

Mining and unearthing hidden biosynthetic potential

Kirstin Scherlach¹ & Christian Hertweck^{1,2}  

Genetically encoded small molecules (secondary metabolites) play eminent roles in ecological interactions, as pathogenicity factors and as drug leads. Yet, these chemical mediators often evade detection, and the discovery of novel entities is hampered by low production and high rediscovery rates. These limitations may be addressed by genome mining for biosynthetic gene clusters, thereby unveiling cryptic metabolic potential. The development of sophisticated data mining methods and genetic and analytical tools has enabled the discovery of an impressive array of previously overlooked natural products. This review shows the newest developments in the field, highlighting compound discovery from unconventional sources and microbiomes.

Natural products are an unparalleled source of bioactive compounds, many of which have found application in medicine or agriculture or are important drivers of organismal interactions^{1,2}. Traditionally, these compounds were isolated from microbes and plants by bioactivity-guided approaches; however, conventional strategies now fail to cover the constant demand for new chemical entities³. Advances in genomics, bioinformatics, and chemical analytics have paved the way for modern genomics-based discovery approaches.

While natural products are chemically extremely diverse, their biosynthetic machineries are often highly conserved. Core biosynthetic enzymes are characterized by high amino-acid-sequence similarity, which allows screening of genomic data for the presence of specific biosynthetic genes that encode the required enzymatic activity. Eminent examples of genetically programmed molecular assembly lines include the major classes of natural products such as polyketides⁴ and nonribosomally synthesized peptides (NRPs)⁵, ribosomally synthesized and posttranslationally modified peptides (RiPPs)⁶, alkaloids⁷, and terpenes⁸.

Analyses of genome sequences indicate that the biosynthetic potential of bacteria, fungi, and even higher organisms is much larger than what is observed under laboratory conditions. This may either be due to a strong downregulation of the biosynthetic genes or due to low production yields, which prevent detection of the compounds by analytical approaches⁹. To provide access to these compounds, specific triggers or stimuli are required to activate silent or downregulated gene clusters and to increase the compound production rates¹⁰.

Today, vast genomic data are available and much progress has been made in data mining, compound monitoring, single-cell techniques, and genetic approaches for pathway activation, providing ideal conditions to access the cryptic metabolome.

This review highlights recent advances (since 2015) in the genomics-guided discovery of secondary metabolites focusing on unconventional natural product sources and ecology-inspired discovery approaches.

¹Department of Biomolecular Chemistry, Leibniz Institute for Natural Product Research and Infection Biology, HKI, Jena, Germany. ²Faculty of Biological Sciences, Friedrich Schiller University Jena, Jena, Germany. ✉email: christian.hertweck@leibniz-hki.de

Genome mining tools and strategies

The advance of modern sequencing technologies has led to huge amounts of genomic sequence data revealing a tremendous reservoir of likely bioactive natural products that wait to be discovered. As most of the encoded chemical diversity still remains untapped, novel tools and strategies are required to access these potential drug candidates or chemical mediators. Major bottlenecks in realizing this potential include the recognition and prioritization of interesting biosynthetic genes, their activation, and, finally, establishing the link between the genes and the encoded secondary metabolites. To address these challenges, a number of strategies were developed to make use of the acquired genomic information (Fig. 1).

Automated bioinformatics. The exponential growth of genomic sequencing data has propelled the development of bioinformatic tools to analyze these valuable data. On the basis of our current understanding of the biosynthetic logic, algorithms were created that allow the prediction of natural product biosynthetic assembly lines and putatively encoded structures from gene sequences. Overviews on available automated software tools were provided in a number of comprehensive reviews^{11–16}. Most predictive tools rely on homologies to already characterized pathways. Thus, the output is likely biased toward common biosynthetic principles and may fail to detect novel pathways. To overcome this limitation, machine learning-based approaches and deep learning strategies were developed that show an improved ability to identify biosynthetic gene clusters (BGCs) of novel classes^{17–21}.

These computational tools proved to be invaluable to mine the huge amount of available genomic information and allow even the nonexpert user to automatically analyze genetic data as a starting point for further experimental investigations.

Structure prediction and chemical synthesis. Once a BGC of interest has been detected, the next challenge is the link to the corresponding natural product. If the microorganisms are recalcitrant to cultivation or the biosynthetic genes remain silent under laboratory conditions, isolation from large-scale bacterial fermentations is not feasible. As an alternative, a culture-independent approach to access cryptic metabolites was developed based on bioinformatics prediction of chemical scaffolds followed by chemical synthesis of the desired compounds (Fig. 1). This approach bypasses time-consuming activation and isolation procedures and may yield novel chemicals; however, it needs to be taken into account that the predicted structures may only resemble the original metabolite, as post-assembly-line modifications cannot be accurately predicted. Using this method, a number of synthetic-bioinformatic natural products (syn-BNPs) were discovered, including the peptide humimycin (1) with potent antimethicillin-resistant *Staphylococcus aureus* (MRSA) activity²², the antibiotic paenimucillins (e.g., paenimucillin A, 2)^{23,24}, and an antifungal peptide²³. Similar methodologies were applied to identify novel RiPPs with antibacterial activity²⁵ and a new class of RiPPs, the pyritides²⁶. Based on the architecture of a BGC in *Micromonospora rosaria*, the corresponding natural products were predicted to undergo a formal, enzymatic [4 + 2]-cycloaddition with subsequent elimination of the leader peptide and water to produce a pyridine-based macrocycle (pyritide A2, 3)²⁶. Chemical synthesis of the predicted structures and chemo-enzymatic reconstitution of the pathway confirmed the validity of the hypothesis and demonstrated the combined power of bioinformatics and chemical synthesis for investigating cryptic gene clusters. Nonetheless, the application of this strategy is currently limited to compound classes for which accurate prediction algorithms are established (e.g., NRPs and certain RiPPs). For the

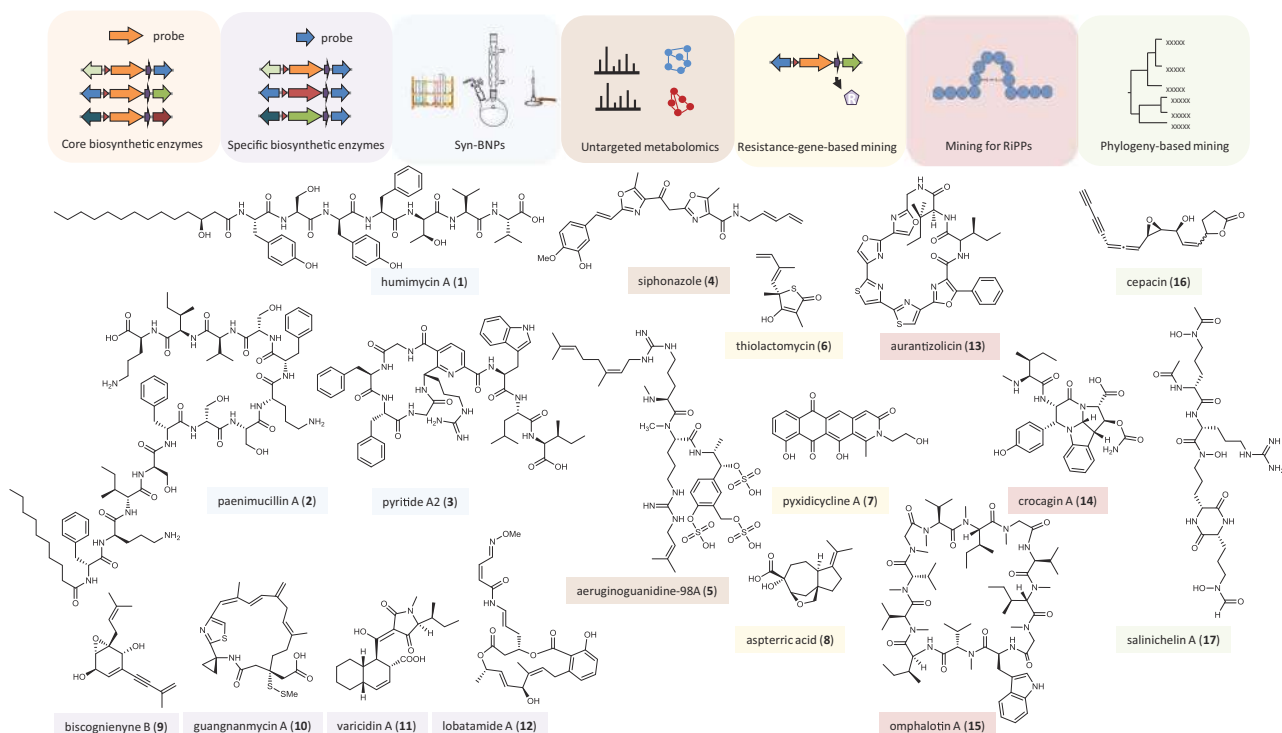


Fig. 1 Genome mining strategies. Tools and strategies for mining genomic data and examples of natural products discovered through these methods. (syn-BNP synthetic-bioinformatic natural products, RiPPs ribosomally synthesized and posttranslationally modified peptides). The color code of the boxes around the compound names and icons indicates the strategy that was applied for the discovery of the respective compound: light orange—mining for core biosynthetic enzymes, light violet—mining for specific biosynthetic enzymes, light blue—bioinformatic structure prediction and chemical synthesis, brown—metabolomic approaches to link genes to metabolites and prioritize cryptic BGCs, yellow—resistance-genes-based genome mining, light red—genome mining for RiPPs, green—phylogeny-based genome mining.

majority of the cryptic BGCs, bioinformatics software is currently not able to predict the exact compound structures. Combined genomic and metabolomic approaches that employ large-scale mass spectrometry-based comparative metabolomics to predict modifications may subsequently overcome these limitations^{27–29}.

Linking genes to metabolites and prioritizing cryptic BGCs.

Various compounds were discovered by traditional approaches like bioactivity-guided isolation in the past, but the molecular bases of their biosynthesis or congeners with other activity profiles remained unknown. Software tools like rBAN³⁰ that simulate the retro-biosynthesis of NRPs from their chemical structure can predict the required enzymatic machinery and also help to prioritize promising BGCs for novel compound discovery. In addition, a number of mass spectrometry-guided genome mining approaches were developed that combine genomics and untargeted metabolomics to assign detected secondary metabolites to orphan BGCs and to prioritize strains^{29,31–33} (Fig. 1). An early application of pattern-based genome mining integrating the analysis of BGCs and molecular networking involved the investigation of a large collection of environmental *Salinispora* isolates, which uncovered a huge metabolic diversity among the strains and led to the characterization of novel compounds³². Siphonazole (4) is an antiplasmodial natural product isolated from a *Herpetosiphon* species. Its biosynthesis has remained elusive for nearly a decade. Through a combination of genome mining, imaging mass spectrometry, and expression studies in the natural producer, the BGC was discovered, revealing that siphonazole originates from a mixed polyketide synthase/nonribosomal peptide synthetase (PKS/NRPS) pathway³⁴. Using a similar approach, the cyanobacterial compound aeruginoguanidine (5) was linked to a cryptic NRPS gene cluster³⁵.

