol.*, 2016, **31**, 15–21.
40 Q. Zhang, X. Yang, H. Wang and W. a. van der Donk, *ACS Chem. Biol.*, 2014, **9**, 2686–2694.
41 M. F. Traxler and R. Kolter, *Nat. Prod. Rep.*, 2015, **32**, 956–970.
42 R. Firn, *Nature's chemicals – The natural products that shaped our world*, 2009.
43 R. D. Firn and C. G. Jones, *J. Exp. Bot.*, 2009, **60**, 719–726.
44 R. D. Firn and C. G. Jones, *Nat. Prod. Rep.*, 2003, **20**, 382.
45 R. C. Hider and X. Kong, *Nat. Prod. Rep.*, 2010, **27**, 637.
46 P. R. Jensen, *Trends Microbiol.*, 2016, **24**, 968–977.
47 G. Moghe and R. L. Last, *Plant Physiol.*, 2015, **169**, 1512–1523.
48 A. Eyre-Walker and P. D. Keightley, *Nat. Rev. Genet.*, 2007, **8**, 610–618.
49 A.-R. Tidjani, J.-N. Lorenzi, M. Toussaint, E. van Dijk, D. Naquin, O. Lespinet, C. Bontemps and P. Leblond, *mBio*, 2019, **10**, 1–12.
50 C. Stuttard, *J. Gen. Microbiol.*, 1982, **128**, 115–121.
51 J. Burke, D. Schneider and J. Westpheling, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 6289–6294.
52 G. S. Pettis, *Environ. Microbiol. Rep.*, 2018, **10**, 503–510.
53 E. Bordeleau, M. G. Ghinet and V. Burrus, *Mobile Genet. Elem.*, 2012, **2**, 119–124.
54 B. Goldman, S. Bhat and L. J. Shimkets, *PLoS One*, 2007, **2**, 1–9.
55 L. C. Clark and P. A. Hoskisson, *PLoS One*, 2011, **6**(10), e25049.
56 U. Bergthorsson, D. I. Andersson and J. R. Roth, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 17004–17009.
57 A. Mira, H. Ochman and N. A. Moran, *Trends Genet.*, 2001, **17**, 589–596.
58 T. Y. Pang and M. J. Lercher, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 187–192.
59 N. Ziemert, A. Lechner, M. Wietz, N. Millan-Aguinaga, K. L. Chavarria and P. R. Jensen, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, E1130–E1139.
60 F. Garcia-Pichel, J. Lombard, T. Soule, S. Dunaj, S. H. Wu and M. F. Wojciechowski, *mBio*, 2019, **10**, 1–14.
61 Y. M. Shi, A. O. Brachmann, M. A. Westphalen, N. Neubacher, N. J. Tobias and H. B. Bode, *Nat. Chem. Biol.*, 2019, **15**, 331–339.
62 S. Wright, in *Proceedings of the VI International Conference on Genetics*, 1932, pp. 356–366.

63 B. J. Shapiro, J. Friedman, O. X. Cordero, S. P. Preheim, S. C. Timberlake, G. Szabo, M. F. Polz and E. J. Alm, *Science*, 2012, **336**, 48–51.
64 M. G. Chevrette and C. R. Currie, *J. Ind. Microbiol. Biotechnol.*, 2019, **46**, 257–271.
65 M. J. Smanski, D. C. Schlatter and L. L. Kinkel, *J. Ind. Microbiol. Biotechnol.*, 2016, **43**, 115–128.
66 L. T. Fernández-Martínez and P. A. Hoskisson, *Curr. Opin. Microbiol.*, 2019, **51**, 16–21.
67 H. Bruns, M. Crüsemann, A.-C. Letzel, M. Alanjary, J. O. McInerney, P. R. Jensen, S. Schulz, B. S. Moore and N. Ziemert, *ISME J.*, 2018, **12**, 320–329.
68 H. Jenke-Kodama, A. Sandmann, R. Müller and E. Dittmann, *Mol. Biol. Evol.*, 2005, **22**, 2027–2039.
69 H. Jenke-Kodama, T. Börner and E. Dittmann, *PLoS Comput. Biol.*, 2006, **2**, e132.
70 H. Jenke-Kodama and E. Dittmann, *Phytochemistry*, 2009, **70**, 1858–1866.
71 T. Nguyen, K. Ishida, H. Jenke-Kodama, E. Dittmann, C. Gurgui, T. Hochmuth, S. Taudien, M. Platzer, C. Hertweck and J. Piel, *Nat. Biotechnol.*, 2008, **26**, 225–233.
72 M. a. Fischbach, C. T. Walsh and J. Clardy, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 4601–4608.
73 C. T. Walsh and M. a. Fischbach, *J. Am. Chem. Soc.*, 2010, **132**, 2469–2493.
74 S. Donadio, M. Staver, J. McAlpine, S. Swanson and L. Katz, *Science*, 1991, **252**, 675–679.
75 W. P. Revill, M. J. Bibb and D. A. Hopwood, *J. Bacteriol.*, 1996, **178**, 5660–5667.
76 B. R. McDonald and C. R. Currie, *mBio*, 2017, **8**, e00644-17.
77 R. Joynt and R. F. Seipke, *Microbiology*, 2018, **164**, 28–39.
78 J. Zucko, P. F. Long, D. Hranueli and J. Cullum, *J. Ind. Microbiol. Biotechnol.*, 2012, **39**, 1541–1547.