A crucial aspect of genomics-based natural product discovery is the prioritization of the most promising BGCs among the huge number of detected genetic loci. Bioinformatics tools that allow to group related genes by sequence similarity networks³⁶, genome neighborhood networks^{36,37}, or BGC family³⁸ may assist in identifying a specific biosynthetic background. Combining the genomic datasets with automated MS-based metabolomics analysis helps to prioritize novel compounds for structure elucidation³². Additionally, target-based genome mining strategies (see below) may accomplish the discovery of natural products with biological/pharmacological potential^{39,40}.

Specialized mining strategies. In most cases, genome mining approaches target core biosynthetic genes of molecular assembly lines.

With the aim to specifically search for compounds with defined bioactivity or with novel structures or even new scaffolds, alternative strategies were established that target, for example, genes encoding resistance information or tailoring enzymes. In addition, a number of phylogeny-guided approaches have been pursued.

Resistance genes-based mining. One strategy to specifically search for antibiotic natural products is to mine microbial genomes for resistance genes (Fig. 1). Bacteria have evolved several strategies to avoid the self-toxicity of their antibiotics, including enzyme-catalyzed antibiotic modifications, bypass of antibiotic targets, and active efflux of drugs from the cell⁴¹. The required resistance genes are often co-localized with the genes encoding the biosynthetic machinery for antibiotic production and can thus serve as a guide to discover putative antibiotics⁴². Mining the genomes of 86 *Salinispora* strains for putative target-modifying resistance genes associated with natural product biosynthetic genes led to the prioritization of an orphan PKS-NRPS hybrid gene cluster

harboring a putative fatty acid synthase resistance gene as a candidate for targeted antibiotic discovery. Heterologous expression of the gene cluster in a *Streptomyces* host led to the identification of a group of thiotetronic acid natural products, including the previously known fatty acid synthase inhibitor thiolactomycin (6)⁴³. Even though the chemical structure of this compound had been long known, this work revealed the molecular basis of thiolactomycin biosynthesis for the first time and demonstrated the feasibility of such an approach. Guided by the presence of genes coding for pentapeptide repeat proteins known for conferring resistance to topoisomerase inhibitors, in the genome of the myxobacterium *Pyxidicoccus fallax*, a cryptic PKS gene cluster was targeted. Its activation in the native host as well as its heterologous expression enabled the structure elucidation of pyxidicyclines (e.g., pyxidicyclin A, 7)⁴⁴. A similar strategy was successfully applied for the targeted discovery of a novel bioactive compound from filamentous fungi. With the aim to discovering a potential herbicide, published fungal genomes were scanned for genes coding for dihydroxyacid dehydratase (DHAD) that are co-localized with core biosynthetic enzymes. DHAD is an essential enzyme in the indispensable branched-chain amino acid biosynthetic pathway in plants and a common target for herbicides. A homolog of a DHAD encoding gene was identified in the vicinity to genes encoding a sesquiterpene cyclase homolog and two cytochrome P450s in *Aspergillus terreus*. Since this set of genes was highly conserved among a number of fungal genomes, it was hypothesized that it might code for a natural product with DHAD inhibitory activity. Heterologous expression of this gene cluster in *Saccharomyces cerevisiae* and subsequent compound isolation and characterization revealed aspterric acid (8) as the encoded natural product and confirmed its DHAD inhibitory activity⁴⁵. These examples demonstrate that genes conferring self-resistance can serve as an indicator of biosynthetic machinery encoding putative antibiotics or toxins. Automated tools to connect genomic and structural information with resistance determinants of known antibiotics will further support resistance-based mining efforts^{40,46–48}.

Mining for genes encoding specific biosynthetic enzymes (other than canonical PKS and NRPS). Whereas the majority of genome mining approaches target core biosynthetic enzymes such as canonical PKSs or NRPSs, also genes encoding tailoring enzymes or unusual modules in biosynthetic assembly lines proved to be promising alternatives for mining efforts (Fig. 1). For example, scanning genomic data for sequences of bacterial acetylenases uncovered the biosynthetic machineries for secondary metabolites bearing terminal alkyne moieties⁴⁹. The characterization of the biosynthetic pathway of the acetylenic meroterpenoid biscognine B (9) allows further genome mining endeavors for the discovery of new compounds with acetylenic prenyl chains⁵⁰.

Using the DUF-SH didomain, responsible for sulfur incorporation in the leinamycin biosynthetic pathway, as a probe to mine for leinamycin analogs, a variety of potential producers of this compound class were discovered (e.g., guannganmycin A, 10)⁵¹. Similar approaches were chosen to discover novel fungal secondary metabolites. Diels-Alderases are a class of enzymes that catalyze pericyclic reactions of a conjugated diene to a dienophile in analogy to a Diels-Alder reaction known from synthetic chemistry. Genes encoding putative Diels-Alderases can be found in various biosynthetic pathways; however, most of the encoded metabolites have remained elusive. Upon mining for genes coding for such enzymes, the BGC for varicidin A (11) in *Penicillium variable* was discovered and the corresponding natural product identified. Varicidin A is a new antifungal natural product containing a *cis*-octahydrodecalin core, biosynthesized by a Diels-Alderase⁵².

Mining genomes for genes encoding noncanonical PKS homologs led to the identification of an architecturally unique *trans*-AT PKS gene cluster in a *Methylobacterium* strain. The gene locus eluded automated prediction due to its unusual and highly fragmented nature. Yet, orthologous clusters could be detected in related species, suggesting that the gene cluster is functional. Comparative screening of culture extracts of a deletion mutant and the wild-type strain uncovered novel polyketides with rare epoxide and cyclopropyl moieties⁵³. Similarly, mining for genes encoding oxygenase-containing modules in *trans*-AT PKS systems led to the discovery of the BGC for lobatamide A (**12**) in the culturable plant symbiont *Gyvuella sunshinyi*⁵⁴.

Mining for ribosomally synthesized peptides. RiPPs are a structurally diverse group of natural products with a wide spectrum of biological activities. Their biosynthesis proceeds via ribosomally assembled precursor peptides that undergo posttranslational modification to gain their biological function¹⁶. A number of bioinformatics tools were developed to detect the biosynthetic prerequisites in microbial genomes initially relying on core biosynthetic enzymes⁵⁵ (Fig. 1). Later on, class-independent RiPP genome mining tools utilizing alternative probes such as the RiPP recognition elements (RRE) were established^{16,25,55–58}. Whereas potential producers of RiPPs can thus be identified through comparative genome mining, additional methods are required to actually find the corresponding metabolite. Software tools that integrate genomic and metabolomic data may additionally support the identification of novel RiPPs^{27,28,58}. For example, bioinformatics prediction using RiPP-PRISM in tandem with automated LC-MS/MS searches led to the identification of aurantizolicin (**13**)⁵⁹. Through a combination of data mining and analytical chemistry, crocagins (e.g., crocagin A, **14**) were discovered from the myxobacterium *Chondromyces crocatus* that form a new class of RiPPs⁶⁰. Polytheonamides are the only characterized members of a unique family of RiPPs termed proteusins (named after the Greek sea god Proteus constantly changing his shape), based on an unusually large leader peptide with homology to nitrile hydratases. These marine sponge-derived peptides are chemically distinct from any other known natural product. Their ribosomal precursor peptide undergoes 49 mostly noncanonical posttranslational modifications, which results in a highly cytotoxic natural product. As the original producer of these hypermodified peptides cannot be cultivated, an alternative producing platform was required. Data mining revealed that closely related pathways are present in taxonomically and ecologically remarkably diverse organisms, including culturable bacteria. Using one candidate species as a host, a platform was established that allows the production of highly modified polytheonamide-like peptides with cytostatic properties⁶¹.

While major progress has been made in understanding the biosynthesis of RiPPs in bacteria, only little is known about the formation of ribosomal peptides in fungi. One example of a RiPP from a fungus is omphalotin (**15**), a cyclopeptide with multiple *N*-methylations. Mining the genome of *Omphalotus olearius* for genes encoding a precursor peptide resulted in the identification of a novel biosynthesis mechanism for a RiPP. An iterative *N*-methyltransferase fused to its peptide substrate catalyzes the auto-methylation of its C-terminus^{62,63}. Due to this unusual mechanism, the term “borosins” was proposed for this novel RiPP family, referring to the ancient mythological symbol Ouroboros depicting a serpent biting its own tail⁶³. Later on, additional members of the borosin class were discovered⁶⁴.

Phylogeny-based mining. Combining classical genome mining with evolutionary aspects can further support bioprospecting and

may also facilitate functional predictions of biosynthetic genes^{38,65} (Fig. 1). The phylogenies of natural product-producing organisms can be applied to infer talented producers. For example, many members of the genus *Burkholderia* are known to produce a high number of antimicrobial agents and are therefore regarded as potential biocontrol organisms. However, at the same time, *Burkholderia* species are also known to infect humans, which hampers their potential application for biocontrol purposes. Phylogeny-led genome mining in combination with chemical and biological profiling revealed the efficacy of *Burkholderia ambifaria* as a biopesticide. Biosynthesis of the acetylenic antibiotic cepacin (**16**) was shown to be responsible for the pesticidal activity. Deletion of a nonessential plasmid associated with virulence resulted in a less infectious mutant with retained pesticidal activity⁶⁶.

Additionally, studying the evolutionary history of secondary metabolite gene clusters by phylogeny-based methods can also expedite the discovery of novel molecules. Using a strategy called EvoMining, which is based on the assumption that most enzymes from secondary metabolism evolved from primary metabolism, the evolutionary history of 23 enzyme families was reviewed, which led to the discovery of arseno-organic metabolites in *Streptomyces* species⁶⁷. Through the reconstruction of the evolutionary history of two different siderophore families, it was shown that certain *Salinispora* strains have functionally replaced an ancient desferrioxamine pathway and acquired the genetic accessories for the biosynthesis of the novel siderophore salinichelin (**17**)⁶⁸.

Accessing silent biosynthetic genes

The finely tuned regulation of secondary metabolism poses a huge challenge to natural product researchers to identify conditions under which biosynthetic genes are expressed. In many cases, biosynthesis is downregulated, and the encoded structures escape detection. Therefore, efforts are required to induce the expression of silent genes and to link chemical structures to orphan biosynthesis gene clusters (Fig. 2 and 3).

Triggering natural product biosynthesis. The production of secondary metabolites by microorganisms critically depends on the cultivation conditions. Often specific triggers (e.g., small molecules) are required to elicit the expression of the biosynthetic genes (Fig. 2). A systematic variation of culture conditions and/or the application of stress conditions can be an initial approach to propagate the formation of the chemical compounds¹⁰. This rather empirical method is likely impractical for a large number of different microbes. The ecological background of the potential producers may inspire alternative approaches. Based on the hypothesis that natural products serve as chemical mediators of microbial interaction², several co-culturing methods were developed with the aim that one organism induces the formation of silent metabolites in the other. For example, co-culturing of the marine invertebrate-associated bacteria *Micromonospora* sp. and *Rhodococcus* sp. afforded the polynitroglycosylated anthracycline keyicin (**18**) which is hypothesized to play a role in microbial communication⁶⁹. Likewise, *Streptomyces rapamycinicus* was shown to induce the formation of fumigermin (**19**) in *Aspergillus fumigatus*. Fumigermin resembles bacterial germination inhibitors such as germicidin and, ultimately, its inhibitory activity on spore germination of *S. rapamycinicus* was demonstrated. This study represents one of the rare examples where it could be demonstrated that a compound whose production was elicited in a mixed culture plays a role in the interaction of the co-cultured partners⁷⁰.

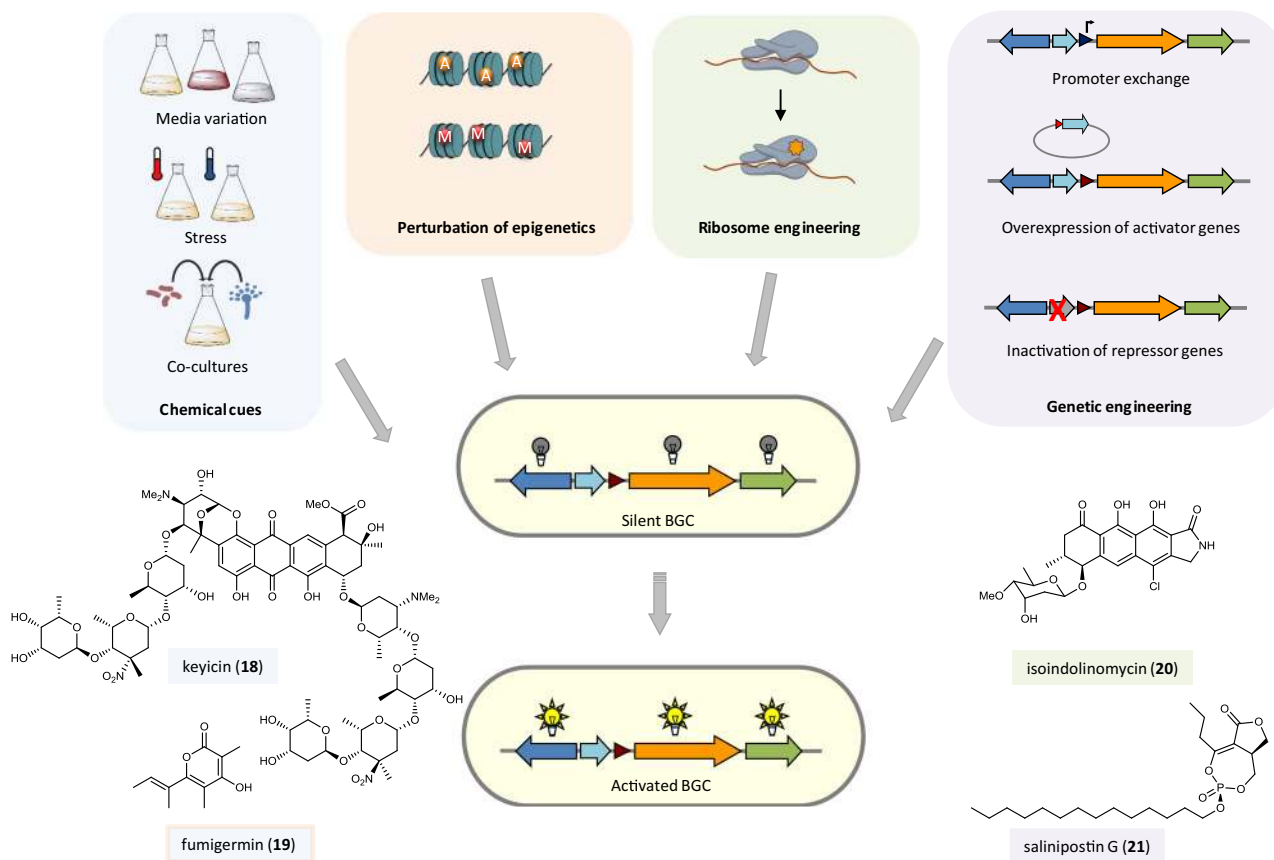


Fig. 2 Accessing cryptic biosynthetic genes in native hosts. Strategies for the activation of silent biosynthetic gene clusters in native hosts and examples of natural products discovered through these methods. (BGC biosynthetic gene cluster). The color code of the boxes around the compound names and icons indicates the strategy that was applied for the discovery of the respective compound: light blue—triggering natural product biosynthesis through chemical cues, light orange—activation of natural product biosynthesis through perturbation of epigenetic regulation, green—activation of natural product biosynthesis through ribosome engineering, light violet—activation of natural product biosynthesis through genetic engineering.

Although microbial co-cultivation has proven successful in inducing secondary metabolite biosynthesis, it is still an arbitrary approach that suffers from low predictability and difficult up-scaling possibilities to increase throughput. Moreover, in the majority of cases, the nature of the elicitor remains unknown. To overcome some of these obstacles, a strategy was developed to screen more systematically for inducers of specific silent biosynthetic pathways (HiTES = high-throughput elicitor screens)⁷¹. This method is based on the assumption that microbes employ small natural products for communication, which may function as elicitors of silent biosynthetic genes. A reporter gene is inserted inside the gene cluster of interest, and the resulting mutant is screened against libraries of secondary metabolites in a high-throughput fashion to find potential inducers. Using this methodology, silent gene clusters in *Burkholderia thailandensis* and *Streptomyces albus* were induced in a targeted fashion and several antibiotic and cytotoxic compounds were identified as potential elicitors of other cryptic biosynthetic pathways^{71,72}. In another study, screening of activation conditions was performed with a reporter-guided mutant selection strategy after genome-scale random mutagenesis⁷³.

Activation of silent pathways through ribosome engineering.

Ribosome engineering is based on the isolation of spontaneously developed drug-resistant mutants. Through the application of the antibiotics streptomycin or rifampicin, strains with mutations in the *rpsL* gene (encoding the ribosomal protein S12) or *rpoB* gene (encoding the RNA polymerase (RNAP) β -subunit) are selected.

Such mutants may show an altered gene expression, which may result in a different metabolite profile. Initially developed in streptomycetes⁷⁴, the method has now been applied to randomly activate silent biosynthetic genes in various strains^{75,76} (Fig. 2). A recent example is the discovery of the polyketide isoindolinomycin through the screening of rifampicin-resistant mutants⁷⁷.

Genetic approaches to activate silent pathways. While some silent BGCs can be activated in the native host after identification of the appropriate cues, some others require genetic manipulation to induce gene expression, for example, the overexpression of regulatory genes or the introduction of promoters (Fig. 2). In a recent example, comparative transcriptomics was used to identify key regulatory genes of silent pathways. Comparing the expression profiles of similar gene clusters in different strains helped to prioritize producer strains and led to the identification of a series of novel compounds (e.g., salinipostin G, 21)⁷⁸. A CRISPR-Cas9-based promoter knock-in strategy was applied to activate multiple BGCs in *Streptomyces* species⁷⁹. Whereas methods to engineer microbial pathways are well established for model organisms, the targeted genetic manipulation of nonmodel strains may be highly challenging. To facilitate in situ promoter insertion in strains less amenable to genetic manipulation, such as environmental *Burkholderia* isolates, a strategy was reported that involves novel bacteriophage recombinases (Red $\alpha\beta$ homologs). The recombinase genes were cloned for transient expression, and optimized for the efficient deletion of chromosomal DNA. The presented