79 M. H. Medema, P. Cimermancic, A. Sali, E. Takano and M. A. Fischbach, *PLoS Comput. Biol.*, 2014, **10**, e1004016.
80 J. F. Barajas, J. M. Blake-Hedges, C. B. Bailey, S. Curran and J. D. Keasling, *Synth. Syst. Biotechnol.*, 2017, **2**, 147–166.
81 S. Kunakom and A. S. Eustáquio, *mSystems*, 2019, **4**, 1–4.
82 J. J. Zhang, X. Tang and B. S. Moore, *Nat. Prod. Rep.*, 2019, **36**(9), 1313–1332.
83 D. a. Hopwood, *Methods Enzymol.*, 2009, **458**, xix–xxi.
84 H.-S. Kang, *J. Ind. Microbiol. Biotechnol.*, 2017, **44**, 285–293.
85 J. Navarro-Muñoz, N. Selem-Mojica, M. Mullowney, S. Kautsar, J. Tryon, E. Parkinson, E. De Los Santos, M. Yeong, P. Cruz-Morales, S. Abubucker, A. Roeters, W. Lokhorst, A. Fernandez-Guerra, L. Teresa Dias Cappelini, R. Thomson, W. Metcalf, N. Kelleher, F. Barona-Gómez and M. H. Medema, bioRxiv.445270, 2018.
86 B. C. Hoefler, R. M. Stubbendieck, N. K. Josyula, S. M. Moisan, E. M. Schulze and P. D. Straight, *Cell Chem. Biol.*, 2017, **24**, 1238–1249.
87 J. R. Lohman, M. Ma, J. Osipiuk, B. Nocek, Y. Kim, C. Chang, M. Cuff, J. Mack, L. Bigelow, H. Li, M. Endres, G. Babnigg, A. Joachimiak, G. N. Phillips and B. Shen, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 12693–12698.
88 A. Vingadassalon, F. Lorieux, M. Juguet, G. Le Goff, C. Gerbaud, J. L. Pernodet and S. Lautru, *ACS Chem. Biol.*, 2015, **10**, 601–610.
89 G. Santoyo and D. Romero, *FEMS Microbiol. Rev.*, 2005, **29**, 169–183.
90 L. Zhang, T. Hashimoto, B. Qin, J. Hashimoto, I. Kozone, T. Kawahara, M. Okada, T. Awakawa, T. Ito, Y. Asakawa, M. Ueki, S. Takahashi, H. Osada, T. Wakimoto, H. Ikeda, K. Shin-ya and I. Abe, *Angew. Chem., Int. Ed.*, 2017, **56**, 1740–1745.
91 A. T. Keatinge-Clay, *Angew. Chem., Int. Ed.*, 2017, **56**, 4658–4660.
92 T. Stachelhaus, H. D. Mootz and M. A. Marahiel, *Chem. Biol.*, 1999, **6**, 493–505.
93 G. L. Challis, J. Ravel and C. A. Townsend, *Chem. Biol.*, 2000, **7**, 211–224.
94 C. Rausch, I. Hoof, T. Weber, W. Wohlleben and D. H. Huson, *BMC Evol. Biol.*, 2007, **7**, 78.
95 J. M. Reimer, I. Harb, O. G. Ovchinnikova, J. Jiang, C. Whitfield and T. M. Schmeing, *ACS Chem. Biol.*, 2018, **13**, 3161–3172.
96 J. M. Reimer, A. S. Haque, M. J. Tarry and T. M. Schmeing, *Curr. Opin. Struct. Biol.*, 2018, **49**, 104–113.
97 S. Mori, A. H. Pang, T. A. Lundy, A. Garzan, O. V. Tsodikov and S. Garneau-Tsodikova, *Nat. Chem. Biol.*, 2018, **14**, 428–430.
98 A. H. Pang, S. Garneau-Tsodikova and O. V. Tsodikov, *J. Struct. Biol.*, 2015, **192**, 349–357.
99 S. Mori, A. H. Pang, N. Thamban Chandrika, S. Garneau-Tsodikova and O. V. Tsodikov, *Nat. Commun.*, 2019, **10**, 1255.
100 S. Meyer, J.-C. Kehr, A. Mainz, D. Dehm, D. Petras, R. D. Süssmuth and E. Dittmann, *Cell Chem. Biol.*, 2016, **23**, 462–471.
101 K. A. J. Bozhüyük, F. Fleischhacker, A. Linck, F. Wesche, A. Tietze, C.-P. Niesert and H. B. Bode, *Nat. Chem.*, 2018, **10**, 275–281.
102 K. A. J. Bozhüyük, A. Linck, A. Tietze, J. Kranz, F. Wesche, S. Nowak, F. Fleischhacker, Y.-N. Shi, P. Grün and H. B. Bode, *Nat. Chem.*, 2019, **11**, 354670.
103 H.-W. Nützmann, A. Huang and A. Osbourn, *New Phytol.*, 2016, **211**, 771–789.
104 Q. Shou, L. Feng, Y. Long, J. Han, J. K. Nunnery, D. H. Powell and R. A. Butcher, *Nat. Chem. Biol.*, 2016, **12**, 770–772.
105 T. Izoré, J. Tailhades, M. H. Hansen, J. A. Kaczmarski, C. J. Jackson and M. J. Cryle, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 2913–2918.
106 J. E. Pool and C. F. Aquadro, *Mol. Ecol.*, 2007, **16**, 2844–2851.
107 M. Rebeiz, J. E. Pool, V. A. Kassner, C. F. Aquadro and S. B. Carroll, *Science*, 2009, **326**, 1663–1667.
108 H. Bastide, A. Yassin, E. J. Johanning and J. E. Pool, *BMC Evol. Biol.*, 2014, **14**, 179.
109 H. Bastide, J. D. Lange, J. B. Lack, A. Yassin and J. E. Pool, *Genetics*, 2016, **204**, 1307–1319.
110 R. Müller, *Chem. Biol.*, 2004, **11**, 4–6.

111 H. Wang, D. P. Fewer, L. Holm, L. Rouhiainen and K. Sivonen, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 9259–9264.
112 J. Gubbens, C. Wu, H. Zhu, D. V. Filippov, B. I. Florea, S. Rigali, H. S. Overkleeft and G. P. van Wezel, *ACS Chem. Biol.*, 2017, **12**, 2756–2766.
113 P. Cruz-Morales, H. E. Ramos-Aboites, C. Licona-Cassani, N. Selem-Mójica, P. M. Mejía-Ponce, V. Souza-Saldívar and F. Barona-Gómez, *FEMS Microbiol. Ecol.*, 2017, **93**, 1–12.
114 S. Kosol, M. Jenner, J. R. Lewandowski and G. L. Challis, *Nat. Prod. Rep.*, 2018, **35**, 1097–1109.
115 R. Ueoka, A. R. Uria, S. Reiter, T. Mori, P. Karbaum, E. E. Peters, E. J. N. Helfrich, B. I. Morinaka, M. Gugger, H. Takeyama, S. Matsunaga and J. Piel, *Nat. Chem. Biol.*, 2015, **11**, 705–712.
116 C. Du and G. P. van Wezel, *Proteomics*, 2018, **18**, 1700332.
117 E. D. James, B. Knuckley, N. Alqahtani, S. Porwal, J. Ban, J. A. Karty, R. Viswanathan and A. L. Lane, *ACS Synth. Biol.*, 2016, **5**, 547–553.
118 Y. Tsunematsu, N. Ishikawa, D. Wakana, Y. Goda, H. Noguchi, H. Moriya, K. Hotta and K. Watanabe, *Nat. Chem. Biol.*, 2013, **9**, 818–825.
119 K. Blin, S. Shaw, K. Steinke, R. Villebro, N. Ziemert, S. Y. Lee, M. H. Medema and T. Weber, *Nucleic Acids Res.*, 2019, 1–7.
120 B. V. B. Reddy, A. Milshteyn, Z. Charlop-Powers and S. F. Brady, *Chem. Biol.*, 2014, **21**, 1023–1033.
121 N. Ziemert, S. Podell, K. Penn, J. H. Badger, E. Allen and P. R. Jensen, *PLoS One*, 2012, **7**, e34064.
122 G. J. Dodge, F. P. Maloney and J. L. Smith, *Nat. Prod. Rep.*, 2018, **35**, 1082–1096.
123 M. Klaus and M. Grininger, *Nat. Prod. Rep.*, 2018, **35**, 1070–1081.
124 T. Robbins, Y.-C. Liu, D. E. Cane and C. Khosla, *Curr. Opin. Struct. Biol.*, 2016, **41**, 10–18.
125 Y. Li, W. Zhang, H. Zhang, W. Tian, L. Wu, S. Wang, M. Zheng, J. Zhang, C. Sun, Z. Deng, Y. Sun, X. Qu and J. Zhou, *Angew. Chem., Int. Ed.*, 2018, **57**, 5823–5827.
126 H. S. T. Bukhari, R. P. Jakob and T. Maier, *Structure*, 2014, **22**, 1775–1785.
127 A. Rittner, K. S. Paithankar, D. J. Drexler, A. Himmler and M. Grininger, *Protein Sci.*, 2019, **28**, 414–428.
128 A. M. Gulick, *Curr. Opin. Chem. Biol.*, 2016, **35**, 89–96.
129 M. Klaus, A. D. D'Souza, A. Nivina, C. Khosla and M. Grininger, *ACS Chem. Biol.*, 2019, **14**, 426–433.
130 M. Klaus, M. P. Ostrowski, J. Austerjost, T. Robbins, B. Lowry, D. E. Cane and C. Khosla, *J. Biol. Chem.*, 2016, **291**, 16404–16415.
131 B. R. Miller, E. J. Drake, C. Shi, C. C. Aldrich and A. M. Gulick, *J. Biol. Chem.*, 2016, **291**, 22559–22571.
132 K. J. Esquilín-Lebrón, T. O. Boynton, L. J. Shimkets and M. G. Thomas, *J. Bacteriol.*, 2018, **200**, 1–13.
133 S. Mori, K. D. Green, R. Choi, G. W. Buchko, M. G. Fried and S. Garneau-Tsodikova, *ChemBioChem*, 2018, **19**, 2186–2194.
134 B. R. Miller, J. A. Sundlov, E. J. Drake, T. A. Makin and A. M. Gulick, *Proteins: Struct., Funct., Bioinf.*, 2014, **82**, 2691–2702.
135 P. Cruz-Morales, J. F. Kopp, C. Martínez-Guerrero, L. A. Yáñez-Guerra, N. Selem-Mojica, H. Ramos-Aboites, J. Feldmann and F. Barona-Gómez, *Genome Biol. Evol.*, 2016, **8**, 1906–1916.
136 A. Wlodek, S. G. Kendrew, N. J. Coates, A. Hold, J. Pogwizd, S. Rudder, L. S. Sheehan, S. J. Higginbotham, A. E. Stanley-Smith, T. Warneck, M. Nur-E-Alam, M. Radzom, C. J. Martin, L. Overvoorde, M. Samborskyy, S. Alt, D. Heine, G. T. Carter, E. I. Graziani, F. E. Koehn, L. McDonald, A. Alanine, R. M. Rodríguez Sarmiento, S. K. Chao, H. Ratni, L. Steward, I. H. Norville, M. Sarkar-Tyson, S. J. Moss, P. F. Leadlay, B. Wilkinson and M. A. Gregory, *Nat. Commun.*, 2017, **8**(1), 1206.
137 E. Kalkreuter, J. M. CroweTipton, A. N. Lowell, D. H. Sherman and G. J. Williams, *J. Am. Chem. Soc.*, 2019, **141**, 1961–1969.
138 T. Awakawa, T. Fujioka, L. Zhang, S. Hoshino, Z. Hu, J. Hashimoto, I. Kozone, H. Ikeda, K. Shin-Ya, W. Liu and I. Abe, *Nat. Commun.*, 2018, **9**, 3534.
139 E. J. Caldera, M. G. Chevrette, B. R. McDonald and C. R. Currie, *Appl. Environ. Microbiol.*, 2019, **85**(24), e01580-19.
140 Y. Shimizu, H. Ogata and S. Goto, *ChemBioChem*, 2017, **18**, 50–65.
141 M. N. Ngaki, G. V. Louie, R. N. Philippe, G. Manning, F. Pojer, M. E. Bowman, L. Li, E. Larsen, E. S. Wurtele and J. P. Noel, *Nature*, 2012, **485**, 530–533.
142 Y. Yamada, T. Kuzuyama, M. Komatsu, K. Shin-ya, S. Omura, D. E. Cane and H. Ikeda, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 857–862.
143 F. Chen, D. Tholl, J. Bohlmann and E. Pichersky, *Plant J.*, 2011, **66**, 212–229.
144 D. Tholl, *Curr. Opin. Plant Biol.*, 2006, **9**, 297–304.
145 C. Schmidt-Dannert, *Biosynthesis of Terpenoid Natural Products in Fungi*, 2014, pp. 19–61.
146 F. Beran, P. Rahfeld, K. Luck, R. Nagel, H. Vogel, N. Wielsch, S. Irmisch, S. Ramasamy, J. Gershenzon, D. G. Heckel and T. G. Köllner, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 2922–2927.
147 X. Chen, T. G. Köllner, Q. Jia, A. Norris, B. Santhanam, P. Rabe, J. S. Dickschat, G. Shaulsky, J. Gershenzon and F. Chen, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 12132–12137.
148 J. S. Dickschat, *Nat. Prod. Rep.*, 2016, **33**, 87–110.
149 M. T. Miller, B. O. Bachmann, C. A. Townsend and A. C. Rosenzweig, *Nat. Struct. Biol.*, 2001, **8**, 684–689.
150 M. T. Miller, B. O. Bachmann, C. A. Townsend and A. C. Rosenzweig, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 14752–14757.
151 K. Tahlan and S. E. Jensen, *J. Antibiot.*, 2013, **66**, 401–410.
152 S. L. Robinson, J. K. Christenson and L. P. Wackett, *Nat. Prod. Rep.*, 2019, **36**, 458–475.
153 F. Kudo and T. Eguchi, *J. Antibiot.*, 2009, **62**, 471–481.