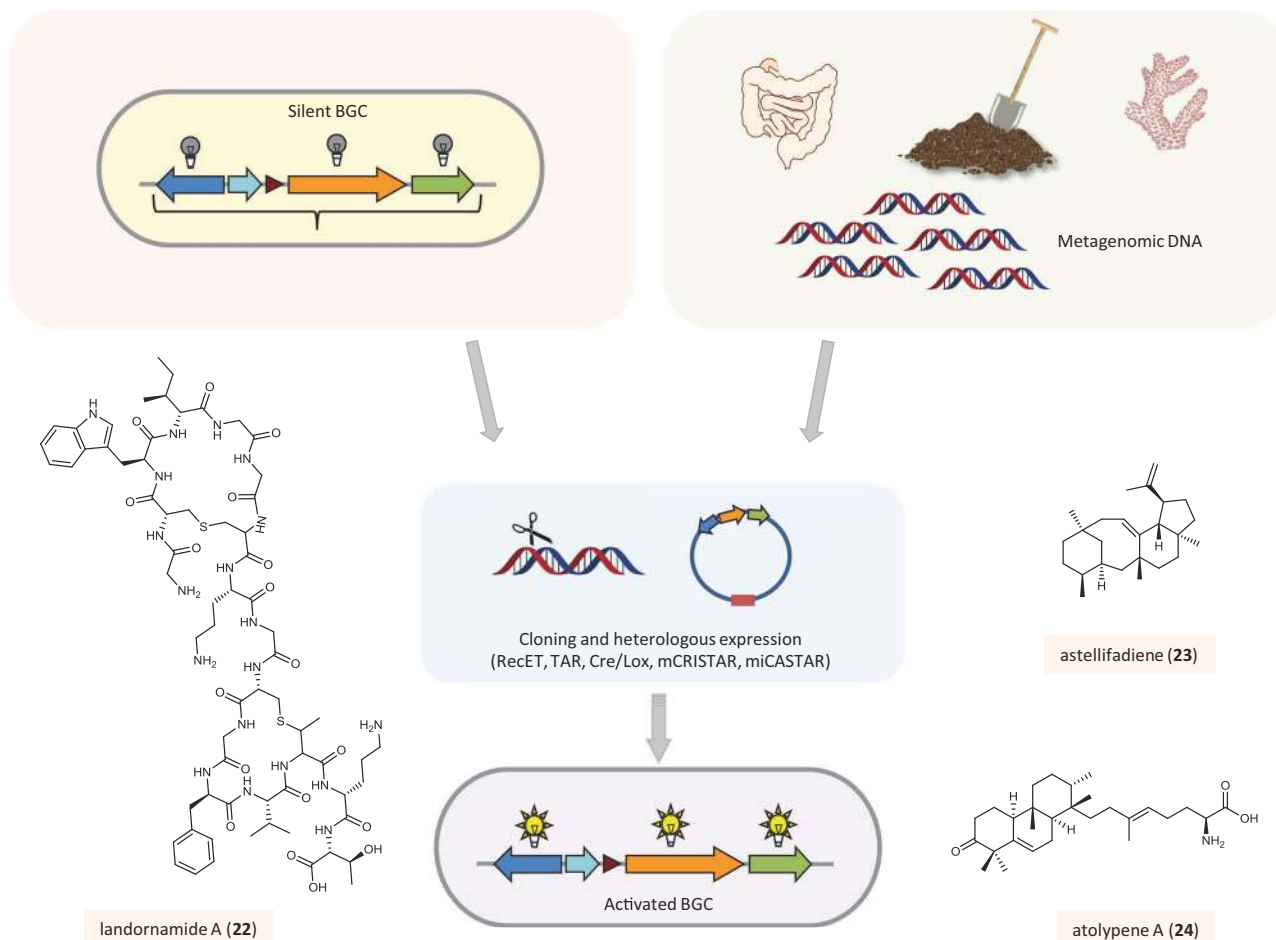


Fig. 3 Accessing cryptic biosynthetic genes in heterologous hosts. Strategies for the activation of silent biosynthetic gene clusters in heterologous hosts and examples of natural products discovered through these methods. (BGC biosynthetic gene cluster). The color code of the boxes around the compound names and icons indicates the strategy that was applied for the discovery of the respective compound: Activation of natural product biosynthesis through heterologous expression of biosynthetic genes that are downregulated in the native producer (light orange) or from metagenomic DNA (green).

workflow allows targeted gene deletions or promoter knock-ins in various Burkholderiales that lack native Red $\alpha\beta$ recombinase homologs⁸⁰.

In eukaryotes, many secondary metabolite gene clusters are located in heterochromatic regions of the genome. Gene transcription is controlled by epigenetic regulation such as histone deacetylation or DNA methylation. Manipulation of epigenetic control may lead to activation of silent biosynthetic genes⁸¹. Deletion of a histone H3 deacetylase resulted in the pleiotropic activation and overexpression of more than 75% of the biosynthetic genes of the endophytic fungus *Calcarisporium arbuscular*⁸². Modification of the chromatin landscape may also be a strategy that bacteria apply to influence gene transcription in fungi. In *A. nidulans*, changes in histone acetylation were monitored upon co-cultivation with *S. rapamycinicus* and related to changes in the fungal transcriptome⁸³.

In strains that are less amenable to genetic manipulation or cannot be cultivated, functional expression of full gene clusters in heterologous hosts may be a feasible alternative (Fig. 3). The prerequisite is the direct capture of the gene cluster by homologous recombination using, for example, the Lambda Red/ET recombination, yeast-based transformation-associated recombination system, or alternative techniques^{84–86}. These technologies together with optimized heterologous hosts were successfully applied for the discovery of novel natural products^{86,87}.

Reconstruction of a silent RiPP pathway in *E. coli* led to the identification of a new antiviral peptide, landornamide A (22), revealing at the same time various biosynthetic novelties⁸⁸. Mining the genome of *Emericella varicolor* NBRC 32302 for prenyl transferase and terpene cyclase encoding genes and functional expression of the identified terpene synthesis genes in *Aspergillus oryzae* afforded the sesterterpene astellifadiene (23)⁸⁹.

The introduction of the CRISPR/Cas9 system also offered new possibilities for the refactoring of BGCs⁸⁷. For example, a yeast-based promoter engineering platform was developed that combines CRISPR/Cas9 and transformation-associated recombination to enable single-marker multiplexed promoter engineering of large gene clusters (mCRISTAR)⁹⁰. The target gene cluster is cut at the native promoter regions using CRISPR/Cas9 to allow the TAR-mediated reassembly to incorporate synthetic promoters. Further development of this strategy resulted in a simplified method (miCASTAR) that allows the targeted activation of BGCs, as was demonstrated by the discovery of the sesterterpene atolypene (24)⁹¹.

Genome-guided discovery of natural products from unconventional sources

As the search for new natural products from traditional bacterial sources like actinomycetes often results in a high rediscovery rate of known compounds, alternative sources are being explored. The high-throughput sequencing of microorganisms from diverse

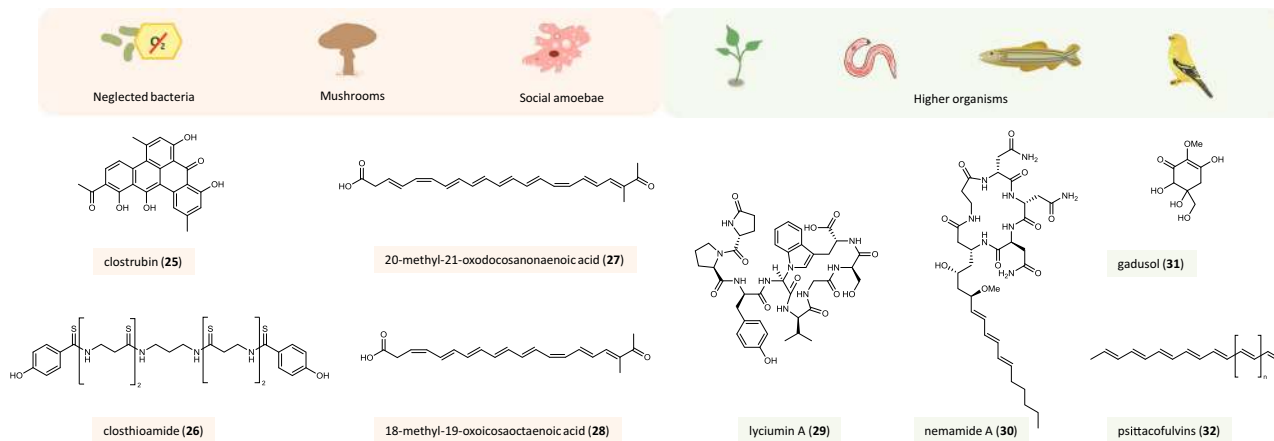


Fig. 4 Genome-guided discovery of natural products from unconventional sources. Neglected bacteria, mushrooms, social amoebae, and higher organisms as sources and examples of natural products discovered from these organisms. The color code of the boxes around the compound names and icons indicates the source of the respective compound: light orange—microbes (neglected bacteria, mushrooms, and amoebae), green—higher organisms (plants, nematodes, fishes, birds).

habitats and genome sequences of higher organisms revealed a huge number of “talented producers” from so far underexplored genera (Fig. 4).

Genome-guided discovery of natural products from neglected bacteria and archaea. Anaerobic bacteria have long been neglected as a potential source of bioactive secondary metabolites. Genomic analyses indicated that these organisms harbor a huge biosynthetic potential that remains to be discovered⁹². Only recently, a handful of secondary metabolites could be identified through combined genomic and analytical approaches^{93–97} (Fig. 4). Mining the genome of *Clostridium puniceum*, an anaerobic plant pathogen causing potato rot, revealed a gene locus coding for the biosynthesis of the pentacyclic polyketides clostrubins (e.g., clostrubin A, **25**) that are essential for the anaerobic bacteria to grow under aerobic/oxic conditions in association with plants. Moreover, clostrubins were found to possess strong antibiotic activity against major potato pathogens, implying that these metabolites play an important role in niche defense⁹⁸. Another potent antibiotic from a *Clostridium* species (*Ruminiclostridium cellulolyticum*) is closthioamide (CTA, **26**)⁹⁹. Its biosynthesis, however, has remained enigmatic until recently. Now it was demonstrated that CTA is biosynthesized via a novel thiotemplated peptide assembly line that differs from canonical NRPSs and therefore escaped previous genomic analyses with automated software tools^{100,101}. This example showcases the limitations of automated genome mining efforts relying on characterized biosynthetic mechanisms.

Archaea are known to generally have small genomes, and analyses of their secondary metabolite BGCs are scarce. A systematic screening of archaeal genome sequences for the presence of putative secondary metabolite BGCs revealed that the majority of archaeal genomes only harbor one or two types of secondary metabolite BGCs, with bacteriocin- and terpene-encoding genes being most abundant¹⁰². NRPS encoding genes are sporadically found¹⁰³. This biosynthetic potential is reflected in the so-far-characterized secondary metabolites¹⁰⁴.

Genome-guided discovery of natural products from mushrooms and amoebae. The majority of characterized secondary metabolites from Basidiomycota constitute terpenoids. Genome analyses, however, indicate that mushrooms can synthesize chemically diverse metabolites (Fig. 4). The first reducing PKS from a mushroom was characterized from the steraceous mushroom

BY1. The PKS PPS1 is highly upregulated upon mycelial damage and synthesizes anti-larval polyenes (**27**, **28**), as was shown by heterologous expression in an *Aspergillus* host¹⁰⁵.

Bioinformatics studies also revealed the presence of a multitude of biosynthetic genes in various social amoebae, qualifying them as a promising source for genomics-based natural product discovery¹⁰⁶ (Fig. 4). Genome mining revealed the presence of classical terpene synthases in six species of amoebae. Functional expression in *E. coli* and metabolic profiling of amoebae cultures demonstrated that these organisms are able to produce a variety of terpenes. The fact that the production is restricted to specific periods during multicellular development suggests a functional role for these compounds in the native habitat or the life cycle¹⁰⁷.