154 A. Sánchez-Rodríguez, H. L. Tytgat, J. Winderickx, J. Vanderleyden, S. Lebeer and K. Marchal, *BMC Genomics*, 2014, **15**, 349.
155 M. H. Medema and M. a. Fischbach, *Nat. Chem. Biol.*, 2015, **11**, 639–648.
156 K.-S. Ju, J. R. Doroghazi and W. W. Metcalf, *J. Ind. Microbiol. Biotechnol.*, 2014, **41**, 345–356.
157 X. Yu, J. R. Doroghazi, S. C. Janga, J. K. Zhang, B. Circello, B. M. Griffin, D. P. Labeda and W. W. Metcalf, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 20759–20764.
158 F. Barona-Gómez, P. Cruz-Morales and L. Noda-García, *Antonie van Leeuwenhoek*, 2012, **101**, 35–43.
159 N. Sélem-Mojica, C. Aguilar, K. Gutiérrez-García, C. E. Martínez-Guerrero and F. Barona-Gómez, *Microb. Genomics*, 2019, 1–16.
160 J. K. Schniete, P. Cruz-Morales, N. Selem-Mojica, L. T. Fernández-Martínez, I. S. Hunter, F. Barona-Gómez and P. A. Hoskisson, *mBio*, 2018, **9**, e02283-17.
161 M. Alanjary, B. Kronmiller, M. Adamek, K. Blin, T. Weber, D. Huson, B. Philmus and N. Ziemert, *Nucleic Acids Res.*, 2017, **45**, W42–W48.
162 M. Knopp, J. S. Gudmundsdottir, T. Nilsson, F. König, O. Warsi, F. Rajer, P. Ädelroth and D. I. Andersson, *mBio*, 2019, **10**, 1–15.
163 D. A. Hopwood, *Heredity*, 2019, **123**, 23–32.
164 D. A. Hodgson, in *Advances in Microbial Physiology*, 2000, vol. 42, pp. 47–238.
165 P. A. Hoskisson and L. T. Fernández-Martínez, *Environ. Microbiol. Rep.*, 2018, **10**, 231–238.
166 Y. A. Chan, A. M. Podevels, B. M. Kevany and M. G. Thomas, *Nat. Prod. Rep.*, 2009, **26**, 90–114.
167 B. A. Pfeifer and C. Khosla, *Chem. Eng. Sci.*, 2001, **65**, 106–118.
168 G. Zhang, Y. Li, L. Fang and B. A. Pfeifer, *Sci. Adv.*, 2015, **1**(4), e1500077.
169 A. L. Juárez-Vázquez, J. N. Edirisinghe, E. A. Verduzco-Castro, K. Michalska, C. Wu, L. Noda-García, G. Babnigg, M. Endres, S. Medina-Ruíz, J. Santoyo-Flores, M. Carrillo-Tripp, H. Ton-That, A. Joachimiak, C. S. Henry and F. Barona-Gómez, *eLife*, 2017, **6**, 1–21.
170 F. Barona-Gómez and D. A. Hodgson, *J. Mol. Microbiol. Biotechnol.*, 2010, **19**, 152–158.
171 Z. Hojati, C. Milne, B. Harvey, L. Gordon, M. Borg, F. Flett, B. Wilkinson, P. J. Sidebottom, B. A. Rudd, M. A. Hayes, C. P. Smith and J. Micklefield, *Chem. Biol.*, 2002, **9**, 1175–1187.
172 J. Davies and D. Davies, *Microbiol. Mol. Biol. Rev.*, 2010, **74**, 417–433.
173 M. N. Thaker, N. Waglechner and G. D. Wright, *Nat. Protoc.*, 2014, **9**, 1469–1479.
174 K. C. Freel, N. Millán-Aguiñaga and P. R. Jensen, *Appl. Environ. Microbiol.*, 2013, **79**, 5997–6005.
175 A. G. McArthur, N. Waglechner, F. Nizam, A. Yan, M. A. Azad, A. J. Baylay, K. Bhullar, M. J. Canova, G. De Pascale, L. Ejim, L. Kalan, A. M. King, K. Koteva, M. Morar, M. R. Mulvey, J. S. O'Brien, A. C. Pawlowski, L. J. V. Piddock, P. Spanogiannopoulos, A. D. Sutherland, I. Tang, P. L. Taylor, M. Thaker, W. Wang, M. Yan, T. Yu and G. D. Wright, *Antimicrob. Agents Chemother.*, 2013, **57**, 3348–3357.
176 X. Tang, J. Li, N. Millán-Aguiñaga, J. J. Zhang, E. C. O'Neill, J. a. Ugalde, P. R. Jensen, S. M. Mantovani and B. S. Moore, *ACS Chem. Biol.*, 2015, **10**(12), 2841–2849.
177 R. M. Peterson, T. Huang, J. D. Rudolf, M. J. Smanski and B. Shen, *Chem. Biol.*, 2014, **21**, 389–397.
178 M. Steffensky, A. Mühlenweg, Z. X. Wang, S. M. Li and L. Heide, *Antimicrob. Agents Chemother.*, 2000, **44**, 1214–1222.
179 A. Kling, P. Lukat, D. V. Almeida, A. Bauer, E. Fontaine, S. Sordello, N. Zaburannyi, J. Herrmann, S. C. Wenzel, C. Konig, N. C. Ammerman, M. B. Barrio, K. Borchers, F. Bordon-Pallier, M. Bronstrup, G. Courtemanche, M. Gerlitz, M. Geslin, P. Hammann, D. W. Heinz, H. Hoffmann, S. Klieber, M. Kohlmann, M. Kurz, C. Lair, H. Matter, E. Nuermberger, S. Tyagi, L. Fraisse, J. H. Grosset, S. Lagrange and R. Muller, *Science*, 2015, **348**, 1106–1112.
180 G. Vigliotta, S. M. Tredici, F. Damiano, M. R. Montinaro, R. Pulimeno, R. Di Summa, D. R. Massardo, G. V. Gnoni and P. Alifano, *Mol. Microbiol.*, 2005, **55**, 396–412.
181 K. J. Laiple, T. Härtner, H. P. Fiedler, W. Wohlleben and T. Weber, *J. Antibiot.*, 2009, **62**, 465–468.
182 J. Ishikawa, A. Yamashita, Y. Mikami, Y. Hoshino, H. Kurita, K. Hotta, T. Shiba and M. Hattori, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 14925–14930.
183 K. Kovács, L. D. Hurst and B. Papp, *PLoS Biol.*, 2009, **7**, 1–9.
184 P. Cimermancic, M. H. Medema, J. Claesen, K. Kurita, L. C. Wieland Brown, K. Mavrommatis, A. Pati, P. A. Godfrey, M. Koehrsen, J. Clardy, B. W. Birren, E. Takano, A. Sali, R. G. Linington and M. A. Fischbach, *Cell*, 2014, **158**, 412–421.
185 F. Del Carratore, K. Zych, M. Cummings, E. Takano, M. H. Medema and R. Breitling, *Commun. Biol.*, 2019, **2**, 83.
186 M. Alanjary, K. Steinke and N. Ziemert, *Nucleic Acids Res.*, 2019, 1–7.
187 R. S. Gokhale, S. Y. Tsuji, D. E. Cane and C. Khosla, *Science*, 1999, **284**, 482–485.
188 R. F. Seipke, *PLoS One*, 2015, **10**, e0116457.
189 C. Vicente, A. Thibessard, J.-N. Lorenzi, M. Benhadj, L. Hôtel, D. Gacemi-Kirane, O. Lespinet, P. Leblond and B. Aigle, *Antibiotics*, 2018, **7**, 86.
190 C. J. Park and C. P. Andam, *Front. Microbiol.*, 2019, **10**, 1–12.
191 P. Cruz-Morales, E. Vijgenboom, F. Iruegas-Bocardo, G. Girard, L. A. Yáñez-Guerra, H. E. Ramos-Aboites, J. L. Pernodet, J. Anné, G. P. Van Wezel and F. Barona-Gómez, *Genome Biol. Evol.*, 2013, **5**, 1165–1175.
192 T. K. S. Richter, C. C. Hughes and B. S. Moore, *Environ. Microbiol.*, 2015, **17**, 2158–2171.
193 J. Lopera, I. J. Miller, K. L. McPhail and J. C. Kwan, *mSystems*, 2017, **2**(6), e00096-17.
194 B. M. Kevany, D. a. Rasko and M. G. Thomas, *Appl. Environ. Microbiol.*, 2009, **75**, 1144–1155.