Genome-guided discovery of natural products from higher organisms. Plant-derived natural products have been appreciated as medicinal agents for a long time, but genomics-guided approaches to discover novel secondary metabolites have been pursued only since the advent of modern sequencing technologies¹⁰⁸ (Fig. 4). Nowadays prediction and characterization of entire pathways are possible. Limonoids are triterpenes contributing to the bitter taste of citrus fruits and are well known for their insecticidal activity and their potential pharmaceutical properties. Through mining the genomes and transcriptomes of three diverse limonoid-producing species and expression studies, the first insight into the biosynthesis of these triterpenes was gained¹⁰⁹. Likewise, targeted mining of available plant genome sequences revealed co-localized prenyl transferase and terpene synthase genes for the biosynthesis of sesterterpenes in the Brassicaceae family. Expression of these genes in *Nicotiana benthamiae* resulted in the formation of fungal-type sesterterpenes¹¹⁰. Analysis of the transcriptome data of Chinese wolfberry (*Lycium chinense*) using the predicted core peptide sequences of three lyciumin isoforms as a probe uncovered the molecular basis of RiPP biosynthesis in plants. The lyciumin precursor gene *LbaLycA* was identified from *L. barbarum* and then characterized by heterologous expression in tobacco leaves to confirm its role in lyciumin (**29**) biosynthesis¹¹¹.

Whereas PKS and NRPS gene clusters are commonly found in many bacterial and fungal genomes, they are underrepresented in animals. Yet, the genome of the model organism *Caenorhabditis elegans* harbors a huge, multi-module hybrid PKS/NRPS and a large multi-module NRPS. To identify the encoded products, LC-MS based, comparative untargeted metabolomics of wild-type

and deletion mutants was performed, which resulted in the identification of nemamides (e.g., nemamide A, **30**), which are important for larval development and represent the first complex PKS-NRPS hybrid metabolites from a metazoan¹¹² (Fig. 4).

In another study, it was found that identical metabolites may be synthesized by microbes and animals via different pathways. Mycosporine-like amino acids and gadusols are UV-vis protective compounds produced by different marine microorganisms. They are also found in corals, marine invertebrates, and fish, and it was hypothesized that marine higher organisms obtain these compounds exclusively from their diet. However, by genome mining, it was found that zebrafish harbor putative gadusol-encoding genes. Expression analysis in the native producer and metabolic profiling of zebrafish embryos demonstrated that this organism actually synthesizes gadusol (**31**) de novo. Using the identified genes as an in silico probe, gadusol biosynthetic genes could also be detected in the genomes of birds, reptiles, and other organisms, raising the question of how this BGC evolved¹¹³ (Fig. 4).

Recently, the first functional evidence for a PKS in vertebrates was gained. In contrast to the majority of birds where red and yellow colors of the feathers are derived from carotenoid pigments obtained from the diet, parrots employ a pigment called psittacofulvin (**32**). Through genome mining, association mapping, and gene expression analysis, the gene responsible for the yellow pigmentation in the feathers of budgerigars was identified. Psittacofulvin pigments are synthesized by the PKS gene *MuPKS*, a pre-existing gene co-opted into developing feathers. Interestingly, homologous genes were identified in other birds¹¹⁴ (Fig. 4).

Genome-guided discovery of natural products from ecological interactions

Secondary metabolites play important roles as mediators of interactions among different organisms. Therefore, taking the ecological context into account can support the discovery of new bioactive compounds¹¹⁵ (Fig. 5).

Genome-guided discovery of natural products from symbiotic bacteria. Microorganisms involved in symbiotic relationships with higher organisms have been increasingly recognized as a promising source for genomics-driven natural product discovery¹¹⁶ (Fig. 5). For example, by a combination of genome mining and chemical analytics, a number of nonribosomally synthesized peptides were identified from the endosymbiont of the plant-pathogenic fungus *Rhizopus microsporus* that are involved in the bacterial–fungal interaction and/or serve an ecological function^{117–119}. The endofungal bacteria also produce cytotoxic necroximes (e.g., necroxime A, **33**) in symbiosis with the fungal host. These benzolactones are biosynthesized by a modular PKS/NRPS assembly line and may contribute to the pathogenic phenotypes ascribed to the fungal host¹²⁰. A genome-guided chemical profiling of a marine bacterium from the rhizosphere of the halophilic plant *Carex scabrifolia* yielded a number of chemically diverse natural products, some of which possess potent cytostatic activity¹²¹. A novel type of siderophore featuring diazeniumdiolate moieties for iron binding (gramibactin, **34**) was identified from the rhizosphere-associated strain *Paraburkholderia graminis*. As the corresponding gene locus is highly conserved in numerous other plant-associated bacteria, it was hypothesized that gramibactin may solubilize iron to make it accessible to the plant¹²². Subsequently, genome mining identified other types of diazeniumdiolate siderophores from other *Burkholderia* species¹²³.

Mining genomes of pathogenic microorganisms. In many cases, natural products serve as virulence factors in pathogenic interactions². Targeting the cryptic metabolome of pathogenic microorganisms (Fig. 5) may not only contribute to the understanding of pathogenicity mechanisms, but also open new opportunities to combat diseases. A salient example for the genome-based discovery of a virulence factor is the cytotoxin colibactin (**35**). From the initial discovery of its biosynthetic locus in the genome of human gut bacteria, it took more than a decade and joint forces of several workgroups until its chemical structure

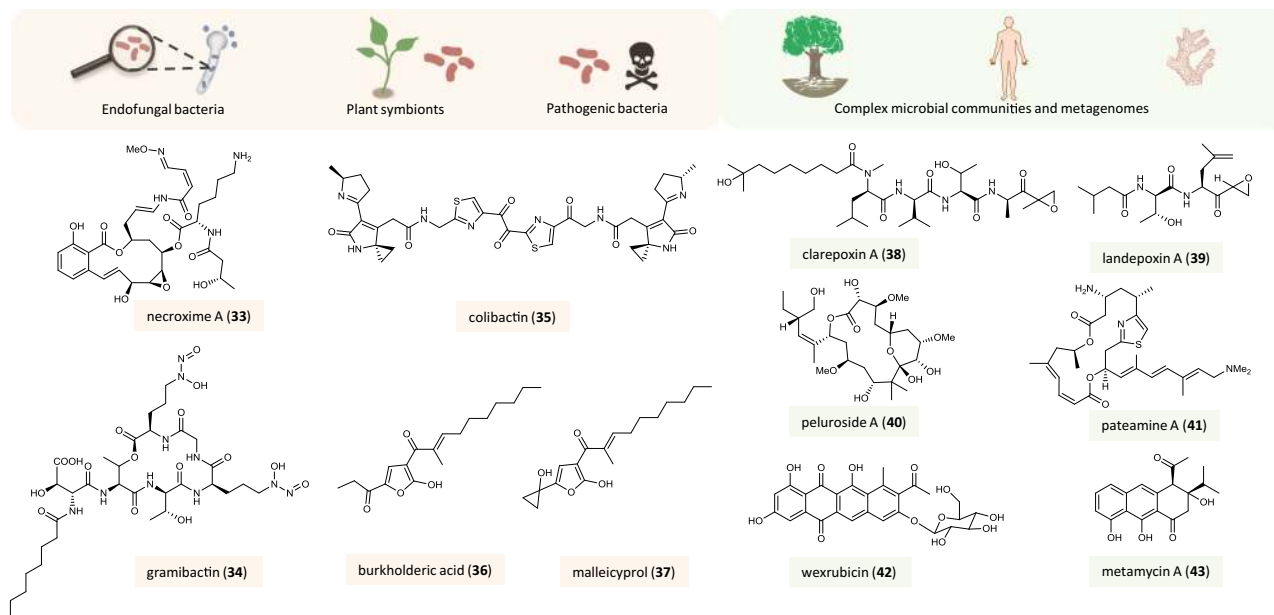


Fig. 5 Genome-guided discovery of natural products from ecological interactions. Symbiotic relationships among few or multiple partners as sources and examples of natural products discovered from these interactions. The color code of the boxes around the compound names and icons indicates the source of the respective compound: light orange—interactions among defined partners (endofungal bacteria, plant symbionts, and pathogenic bacteria), green—interactions among complex microbial communities (plant and human microbiota as well as marine sponges).

was elucidated^{124,125}. Another example concerns the infectious diseases glanders and melioidosis caused by *Burkholderia* species. Through targeted promoter exchange a cryptic NRPS/PKS hybrid gene cluster had been activated, which led to the identification of burkholderic acid (**36**, *syn.* malleilactone) in the pathogens *B. thailandensis*/*B. mallei*^{126,127}. Since this metabolite did not account for the observed pathogenic phenotype, additional compounds of this gene cluster were identified in a follow-up study. By metabolic profiling and molecular network analyses of the model organism *B. thailandensis*, unusual cyclopropanol-substituted polyketides (e.g., malleicyprol, **37**) were identified as the primary products of the cryptic pathway and shown to be highly active in a nematode infection model¹²⁸.

Mining microbiomes and metagenomes. The majority of the bacterial diversity in the environment has remained undetected due to the limitations of culturing techniques. Rapid and inexpensive sequencing technologies and bioinformatics mining of the acquired sequencing data have allowed a glimpse of the wealth of the encoded chemical diversity. It has become apparent that complex microbial communities, as encountered for example in the soil, the marine environment, or in animals/humans, are a promising source of novel natural products. Metagenomics approaches may bypass cultivation or expression problems and provide access to these molecules (Fig. 5).

For example, a metagenome exploration of *Dysideidae* sponges resulted in the identification of the BGCs responsible for the formation of cytotoxic polybrominated diphenyl ethers. The functionality of the biosynthetic genes was experimentally proven by heterologous expression in a cyanobacterial host and the origin of the compounds from the primary cyanobacterial symbiont *Hormoscilla spongelliae* was demonstrated¹²⁹. Recovering environmental DNA sequences related to known epoxyketone biosynthesis genes and heterologous expression of the complete gene clusters led to the identification of the epoxyketone proteasome inhibitors clarepoxcin A (**38**) and landepoxcin A (**39**)¹³⁰. The sponge *Mycale hentscheli* is known for the production of highly active secondary metabolites such as the microtubule-stabilizing pelorusides (e.g., peluroside A, **40**), mycalamide-type contact poisons, and the pateamines (e.g., pateamine A, **41**) with anticancer activity. Investigations of the sponge microbiome established the biosynthetic background of these potent natural products and revealed that in contrast to other sponges with known ‘super-producer’ symbionts, the huge chemical diversity is likely created by multiple bacterial species^{131,132}.

A systematic analysis of the microbiome associated with the model plant *Arabidopsis thaliana* demonstrated the huge biosynthetic potential of its symbionts. Through a high-throughput interaction screening system of a strain collection of more than 200 leaf isolates and genome mining, more than 1000 BGCs for compounds with a possible ecological relevance were identified. To demonstrate the rationality of this approach, several bioactive metabolites were characterized from a selected bacterium by bioactivity- and genomics-guided approaches¹³³.