195 N. A. Moss, G. Seiler, T. F. Leão, G. Castro-Falcón, L. Gerwick, C. C. Hughes and W. H. Gerwick, *Angew. Chem., Int. Ed.*, 2019, 1–6.
196 L. Chao and S. A. K. Jongkees, *Angew. Chem.*, 2019, **131**(37), 12880–12890.
197 C. Wu, M. H. Medema, R. M. Läkamp, L. Zhang, P. C. Dorrestein, Y. H. Choi and G. P. Van Wezel, *ACS Chem. Biol.*, 2016, **11**, 478–490.
198 M. Adamek, M. Alanjary, H. Sales-Ortells, M. Goodfellow, A. T. Bull, A. Winkler, D. Wibberg, J. Kalinowski and N. Ziemert, *BMC Genomics*, 2018, **19**, 426.
199 K. Gutiérrez-García, A. Neira-González, R. M. Pérez-Gutiérrez, G. Granados-Ramírez, R. Zarraga, K. Wrobel, F. Barona-Gómez and L. B. Flores-Cotera, *J. Nat. Prod.*, 2017, **80**, 1955–1963.
200 I. Nouioui, L. Carro, M. García-López, J. P. Meier-Kolthoff, T. Woyke, N. C. Kyrpides, R. Pukall, H. P. Klenk, M. Goodfellow and M. Göker, *Front. Microbiol.*, 2018, **9**, 1–119.
201 N. J. Tobias and H. B. Bode, *J. Mol. Biol.*, 2019, DOI: 10.1016/j.jmb.2019.04.042.
202 R. A. McClure, A. W. Goering, K.-S. Ju, J. A. Baccile, F. C. Schroeder, W. W. Metcalf, R. J. Thomson and N. L. Kelleher, *ACS Chem. Biol.*, 2016, **11**, 3452–3460.
203 J. P. Griffith, J. L. Kim, E. E. Kim, M. D. Sintchak, J. A. Thomson, M. J. Fitzgibbon, M. A. Fleming, P. R. Caron, K. Hsiao and M. A. Navia, *Cell*, 1995, **82**, 507–522.
204 J. Choi, J. Chen, S. L. Schreiber and J. Clardy, *Science*, 1996, **273**, 239–242.
205 T. Schwecke, J. F. Aparicio, I. Molnár, A. König, L. E. Khaw, S. F. Haydock, M. Oliynyk, P. Caffrey, J. Cortés and J. B. Lester, *Proc. Natl. Acad. Sci. U. S. A.*, 1995, **92**, 7839–7843.
206 G. Gatto and M. Boyne, *J. Am. Chem. Soc.*, 2006, **12**, 3838–3847.
207 J. N. Andexer, S. G. Kendrew, M. Nur-e-Alam, O. Lazos, T. A. Foster, A.-S. Zimmermann, T. D. Warneck, D. Suthar, N. J. Coates, F. E. Koehn, J. S. Skotnicki, G. T. Carter, M. A. Gregory, C. J. Martin, S. J. Moss, P. F. Leadlay and B. Wilkinson, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 4776–4781.
208 J. Zucko, J. Cullum, D. Hranueli and P. F. Long, *J. Antibiot.*, 2011, **64**, 89–92.
209 P. Borgman, R. D. Lopez and A. L. Lane, *Org. Biomol. Chem.*, 2019, **17**, 2305–2314.
210 M. H. Medema, R. Kottmann, P. Yilmaz, M. Cummings, J. B. Biggins, K. Blin, I. de Bruijn, Y. H. Chooi, J. Claesen, R. C. Coates, P. Cruz-Morales, S. Duddela, S. Düsterhus, D. J. Edwards, D. P. Fewer, N. Garg, C. Geiger, J. P. Gomez-Escribano, A. Greule, M. Hadjithomas, A. S. Haines, E. J. N. Helfrich, M. L. Hillwig, K. Ishida, A. C. Jones, C. S. Jones, K. Jungmann, C. Kegler, H. U. Kim, P. Kötter, D. Krug, J. Masschelein, A. V. Melnik, S. M. Mantovani, E. a. Monroe, M. Moore, N. Moss, H.-W. Nützmann, G. Pan, A. Pati, D. Petras, F. J. Reen, F. Rosconi, Z. Rui, Z. Tian, N. J. Tobias, Y. Tsunematsu, P. Wiemann, E. Wyckoff, X. Yan, G. Yim, F. Yu, Y. Xie, B. Aigle, A. K. Apel, C. J. Balibar, E. P. Balskus, F. Barona-Gómez, A. Bechthold, H. B. Bode, R. Borriss, S. F. Brady, A. a. Brakhage, P. Caffrey, Y.-Q. Cheng, J. Clardy, R. J. Cox, R. De Mot, S. Donadio, M. S. Donia, W. a. van der Donk, P. C. Dorrestein, S. Doyle, A. J. M. Driessen, M. Ehling-Schulz, K.-D. Entian, M. a. Fischbach, L. Gerwick, W. H. Gerwick, H. Gross, B. Gust, C. Hertweck, M. Höfte, S. E. Jensen, J. Ju, L. Katz, L. Kaysser, J. L. Klassen, N. P. Keller, J. Kormanec, O. P. Kuipers, T. Kuzuyama, N. C. Kyrpides, H.-J. Kwon, S. Lautru, R. Lavigne, C. Y. Lee, B. Linquan, X. Liu, W. Liu, A. Luzhetskyy, T. Mahmud, Y. Mast, C. Méndez, M. Metsä-Ketelä, J. Micklefield, D. a. Mitchell, B. S. Moore, L. M. Moreira, R. Müller, B. a. Neilan, M. Nett, J. Nielsen, F. O'Gara, H. Oikawa, A. Osbourn, M. S. Osburne, B. Ostash, S. M. Payne, J.-L. Pernodet, M. Petricek, J. Piel, O. Ploux, J. M. Raaijmakers, J. a. Salas, E. K. Schmitt, B. Scott, R. F. Seipke, B. Shen, D. H. Sherman, K. Sivonen, M. J. Smanski, M. Sosio, E. Stegmann, R. D. Süssmuth, K. Tahlan, C. M. Thomas, Y. Tang, A. W. Truman, M. Viaud, J. D. Walton, C. T. Walsh, T. Weber, G. P. van Wezel, B. Wilkinson, J. M. Willey, W. Wohlleben, G. D. Wright, N. Ziemert, C. Zhang, S. B. Zotchev, R. Breitling, E. Takano and F. O. Glöckner, *Nat. Chem. Biol.*, 2015, **11**, 625–631.
211 N. Guttenberger, W. Blankenfeldt and R. Breinbauer, *Bioorg. Med. Chem.*, 2017, **25**, 6149–6166.
212 Q. Yan, B. Philmus, J. H. Chang and J. E. Loper, *eLife*, 2017, **6**, 1–24.
213 Q. Yan, J. H. Chang, J. E. Loper, L. D. Lopes, F. D. Andreote, B. T. Shaffer, T. A. Kidarsa, V. O. Stockwell, O. Vining, B. Philmus, K. L. McPhail, C. Song and J. M. Raaijmakers, *mBio*, 2018, **9**, 1–18.
214 K. Tahlan, H. U. Park and S. E. Jensen, *Can. J. Microbiol.*, 2004, **50**, 803–810.
215 S. E. Jensen, *J. Ind. Microbiol. Biotechnol.*, 2012, **39**, 1407–1419.
216 P. Liras, *Antonie van Leeuwenhoek*, 1999, **75**, 109–124.
217 R. B. Hamed, J. R. Gomez-Castellanos, L. Henry, C. Ducho, M. a. McDonough and C. J. Schofield, *Nat. Prod. Rep.*, 2013, **30**, 21–107.
218 M. A. Fischbach, *Curr. Opin. Microbiol.*, 2009, **12**, 520–527.
219 N. F. AbuSara, B. M. Piercey, M. A. Moore, A. A. Shaikh, L. Nothias, S. K. Srivastava, P. Cruz-Morales, P. C. Dorrestein, F. Barona-Gómez and K. Tahlan, *Front. Microbiol.*, 2019, **10**, 1–17.
220 S. K. Srivastava, K. S. King, N. F. AbuSara, C. J. Malayny, B. M. Piercey, J. A. Wilson and K. Tahlan, *PLoS One*, 2019, **14**, e0215960.
221 S. W. Elson, J. Gillett, N. H. Nicholson and J. W. Tyler, *J. Chem. Soc., Chem. Commun.*, 1988, 979.
222 D. C. Alexander and S. E. Jensen, *J. Bacteriol.*, 1998, **180**, 4068–4079.
223 L. A. Egan, R. W. Busby, D. Iwata-Reuyl and C. A. Townsend, *J. Am. Chem. Soc.*, 1997, **119**, 2348–2355.