The human microbiota have also been recognized as a fruitful source for novel metabolites, especially for antibiotics^{134–136}. The growing availability of human microbiome sequence data has spurred efforts to develop efficient systems to leverage the encoded diversity. Assemblies of complex metagenomic sequencing data frequently consist of fragmented BGCs and sequences of the most abundant members of the microbiome are over-represented. Thus, the biosynthetic capabilities of less abundant bacteria often remain hidden. Therefore, an assembly-independent approach was developed that allows the direct

identification of BGCs from metagenomics reads. With this method, a huge human microbiome data set was analyzed, which resulted in the identification of type II PKS gene clusters that are widely distributed among gut, oral, and skin microbiomes. Cloning and heterologous expression of selected genes led to the discovery of wexrubicin (**42**) and metamyxins (e.g., metamyxin A, **43**) with antibiotic activities¹³⁷. Techniques of single-cell genomics are currently being explored by natural product researchers as an alternative option to address these problems¹³⁸.

Conclusions and Perspectives

Advances in genomics and bioinformatics have reinvigorated natural product research to become a more targeted and systematic research endeavor with the genomic information as the starting point. Rapid and inexpensive sequencing technologies along with powerful bioinformatics tools have furthered our insights into microbial and structural diversity, revealing nearly unlimited natural product discovery potential. Besides the targeted genome-guided reinvestigation of well-known producers, the exploration of unconventional sources such as anaerobic bacteria or higher organisms promises exciting discoveries. Likewise, the investigation of natural products in their native environment (as mediators of pairwise or complex organismal interaction) will not only help to understand natural product functions, but also uncover novel drug candidates inspired from the ecological function of secondary metabolites. Major hurdles in accessing the hidden chemical diversity will be subsequently overcome by the ongoing development of innovative culturing methods, efficient genome editing, and optimized expression systems. Along with more sensitive chemical analytics, this will especially apply leverage to the discovery of novel metabolites from metagenomics data. Single-cell-based technologies will open additional possibilities to study biosynthetic capabilities of individual members of microbiomes.

Data availability

Data sharing not applicable as no original research is reported.

Received: 18 December 2020; Accepted: 4 June 2021;
Published online: 23 June 2021

References

1. Newman, D. J. & Cragg, G. M. Natural products as sources of new drugs from 1981 to 2014. *J. Nat. Prod.* **79**, 629–661 (2016).
2. Scherlach, K. & Hertweck, C. Chemical mediators at the bacterial-fungal interface. *Annu. Rev. Microbiol.* **74**, 267–290 (2020).
3. Katz, L. & Baltz, R. H. Natural product discovery: past, present, and future. *J. Ind. Microbiol. Biotechnol.* **43**, 155–176 (2016).
4. Hertweck, C. The biosynthetic logic of polyketide diversity. *Angew. Chem. Int. Ed.* **48**, 4688–4716 (2009).
5. Walsh, C. T. Insights into the chemical logic and enzymatic machinery of NRPS assembly lines. *Nat. Prod. Rep.* **33**, 127–135 (2016).
6. Hudson, G. A. & Mitchell, D. A. RiPP antibiotics: biosynthesis and engineering potential. *Curr. Opin. Microbiol.* **45**, 61–69 (2018).
7. Mullowney, M. W., McClure, R. A., Robey, M. T., Kelleher, N. L. & Thomson, R. J. Natural products from thioester reductase containing biosynthetic pathways. *Nat. Prod. Rep.* **35**, 847–878 (2018).
8. Baunach, M., Franke, J. & Hertweck, C. Terpenoid biosynthesis off the beaten track: unconventional cyclases and their impact on biomimetic synthesis. *Angew. Chem. Int. Ed.* **54**, 2604–2626 (2015).
9. Zhang, M. M., Qiao, Y., Ang, E. L. & Zhao, H. Using natural products for drug discovery: the impact of the genomics era. *Expert Opin. Drug Discov.* **12**, 475–487 (2017).
10. Scherlach, K. & Hertweck, C. Triggering cryptic natural product biosynthesis in microorganisms. *Org. Biomol. Chem.* **7**, 1753–1760 (2009).
11. Ren, H., Shi, C. & Zhao, H. Computational tools for discovering and engineering natural product biosynthetic pathway. *iScience* **23**, 100795 (2020).

12. van der Hoof, J. J. J. et al. Linking genomics and metabolomics to chart specialized metabolic diversity. *Chem. Soc. Rev.* **49**, 3297–3314 (2020).
13. van der Lee, T. A. J. & Medema, M. H. Computational strategies for genome-based natural product discovery and engineering in fungi. *Fungal Genet. Biol.* **89**, 29–36 (2016).
14. Alanjary, M., Cano-Prieto, C., Gross, H. & Medema, M. H. Computer-aided re-engineering of nonribosomal peptide and polyketide biosynthetic assembly lines. *Nat. Prod. Rep.* **36**, 1249–1261 (2019).
15. Kim, H. U., Blin, K., Lee, S. Y. & Weber, T. Recent development of computational resources for new antibiotics discovery. *Curr. Opin. Microbiol.* **39**, 113–120 (2017).
16. Montalbán-López, M. et al. New developments in RiPP discovery, enzymology and engineering. *Nat. Prod. Rep.* **38**, 130–239 (2020).
17. Cimermancic, P. et al. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **158**, 412–421 (2014).
18. Hannigan, G. D. et al. A deep learning genome-mining strategy for biosynthetic gene cluster prediction. *Nucleic Acids Res.* **47**, e110 (2019).
19. Kloosterman, A. M. et al. Expansion of RiPP biosynthetic space through integration of pan-genomics and machine learning uncovers a novel class of lanthipeptides. *PLoS Biol.* **18**, e3001026 (2020).
20. Kautsar, S. A., van der Hoof, J. J. J., de Ridder, D. & Medema, M. H. BiG-SLiCE: a highly scalable tool maps the diversity of 1.2 million biosynthetic gene clusters. *Gigascience* **10**, gaa154 (2021).
21. Zierp, P. F., Ceci, A. T., Dobrusin, I., Rockwell-Kollmann, S. C. & Günther, S. SeMPI 2.0-A web server for PKS and NRPS predictions combined with metabolite screening in natural product databases. *Metabolites* **11**, 13 (2020).
22. Chu, J. et al. Discovery of MRSA active antibiotics using primary sequence from the human microbiome. *Nat. Chem. Biol.* **12**, 1004–1006 (2016).
23. Vila-Farres, X. et al. Antimicrobials inspired by nonribosomal peptide synthetase gene clusters. *J. Am. Chem. Soc.* **139**, 1404–1407 (2017).
24. Vila-Farres, X. et al. An optimized synthetic-bioinformatic natural product antibiotic sterilizes multidrug-resistant *Acinetobacter baumannii*-infected wounds. *mSphere* **3**, e00528-17 (2018).
25. Fields, F. R. et al. Novel antimicrobial peptide discovery using machine learning and biophysical selection of minimal bacteriocin domains. *Drug Dev. Res.* **81**, 43–51 (2020).
26. Hudson, G. A., Hooper, A. R., DiCaprio, A. J., Sarlah, D. & Mitchell, D. A. Structure prediction and synthesis of pyridine-based macrocyclic peptide natural products. *Org. Lett.* <https://doi.org/10.1021/acs.orglett.1020c02699> (2020).
27. Merwin, N. J. et al. DeepRiPP integrates multiomics data to automate discovery of novel ribosomally synthesized natural products. *Proc. Natl Acad. Sci. USA* **117**, 371–380 (2020).
28. Cao, L. et al. MetaMiner: A scalable peptidogenomics approach for discovery of ribosomal peptide natural products with blind modifications from microbial communities. *Cell Syst.* **9**, 600–608 e604 (2019).
29. Mohimani, H. et al. NRPquest: Coupling mass spectrometry and genome mining for nonribosomal peptide discovery. *J. Nat. Prod.* **77**, 1902–1909 (2014).
30. Ricart, E. et al. rBAN: retro-biosynthetic analysis of nonribosomal peptides. *J. Cheminform.* **11**, 13 (2019).
31. Doroghazi, J. R. et al. A roadmap for natural product discovery based on large-scale genomics and metabolomics. *Nat. Chem. Biol.* **10**, 963–968 (2014).
32. Duncan, K. R. et al. Molecular networking and pattern-based genome mining improves discovery of biosynthetic gene clusters and their products from *Salinispora* species. *Chem. Biol.* **22**, 460–471 (2015).
33. Medema, M. H. et al. Pep2Path: automated mass spectrometry-guided genome mining of peptidic natural products. *PLoS Comput. Biol.* **10**, e1003822 (2014).
34. Mir Mohseni, M. et al. Discovery of a mosaic-like biosynthetic assembly line with a decarboxylative off-loading mechanism through a combination of genome mining and imaging. *Angew. Chem. Int. Ed.* **55**, 13611–13614 (2016).
35. Pancrace, C. et al. Unique biosynthetic pathway in bloom-forming cyanobacterial genus *Microcystis* jointly assembles cytotoxic aeruginoguanidines and microguanidines. *ACS Chem. Biol.* **14**, 67–75 (2019).
36. Zhao, S. et al. Prediction and characterization of enzymatic activities guided by sequence similarity and genome neighborhood networks. *eLife* **3**, e03275 (2014).
37. Rudolf, J. D., Yan, X. & Shen, B. Genome neighborhood network reveals insights into enediynes biosynthesis and facilitates prediction and prioritization for discovery. *J. Ind. Microbiol. Biotechnol.* **43**, 261–276 (2016).
38. Navarro-Muñoz, J. C. et al. A computational framework to explore large-scale biosynthetic diversity. *Nat. Chem. Biol.* **16**, 60–68 (2020).
39. Alanjary, M. et al. The Antibiotic Resistant Target Seeker (ARTS), an exploration engine for antibiotic cluster prioritization and novel drug target discovery. *Nucleic Acids Res.* **45**, W42–W48 (2017).
40. Mungan, M. D. et al. ARTS 2.0: feature updates and expansion of the Antibiotic Resistant Target Seeker for comparative genome mining. *Nucleic Acids Res.* **48**, W546–W552 (2020).
41. Wright, G. D. Molecular mechanisms of antibiotic resistance. *Chem. Commun.* **47**, 4055–4061 (2011).
42. O'Neill, E. C., Schorn, M., Larson, C. B. & Millán-Aguíñaga, N. Targeted antibiotic discovery through biosynthesis-associated resistance determinants: target directed genome mining. *Crit. Rev. Microbiol.* **45**, 255–277 (2019).
43. Tang, X. et al. Identification of thiotetronic acid antibiotic biosynthetic pathways by target-directed genome mining. *ACS Chem. Biol.* **10**, 2841–2849 (2015).
44. Panter, F., Krug, D., Baumann, S. & Müller, R. Self-resistance guided genome mining uncovers new topoisomerase inhibitors from myxobacteria. *Chem. Sci.* **9**, 4898–4908 (2018).
45. Yan, Y. et al. Resistance-gene-directed discovery of a natural-product herbicide with a new mode of action. *Nature* **559**, 415–418 (2018). **The authors propose a new fungal genome mining strategy to discover a herbicidal natural product based on the search for associated resistance genes.**
46. Johnston, C. W. et al. Assembly and clustering of natural antibiotics guides target identification. *Nat. Chem. Biol.* **12**, 233–239 (2016).
47. Skinnider, M. A., Merwin, N. J., Johnston, C. W. & Magarvey, N. A. PRISM 3: expanded prediction of natural product chemical structures from microbial genomes. *Nucleic Acids Res.* **45**, W49–w54 (2017).
48. Kjørboelling, I., Vesth, T. & Andersen, M. R. Resistance gene-directed genome mining of 50 *Aspergillus* species. *mSystems* **4** (2019).
49. Zhu, X., Su, M., Manickam, K. & Zhang, W. Bacterial genome mining of enzymatic tools for alkyne biosynthesis. *ACS Chem. Biol.* **10**, 2785–2793 (2015).
50. Lv, J. M. et al. Biosynthesis of biscognienyne B involving a cytochrome P450-dependent alkynylation. *Angew. Chem. Int. Ed.* **59**, 13531–13536 (2020).
51. Pan, G. et al. Discovery of the leinamycin family of natural products by mining actinobacterial genomes. *Proc. Natl Acad. Sci. USA* **114**, E11131–E11140 (2017).
52. Tan, D. et al. Genome-mined Diels-Alderase catalyzes formation of the *cis*-Octahydrodecalins of varicidin A and B. *J. Am. Chem. Soc.* **141**, 769–773 (2019).
53. Ueoka, R., Bortfeld-Miller, M., Morinaka, B. I., Vorholt, J. A. & Piel, J. Toblerols: cyclopropanol-containing polyketide modulators of antibiosis in *Methylobacteria*. *Angew. Chem. Int. Ed.* **57**, 977–981 (2018).
54. Ueoka, R. et al. Genome mining of oxidation modules in trans-Acyltransferase polyketide synthases reveals a culturable source for lobatamides. *Angew. Chem. Int. Ed.* **59**, 7761–7765 (2020).
55. Hetrick, K. J. & van der Donk, W. A. Ribosomally synthesized and post-translationally modified peptide natural product discovery in the genomic era. *Curr. Opin. Chem. Biol.* **38**, 36–44 (2017).
56. Russell, A. H. & Truman, A. W. Genome mining strategies for ribosomally synthesised and post-translationally modified peptides. *Comput. Struct. Biotechnol. J.* **18**, 1838–1851 (2020).
57. Kloosterman, A. M., Shelton, K. E., van Wezel, G. P., Medema, M. H. & Mitchell, D. A. RRE-finder: a genome-mining tool for class-independent RiPP discovery. *mSystems* **5**, e00267-20 (2020).
58. de Los Santos, E. L. C. NeuRiPP: Neural network identification of RiPP precursor peptides. *Sci. Rep.* **9**, 13406 (2019).
59. Skinnider, M. A. et al. Genomic charting of ribosomally synthesized natural product chemical space facilitates targeted mining. *Proc. Natl. Acad. Sci. USA* **113**, E6343–E6351 (2016).
60. Viehrig, K. et al. Structure and biosynthesis of crocagins: polycyclic posttranslationally modified ribosomal peptides from *Chondromyces crocatus*. *Angew. Chem. Int. Ed.* **56**, 7407–7410 (2017).
61. Bhushan, A., Egli, P. J., Peters, E. E., Freeman, M. F. & Piel, J. Genome mining- and synthetic biology-enabled production of hypermodified peptides. *Nat. Chem.* **11**, 931–939 (2019).
62. Ramm, S. et al. A self-sacrificing *N*-methyltransferase is the precursor of the fungal natural product omphalotin. *Angew. Chem. Int. Ed.* **56**, 9994–9997 (2017).
63. van der Velden, N. S. et al. Autocatalytic backbone *N*-methylation in a family of ribosomal peptide natural products. *Nat. Chem. Biol.* **13**, 833–835 (2017).
64. Quijano, M. R. et al. Distinct autocatalytic α -*N*-methylating precursors expand the borosin RiPP family of peptide natural products. *J. Am. Chem. Soc.* **141**, 9637–9644 (2019).
65. Adamek, M., Alanjary, M. & Ziemert, N. Applied evolution: phylogeny-based approaches in natural products research. *Nat. Prod. Rep.* **36**, 1295–1312 (2019).
66. Mullins, A. J. et al. Genome mining identifies cepacin as a plant-protective metabolite of the biopesticidal bacterium *Burkholderia ambifaria*. *Nat. Microbiol.* **4**, 996–1005 (2019).
67. Cruz-Morales, P. et al. Phylogenomic analysis of natural products biosynthetic gene clusters allows discovery of arseno-organic metabolites in model streptomycetes. *Genome Biol. Evol.* **8**, 1906–1916 (2016).
68. Bruns, H. et al. Function-related replacement of bacterial siderophore pathways. *ISME J.* **12**, 320–329 (2018).