224 R. H. Mosher, A. S. Paradkar, C. Anders, B. Barton and S. E. Jensen, *Antimicrob. Agents Chemother.*, 1999, **43**, 1215–1224.
225 K. Tahlan, C. Anders, A. Wong, R. H. Mosher, P. H. Beatty, M. J. Brumlik, A. Griffin, C. Hughes, J. Griffin, B. Barton and S. E. Jensen, *Chem. Biol.*, 2007, **14**, 131–142.
226 P. Liras, J. P. Gomez-Escribano and I. Santamarta, *J. Ind. Microbiol. Biotechnol.*, 2008, **35**, 667–676.
227 N. J. Zelyas, H. Cai, T. Kwong and S. E. Jensen, *J. Bacteriol.*, 2008, **190**, 7957–7965.
228 D. M. Troppens, J. A. Moynihan, M. Barret, F. O'Gara and J. P. Morrissey, in *Molecular Microbial Ecology of the Rhizosphere*, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2013, pp. 593–605.
229 V. K. Veena, R. N. Popavath, K. Kennedy and N. Sakthivel, *Apoptosis*, 2015, **20**, 1281–1295.
230 D. V. Mavrodi, O. V. Mavrodi, J. A. Parejko, D. M. Weller and L. S. Thomashow, in *Bacteria in Agrobiology: Plant Nutrient Management*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 267–283.
231 A. Ramette, M. Frapolli, M. F.-L. Saux, C. Gruffaz, J.-M. Meyer, G. Défago, L. Sutra and Y. Moënne-Loccoz, *Syst. Appl. Microbiol.*, 2011, **34**, 180–188.
232 O. V. Mavrodi, B. B. McSpadden Gardener, D. V. Mavrodi, R. F. Bonsall, D. M. Weller and L. S. Thomashow, *Phytopathology*, 2001, **91**, 35–43.
233 F. Barona-Gómez, U. Wong, A. E. Giannakopulos, P. J. Derrick and G. L. Challis, *J. Am. Chem. Soc.*, 2004, **126**, 16282–16283.
234 F. Barona-Gómez, S. Lautru, F. Francois-Xavier, P. Leblond, J. L. Pernodet and G. L. Challis, *Microbiology*, 2006, **152**, 3355–3366.
235 H. Ramos-Aboites, A. Yáñez-Olvera and F. Barona-Gómez, in *Ecosystem Ecology and Geochemistry of Cuatro Cienegas*, 2018, pp. 123–140.
236 V. H. Tierrafría, H. E. Ramos-Aboites, G. Gosset and F. Barona-Gómez, *Microb. Biotechnol.*, 2011, **4**, 275–285.
237 N. Kadi, D. Oves-Costales, F. Barona-Gomez and G. L. Challis, *Nat. Chem. Biol.*, 2007, **3**, 652–656.
238 S. Tunca, C. Barreiro, A. Sola-Landa, J. J. R. Coque and J. F. Martín, *FEBS J.*, 2007, **274**, 1110–1122.
239 S. Tunca, C. Barreiro, J.-J. R. Coque and J. F. Martín, *FEBS J.*, 2009, **276**, 4814–4827.
240 J. L. Ronan, N. Kadi, S. A. McMahon, J. H. Naismith, L. M. Alkhalaf and G. L. Challis, *Philos. Trans. R. Soc., B*, 2018, **373**, 20170068.
241 A. A. Roberts, A. W. Schultz, R. D. Kersten, P. C. Dorrestein and B. S. Moore, *FEMS Microbiol. Lett.*, 2012, **335**, 95–103.
242 P. Huang, F. Xie, B. Ren, Q. Wang, J. Wang, Q. Wang, W. M. Abdel-Mageed, M. Liu, J. Han, A. Oyeleye, J. Shen, F. Song, H. Dai, X. Liu and L. Zhang, *Appl. Microbiol. Biotechnol.*, 2016, **100**, 7437–7447.
243 N. Ejje, C. Z. Soe, J. Gu and R. Codd, *Metallomics*, 2013, **5**, 1519.
244 S. A. Essén, A. Johnsson, D. Bylund, K. Pedersen and U. S. Lundström, *Appl. Environ. Microbiol.*, 2007, **73**, 5857–5864.
245 T. H. M. Smits and B. Duffy, *Arch. Microbiol.*, 2011, **193**, 693–699.
246 A. Sen, V. Daubin, D. Abrouk, I. Gifford, A. M. Berry and P. Normand, *Int. J. Syst. Evol. Microbiol.*, 2014, **64**, 3821–3832.
247 R. Codd, T. Richardson-Sanchez, T. J. Telfer and M. P. Gotsbacher, *ACS Chem. Biol.*, 2018, **13**, 11–25.
248 T. J. Telfer and R. Codd, *ACS Chem. Biol.*, 2018, **13**, 2456–2471.
249 J. Meiwes, H. P. Fiedler, H. Zähner, S. Konetschny-Rapp and G. Jung, *Appl. Microbiol. Biotechnol.*, 1990, **32**, 505–510.
250 A. M. P. de Almeida, A. Guiyoule, I. Guilvout, I. Iteman, G. Baranton and E. Carniel, *Microb. Pathog.*, 1993, **14**, 9–21.
251 C. Pelludat, A. Rakin, C. A. Jacobi, S. Schubert and J. Heesemann, *J. Bacteriol.*, 1998, **180**, 538–546.
252 K. Yamanaka, H. Oikawa, H. Ogawa, K. Hosono, F. Shinmachi, H. Takano, S. Sakuda, T. Beppu and K. Ueda, *Microbiology*, 2005, **151**, 2899–2905.
253 A. D'Onofrio, J. M. Crawford, E. J. Stewart, K. Witt, E. Gavrish, S. Epstein, J. Clardy and K. Lewis, *Chem. Biol.*, 2010, **17**, 254–264.
254 O. Sekurova, I. Pérez-Victoria, J. Martín, K. Degnes, H. Sletta, F. Reyes and S. Zotchev, *Molecules*, 2016, **21**, 1131.
255 C. H. R. Senges, A. Al-Dilaimi, D. H. Marchbank, D. Wibberg, A. Winkler, B. Haltli, M. Nowrousian, J. Kalinowski, R. G. Kerr and J. E. Bandow, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, 2490–2495.
256 M. F. Traxler, M. R. Seyedsayamdost, J. Clardy and R. Kolter, *Mol. Microbiol.*, 2012, **86**, 628–644.
257 A. A. Arias, S. Lambert, L. Martinet, D. Adam, E. Tenconi, M.-P. Hayette, M. Ongena and S. Rigali, *FEMS Microbiol. Ecol.*, 2015, **91**, fiv080.
258 J. Galet, A. Deveau, L. Hôtel, P. Frey-Klett, P. Leblond and B. Aigle, *Appl. Environ. Microbiol.*, 2015, **81**, 3132–3141.
259 S. Rütschlin, S. Gunesch and T. Böttcher, *Cell Chem. Biol.*, 2017, **24**, 598–604.
260 S. Rütschlin, S. Gunesch and T. Böttcher, *ACS Chem. Biol.*, 2018, **13**, 1153–1158.
261 S. Lambert, M. F. Traxler, M. Craig, M. Maciejewska, M. Ongena, G. P. van Wezel, R. Kolter and S. Rigali, *Metallomics*, 2014, **6**, 1390–1399.
262 P. Patel, L. Song and G. L. Challis, *Biochemistry*, 2010, **49**, 8033–8042.
263 B. R. Levin, *Philos. Trans. R. Soc., B*, 1988, **319**, 459–472.
264 A. C. Ruzzini and J. Clardy, *Curr. Biol.*, 2016, **26**, R859–R864.
265 L. G. M. Baas Becking, *Geobiologie of inleiding tot de milieukunde*, W. P. Van Stockum & Zoon, Den Haag, 1934.
266 Z. Charlop-Powers, J. G. Owen, B. V. B. Reddy, M. A. Ternei, D. O. Guimarães, U. A. de Frias, M. T. Pupo, P. Seepe, Z. Feng and S. F. Brady, *eLife*, 2015, **4**, e05048.
267 B. R. Mcdonald, M. G. Chevrette, J. L. Klassen, H. A. Horn, E. J. Caldera, E. Wendt-Pienkowski, M. J. Cafaro, A. C. Ruzzini, E. B. Van Arnam, G. M. Weinstock, N. M. Gerardo, M. Poulsen, G. Suen, J. Clardy and C. R. Currie, bioRxiv, 2019, DOI: 10.1101/545640.
268 P. A. Hoskisson, P. Sumby and M. C. M. Smith, *Virology*, 2015, **477**, 100–109.