69. Adnani, N. et al. Coculture of marine invertebrate-associated bacteria and interdisciplinary technologies enable biosynthesis and discovery of a new antibiotic, keyicin. *ACS Chem. Biol.* **12**, 3093–3102 (2017).
70. Stroe, M. C. et al. Targeted induction of a silent fungal gene cluster encoding the bacteria-specific germination inhibitor fumigermin. *eLife* **9**, e52541 (2020).
71. Seyedsayamdost, M. R. High-throughput platform for the discovery of elicitors of silent bacterial gene clusters. *Proc. Natl Acad. Sci. USA* **111**, 7266–7271 (2014).
72. Xu, F., Nazari, B., Moon, K., Bushin, L. B. & Seyedsayamdost, M. R. Discovery of a cryptic antifungal compound from *Streptomyces albus* J1074 using high-throughput elicitor screens. *J. Am. Chem. Soc.* **139**, 9203–9212 (2017).
73. Guo, F. et al. Targeted activation of silent natural product biosynthesis pathways by reporter-guided mutant selection. *Metab. Eng.* **28**, 134–142 (2015).
74. Hosaka, T. et al. Antibacterial discovery in actinomycetes strains with mutations in RNA polymerase or ribosomal protein S12. *Nat. Biotechnol.* **27**, 462–464 (2009).
75. Zhu, S., Duan, Y. & Huang, Y. The application of ribosome engineering to natural product discovery and yield improvement in *Streptomyces*. *Antibiotics* **8**, 133 (2019).
76. Ochi, K. Insights into microbial cryptic gene activation and strain improvement: principle, application and technical aspects. *J. Antibiot.* **70**, 25–40 (2017).
77. Thong, W. L., Shin-Ya, K., Nishiyama, M. & Kuzuyama, T. Discovery of an antibacterial isoidolinone-containing tetracyclic polyketide by cryptic gene activation and characterization of its biosynthetic gene cluster. *ACS Chem. Biol.* **13**, 2615–2622 (2018).
78. Amos, G. C. A. et al. Comparative transcriptomics as a guide to natural product discovery and biosynthetic gene cluster functionality. *Proc. Natl Acad. Sci. USA* **114**, E11121–E11130 (2017).
79. Zhang, M. M. et al. CRISPR-Cas9 strategy for activation of silent *Streptomyces* biosynthetic gene clusters. *Nat. Chem. Biol.* (2017).
80. Wang, X. et al. Discovery of recombinases enables genome mining of cryptic biosynthetic gene clusters in Burkholderiales species. *Proc. Natl Acad. Sci. USA* **115**, E4255–e4263 (2018).
81. Brakhage, A. A. Regulation of fungal secondary metabolism. *Nat. Rev. Microbiol.* **11**, 21–32 (2013).
82. Mao, X. M. et al. Epigenetic genome mining of an endophytic fungus leads to the pleiotropic biosynthesis of natural products. *Angew. Chem. Int. Ed.* **54**, 7592–7596 (2015).
83. Fischer, J. et al. Chromatin mapping identifies BasR, a key regulator of bacteria-triggered production of fungal secondary metabolites. *eLife* **7**, e40969 (2018).
84. Greunke, C. et al. Direct Pathway Cloning (DiPaC) to unlock natural product biosynthetic potential. *Metab. Eng.* **47**, 334–345 (2018).
85. Jiang, W. et al. Cas9-Assisted Targeting of chromosome segments CATCH enables one-step targeted cloning of large gene clusters. *Nat. Commun.* **6**, 8101 (2015).
86. Huo, L. et al. Heterologous expression of bacterial natural product biosynthetic pathways. *Nat. Prod. Rep.* **36**, 1412–1436 (2019).
87. Chan, A. N., Santa Maria, K. C. & Li, B. Direct capture technologies for genomics-guided discovery of natural products. *Curr. Top. Med. Chem.* **16**, 1695–1704 (2016).
88. Bösch, N. M. et al. Landornamides: antiviral ornithine-containing ribosomal peptides discovered through genome mining. *Angew. Chem. Int. Ed.* **59**, 11763–11768 (2020).
89. Matsuda, Y. et al. Astellifadiene: structure determination by NMR spectroscopy and crystalline sponge method, and elucidation of its biosynthesis. *Angew. Chem. Int. Ed.* **55**, 5785–5788 (2016).
90. Kang, H. S., Charlop-Powers, Z. & Brady, S. F. Multiplexed CRISPR/Cas9- and TAR-mediated promoter engineering of natural product biosynthetic gene clusters in yeast. *ACS Synth. Biol.* **5**, 1002–1010 (2016). **The report introduces a yeast-based promoter engineering platform that was developed to enable single-marker multiplexed promoter engineering to aid the refactoring of biosynthetic gene clusters.**
91. Kim, S. H. et al. Atolypenes, tricyclic bacterial sesterterpenes discovered using a multiplexed in vitro Cas9-TAR gene cluster refactoring approach. *ACS Synth. Biol.* **8**, 109–118 (2019).
92. Letzel, A. C., Pidot, S. J. & Hertweck, C. A genomic approach to the cryptic secondary metabolome of the anaerobic world. *Nat. Prod. Rep.* **30**, 392–428 (2013).
93. Schieferdecker, S. et al. Biosynthesis of diverse antimicrobial and antiproliferative acylolins in anaerobic bacteria. *ACS Chem. Biol.* **14**, 1490–1497 (2019).
94. Ishida, K. et al. Oak-associated negaticute equipped with ancestral aromatic polyketide synthase produces antimycobacterial dendrubsins. *Chem. Eur. J.* **26**, 13147–13151 (2020).
95. Li, J. S., Barber, C. C. & Zhang, W. Natural products from anaerobes. *J. Ind. Microbiol. Biotechnol.* **46**, 375–383 (2019).
96. Rischer, M. et al. Biosynthesis, synthesis, and activities of barnesin A, a NRPS-PKS hybrid produced by an anaerobic epsilonproteobacterium. *ACS Chem. Biol.* **13**, 1990–1995 (2018).
97. Herman, N. A. et al. The industrial anaerobe *Clostridium acetobutylicum* uses polyketides to regulate cellular differentiation. *Nat. Commun.* **8**, 1514 (2017).
98. Shabuer, G. et al. Plant pathogenic anaerobic bacteria use aromatic polyketides to access aerobic territory. *Science* **350**, 670–674 (2015).
99. Lincke, T., Behnken, S., Ishida, K., Roth, M. & Hertweck, C. Closthioamide: an unprecedented polythioamide antibiotic from the strictly anaerobic bacterium *Clostridium cellulolyticum*. *Angew. Chem. Int. Ed.* **49**, 2011–2013 (2010).
100. Dunbar, K. L. et al. Genome editing reveals novel thiotemplated assembly of polythioamide antibiotics in anaerobic bacteria. *Angew. Chem. Int. Ed.* **57**, 14080–14084 (2018). **The paper describes the genome-mining-based discovery of the biosynthetic gene cluster of the antibiotic closthioamide. A novel mechanism for the NRPS-independent assembly of thioamide-containing nonribosomal peptides is presented.**
101. Dunbar, K. L., Dell, M., Gude, F. & Hertweck, C. Reconstitution of polythioamide antibiotic backbone formation reveals unusual thiotemplated assembly strategy. *Proc. Natl Acad. Sci. USA* **117**, 8850–8858 (2020).
102. Wang, S., Zheng, Z., Zou, H., Li, N. & Wu, M. Characterization of the secondary metabolite biosynthetic gene clusters in archaea. *Comput. Biol. Chem.* **78**, 165–169 (2019).
103. Wang, H., Fewer, D. P., Holm, L., Rouhiainen, L. & Sivonen, K. Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. *Proc. Natl Acad. Sci. USA* **111**, 9259–9264 (2014).
104. Charlesworth, J. C. & Burns, B. P. Untapped resources: Biotechnological potential of peptides and secondary metabolites in archaea. *Archaea* **2015**, 282035 (2015).
105. Brandt, P., Garcia-Altare, M., Nett, M., Hertweck, C. & Hoffmeister, D. Induced chemical defense of a mushroom by a double-bond-shifting polyene synthase. *Angew. Chem. Int. Ed.* **56**, 5937–5941 (2017). **The authors show that certain mushrooms produce antilarval polyenes upon injury. The study represents the first characterization of a reducing polyketide synthase from a mushroom.**
106. Barnett, R. & Stallforth, P. Natural products from social amoebae. *Chemistry* **24**, 4202–4214 (2018).
107. Chen, X. et al. Terpene synthase genes in eukaryotes beyond plants and fungi: Occurrence in social amoebae. *Proc. Natl Acad. Sci. USA* **113**, 12132–12137 (2016).
108. Nützmann, H. W., Huang, A. & Osbourn, A. Plant metabolic clusters - from genetics to genomics. *N. Phytol.* **211**, 771–789 (2016).
109. Hodgson, H. et al. Identification of key enzymes responsible for protolimonoid biosynthesis in plants: Opening the door to azadirachtin production. *Proc. Natl Acad. Sci. USA* **116**, 17096–17104 (2019).
110. Huang, A. C. et al. Unearthing a sesterterpene biosynthetic repertoire in the Brassicaceae through genome mining reveals convergent evolution. *Proc. Natl Acad. Sci. USA* **114**, E6005–E6014 (2017).
111. Kersten, R. D. & Weng, J. K. Gene-guided discovery and engineering of branched cyclic peptides in plants. *Proc. Natl Acad. Sci. USA* **115**, E10961–E10969 (2018).
112. Shou, Q. et al. A hybrid polyketide-nonribosomal peptide in nematodes that promotes larval survival. *Nat. Chem. Biol.* **12**, 770–772 (2016). **The first functional characterization of an animal PKS-NRPS in a nematode is reported.**
113. Osborn, A. R. et al. De novo synthesis of a sunscreen compound in vertebrates. *eLife* **4**, e05919 (2015). **This work uncovers a novel pathway to biologically important sunscreen compounds used by zebrafish and likely other higher organisms.**
114. Cooke, T. F. et al. Genetic mapping and biochemical basis of yellow feather pigmentation in budgerigars. *Cell* **171**, 427–439 (2017). **The genetic and biochemical background of psittacofulvin formation in budgerigars is described. The study highlights the diversity of polyketide synthase functions across animals.**
115. Molloy, E. M. & Hertweck, C. Antimicrobial discovery inspired by ecological interactions. *Curr. Opin. Microbiol.* **39**, 121–127 (2017).
116. Adnani, N., Rajski, S. R. & Bugni, T. S. Symbiosis-inspired approaches to antibiotic discovery. *Nat. Prod. Rep.* **34**, 784–814 (2017).
117. Niehs, S. P. et al. Genome mining reveals endopyrroles from a nonribosomal peptide assembly line triggered in fungal-bacterial symbiosis. *ACS Chem. Biol.* **14**, 1811–1818 (2019).
118. Niehs, S. P., Dose, B., Scherlach, K., Roth, M. & Hertweck, C. Genomics-driven discovery of a symbiont-specific cyclopeptide from bacteria residing in the rice seedling blight fungus. *Chem Bio Chem.* **19**, 2167–2172 (2018).
119. Niehs, S. P., Scherlach, K. & Hertweck, C. Genomics-driven discovery of a linear lipopeptide promoting host colonization by endofungal bacteria. *Org. Biomol. Chem.* **16**, 8345–8352 (2018).