269 P. A. Hoskisson and M. C. Smith, *Curr. Opin. Microbiol.*, 2007, **10**, 396–400.
270 B. J. Shapiro, J. B. Leducq and J. Mallet, *PLoS Genet.*, 2016, **12**, 1–14.
271 P. E. Galand, E. O. Casamayor, D. L. Kirchman and C. Lovejoy, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 22427–22432.
272 M. G. Chevrette, C. M. Carlson, H. E. Ortega, C. Thomas, G. E. Ananiev, K. J. Barns, A. J. Book, J. Cagnazzo, C. Carlos, W. Flanigan, K. J. Grubbs, H. A. Horn, F. M. Hoffmann, J. L. Klassen, J. J. Knack, G. R. Lewin, B. R. McDonald, L. Muller, W. G. P. Melo, A. A. Pinto-Tomás, A. Schmitz, E. Wendt-Pienkowski, S. Wildman, M. Zhao, F. Zhang, T. S. Bugni, D. R. Andes, M. T. Pupo and C. R. Currie, *Nat. Commun.*, 2019, **10**, 516.
273 G. R. Lewin, C. Carlos, M. G. Chevrette, H. A. Horn, B. R. McDonald, R. J. Stankey, B. G. Fox and C. R. Currie, *Annu. Rev. Microbiol.*, 2016, **70**, 235–254.
274 K. Vetsigian, R. Jajoo and R. Kishony, *PLoS Biol.*, 2011, **9**, e1001184.
275 M. Vos, A. B. Wolf, S. J. Jennings and G. A. Kowalchuk, *FEMS Microbiol. Rev.*, 2013, **37**, 936–954.
276 R. H. Baltz, *J. Ind. Microbiol. Biotechnol.*, 2017, **44**, 573–588.
277 J. R. Doroghazi, J. C. Albright, A. W. Goering, K.-S. Ju, R. R. Haines, K. a. Tchalukov, D. P. Labeda, N. L. Kelleher and W. W. Metcalf, *Nat. Chem. Biol.*, 2014, **10**, 963–968.
278 G. L. Lozano, J. I. Bravo, M. F. Garavito Diago, H. B. Park, A. Hurley, S. B. Peterson, E. V. Stabb, J. M. Crawford, N. A. Broderick and J. Handelsman, *mBio*, 2019, **10**, 1–14.
279 P. D. Straight and R. Kolter, *Annu. Rev. Microbiol.*, 2009, **63**, 99–118.
280 L. C. Clark, R. F. Seipke, P. Prieto, J. Willemse, G. P. Van Wezel, M. I. Hutchings and P. A. Hoskisson, *Sci. Rep.*, 2013, **3**, 3–10.
281 A. Buckling, F. Harrison, M. Vos, M. A. Brockhurst, A. Gardner, S. A. West and A. Griffin, *FEMS Microbiol. Ecol.*, 2007, **62**, 135–141.
282 Z. Bódi, Z. Farkas, D. Nevozhay, D. Kalapis, V. Lázár, B. Csörgő, Á. Nyerges, B. Szamecz, G. Fekete, B. Papp, H. Araújo, J. L. Oliveira, G. Moura, M. A. S. Santos, T. Székely, G. Balázsi and C. Pál, *PLoS Biol.*, 2017, **15**, 1–26.
283 B. J. Shapiro and M. F. Polz, *Trends Microbiol.*, 2014, **22**, 235–247.
284 E. Pichersky and E. Lewinsohn, *Annu. Rev. Plant Biol.*, 2011, **62**, 549–566.
285 E. Breukink and B. de Kruijff, *Nat. Rev. Drug Discovery*, 2006, **5**, 321–332.
286 P. Sarkar, V. Yarlagadda, C. Ghosh and J. Haldar, *MedChemComm*, 2017, **8**, 516–533.
287 S. Y. Kim, K.-S. Ju, W. W. Metcalf, B. S. Evans, T. Kuzuyama and W. A. van der Donk, *Antimicrob. Agents Chemother.*, 2012, **56**, 4175–4183.
288 C. S. Sit, A. C. Ruzzini, E. B. Van Arnam, T. R. Ramadhar, C. R. Currie and J. Clardy, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, 201515348.
289 P. Marttinen and W. P. Hanage, *PLoS Comput. Biol.*, 2017, **13**, 1–15.
290 A. M. Dean and J. W. Thornton, *Nat. Rev. Genet.*, 2007, **8**, 675–688.
291 J. Davies and K. S. Ryan, *ACS Chem. Biol.*, 2012, **7**, 252–259.
292 J. Davies, *J. Antibiot.*, 2013, **66**, 361–364.
293 L. Chae, T. Kim, R. Nilo-Poyanco and S. Y. Rhee, *Science*, 2014, **344**, 510–513.
294 A. Rokas, J. H. Wisecaver and A. L. Lind, *Nat. Rev. Microbiol.*, 2018, **16**, 731–744.
295 G. I. Guzmán, T. E. Sandberg, R. A. LaCroix, Á. Nyerges, H. Papp, M. de Raad, Z. A. King, Y. Hefner, T. R. Northen, R. A. Notebaart, C. Pál, B. O. Palsson, B. Papp and A. M. Feist, *Mol. Syst. Biol.*, 2019, **15**, e8462.
296 J. Rosenberg and F. M. Commichau, *Trends Biotechnol*, 2019, **37**, 29–37.
297 L. Noda-García, D. Davidi, E. Korenblum, A. Elazar, E. Putintseva, A. Aharoni and D. S. Tawfik, *Nat. Microbiol.*, 2019, **4**, 1221–1230.
298 J. Zan, Z. Li, M. D. Tianero, J. Davis, R. T. Hill and M. S. Donia, *Science*, 2019, **364**, eaaw6732.
299 J. C. Kwan, M. S. Donia, A. W. Han, E. Hirose, M. G. Haygood and E. W. Schmidt, *Proc. Natl. Acad. Sci.*, 2012, **109**, 20655–20660.
300 G. L. Lozano, H. B. Park, J. I. Bravo, E. A. Armstrong, J. M. Denu, E. V. Stabb, N. A. Broderick, J. M. Crawford and J. Handelsman, *Appl. Environ. Microbiol.*, 2019, **85**, 1–13.
301 N. Adnani, M. G. Chevrette, S. N. Adibhatla, F. Zhang, Q. Yu, D. R. Braun, J. Nelson, S. W. Simpkins, B. R. McDonald, C. L. Myers, J. S. Piotrowski, C. J. Thompson, C. R. Currie, L. Li, S. R. Rajski and T. S. Bugni, *ACS Chem. Biol.*, 2017, **12**, 3093–3102.
302 M. G. Chevrette, J. R. Bratburd, C. R. Currie and R. M. Stubbendieck, *mSystems*, 2019, **4**, 1–8.
303 H. B. Bode, B. Bethe, R. Höfs and A. Zeeck, *ChemBioChem*, 2002, **3**, 619.
304 L. Martinet, A. Naômé, B. Deflandre, M. Maciejewska, D. Tellatin, E. Tenconi, N. Smargiasso, E. de Pauw, G. P. van Wezel and S. Rigali, *mBio*, 2019, **10**, 1–15.
305 R. Pan, X. Bai, J. Chen, H. Zhang and H. Wang, *Front. Microbiol.*, 2019, **10**, 1–20.
306 H. L. Robertsen, E. M. Musiol-Kroll, L. Ding, K. J. Laiple, T. Hofeditz, W. Wohlleben, S. Y. Lee, S. Grond and T. Weber, *Sci. Rep.*, 2018, **8**, 3230.
307 M. N. Thaker, W. Wang, P. Spanogiannopoulos, N. Waglechner, A. M. King, R. Medina and G. D. Wright, *Nat. Biotechnol.*, 2013, **31**, 922–927.
308 D. M. McCandlish, *Evolution*, 2011, **65**, 1544–1558.
309 D. L. Des Marais and M. D. Rausher, *Nature*, 2008, **454**, 762–765.
310 T. Sikosek, H. S. Chan and E. Bornberg-Bauer, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 14888–14893.
311 C. H. Eng, T. W. H. Backman, C. B. Bailey, C. Magnan, H. García Martín, L. Katz, P. Baldi and J. D. Keasling, *Nucleic Acids Res.*, 2018, **46**, D509–D515.
312 N. Ziemert, M. Alanjary and T. Weber, *Nat. Prod. Rep.*, 2016, **33**, 988–1005.