120. Niehs, S. P. et al. Mining symbionts of a spider-transmitted fungus illuminates uncharted biosynthetic pathways to cytotoxic benzolactones. *Angew. Chem. Int. Ed.* **59**, 7766–7771 (2020).
121. Ueoka, R. et al. Genome-based identification of a plant-associated marine bacterium as a rich natural product source. *Angew. Chem. Int. Ed.* **57**, 14519–14523 (2018).
122. Hermenau, R. et al. Gramibactin is a bacterial siderophore with a diazeniumdiolate ligand system. *Nat. Chem. Biol.* **14**, 841–843 (2018).
123. Hermenau, R. et al. Genomics-driven discovery of NO-donating diazeniumdiolate siderophores in diverse plant-associated bacteria. *Angew. Chem. Int. Ed.* **58**, 13024–13029 (2019).
124. Wernke, K. M. et al. Structure and bioactivity of colibactin. *Bioorg. Med. Chem. Lett.* **30**, 127280 (2020).
125. Xue, M. et al. Structure elucidation of colibactin and its DNA cross-links. *Science* **365**, eaax2685 (2019).
126. Franke, J., Ishida, K. & Hertweck, C. Genomics-driven discovery of burkholderic acid, a noncanonical, cryptic polyketide from human pathogenic *Burkholderia* species. *Angew. Chem. Int. Ed.* **51**, 11611–11615 (2012).
127. Biggins, J. B., Ternei, M. A. & Brady, S. F. Malleilactone, a polyketide synthase-derived virulence factor encoded by the cryptic secondary metabolome of *Burkholderia pseudomallei* group pathogens. *J. Am. Chem. Soc.* **134**, 13192–13195 (2012).
128. Trottmann, F. et al. Cyclopropanol warhead in malleicyprol confers virulence of human- and animal-pathogenic *Burkholderia* species. *Angew. Chem. Int. Ed.* **58**, 14129–14133 (2019).
129. Agarwal, V. et al. Metagenomic discovery of polybrominated diphenyl ether biosynthesis by marine sponges. *Nat. Chem. Biol.* **13**, 537–543 (2017).
130. Owen, J. G. et al. Multiplexed metagenome mining using short DNA sequence tags facilitates targeted discovery of epoxyketone proteasome inhibitors. *Proc. Natl Acad. Sci. USA* **112**, 4221–4226 (2015).
131. Rust, M. et al. A multiproducer microbiome generates chemical diversity in the marine sponge *Mycale hentscheli*. *Proc. Natl Acad. Sci. USA* **117**, 9508–9518 (2020).
132. Storey, M. A. et al. Metagenomic exploration of the marine sponge *Mycale hentscheli* uncovers multiple polyketide-producing bacterial symbionts. *mBio* **11**, e02997–02919 (2020).
133. Helfrich, E. J. N. et al. Bipartite interactions, antibiotic production and biosynthetic potential of the *Arabidopsis* leaf microbiome. *Nat. Microbiol.* **3**, 909–919 (2018).
134. Donia, M. S. et al. A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. *Cell* **158**, 1402–1414 (2014).
135. Zipperer, A. et al. Human commensals producing a novel antibiotic impair pathogen colonization. *Nature* **535**, 511–516 (2016).
136. Donia, M. S. & Fischbach, M. A. Small molecules from the human microbiota. *Science* **349**, 1254766 (2015).
137. Sugimoto, Y. et al. A metagenomic strategy for harnessing the chemical repertoire of the human microbiome. *Science* **366**, eaax9176 (2019). **The authors describe a computational approach to identify secondary**

metabolite biosynthetic gene clusters from human microbiome metagenomic sequences and use this strategy to identify novel antibiotics.

138. Piel, J. & Cahn, J. Opening up the single-cell toolbox for microbial natural products research. *Angew. Chem. Int. Ed.* <https://doi.org/10.1002/anie.201900532> (2019).

Acknowledgements

The authors would like to acknowledge the financial support of their original work by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Project-ID 239748522—SFB 1127 ChemBioSys, Cluster of Excellence, Balance of the Microverse, and Leibniz Award.

Author contributions

K.S. and C.H. developed the concept for the review and drafted and edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to C.H.

Peer review information *Nature Communications* thanks Hosein Mohimani, Tilmann Weber and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021