313 M. Alanjary, C. Cano-Prieto, H. Gross and M. M. Medema, *Nat. Prod. Rep.*, 2019, **36**(9), 1249–1261.

314 M. J. Smanski, H. Zhou, J. Claesen, B. Shen, M. A. Fischbach and C. A. Voigt, *Nat. Rev. Microbiol.*, 2016, **14**, 135–149.

315 P. Calero and P. I. Nikel, *Microb. Biotechnol.*, 2019, **12**, 98–124.

316 C. W. Johnston, A. D. Connaty, M. a. Skinnider, Y. Li, A. Grunwald, M. a. Wyatt, R. G. Kerr and N. a. Magarvey, *J. Ind. Microbiol. Biotechnol.*, 2016, **43**(2–3), 293–298.

317 M. Wang, J. J. Carver, V. V. Phelan, L. M. Sanchez, N. Garg, Y. Peng, D. D. Nguyen, J. Watrous, C. A. Kapono, T. Luzzatto-Knaan, C. Porto, A. Bouslimani, A. V. Melnik, M. J. Meehan, W.-T. Liu, M. Crüsemann, P. D. Boudreau, E. Esquenazi, M. Sandoval-Calderón, R. D. Kersten, L. A. Pace, R. A. Quinn, K. R. Duncan, C.-C. Hsu, D. J. Floros, R. G. Gavilan, K. Kleigrewe, T. Northen, R. J. Dutton, D. Parrot, E. E. Carlson, B. Aigle, C. F. Michelsen, L. Jelsbak, C. Sohlenkamp, P. Pevzner, A. Edlund, J. McLean, J. Piel, B. T. Murphy, L. Gerwick, C.-C. Liaw, Y.-L. Yang, H.-U. Humpf, M. Maansson, R. A. Keyzers, A. C. Sims, A. R. Johnson, A. M. Sidebottom, B. E. Sedio, A. Klitgaard, C. B. Larson, C. A. P. Boya, D. Torres-Mendoza, D. J. Gonzalez, D. B. Silva, L. M. Marques, D. P. Demarque, E. Pociute, E. C. O'Neill, E. Briand, E. J. N. Helfrich, E. A. Granatosky, E. Glukhov, F. Ryffel, H. Houson, H. Mohimani, J. J. Kharbush, Y. Zeng, J. A. Vorholt, K. L. Kurita, P. Charusanti, K. L. McPhail, K. F. Nielsen, L. Vuong, M. Elfeki, M. F. Traxler, N. Engene, N. Koyama, O. B. Vining, R. Baric, R. R. Silva, S. J. Mascuch, S. Tomasi, S. Jenkins, V. Macherla, T. Hoffman, V. Agarwal, P. G. Williams, J. Dai, R. Neupane, J. Gurr, A. M. C. Rodríguez, A. Lamsa, C. Zhang, K. Dorrestein, B. M. Duggan, J. Almaliti, P.-M. Allard, P. Phapale, L.-F. Nothias, T. Alexandrov, M. Litaudon, J.-L. Wolfender, J. E. Kyle, T. O. Metz, T. Peryea, D.-T. Nguyen, D. VanLeer, P. Shinn, A. Jadhav, R. Müller, K. M. Waters, W. Shi, X. Liu, L. Zhang, R. Knight, P. R. Jensen, B. Ø. Palsson, K. Pogliano, R. G. Linington, M. Gutiérrez, N. P. Lopes, W. H. Gerwick, B. S. Moore, P. C. Dorrestein and N. Bandeira, *Nat. Biotechnol.*, 2016, **34**, 828–837.

318 A. Gurevich, A. Mikheenko, A. Shlemov, A. Korobeynikov, H. Mohimani and P. A. Pevzner, *Nat. Microbiol.*, 2018, **3**, 319–327.

319 K. J. Hetrick and W. A. van der Donk, *Curr. Opin. Chem. Biol.*, 2017, **38**, 36–44.

320 J. I. Tietz, C. J. Schwalen, P. S. Patel, T. Maxson, P. M. Blair, H.-C. Tai, U. I. Zakai and D. A. Mitchell, *Nat. Chem. Biol.*, 2017, **13**, 470–478.

321 H. Hasegawa and L. Holm, *Curr. Opin. Struct. Biol.*, 2009, **19**, 341–348.

