Thieme

Biological Dark Matter Exploration using Data Mining for the Discovery of Antimicrobial Natural Products#

Authors
José Rivera-Chávez[1]*, Corina-Diana Ceapă[1]*, Mario Figueroa[2]

Affiliations
1 Instituto de Química, Universidad Nacional Autónoma de México, Ciudad de México, México
2 Facultad de Química, Universidad Nacional Autónoma de México, Ciudad de México, México

Correspondence
Dr. Mario Figueroa
Facultad de Química, Universidad Nacional Autónoma de México
Avenida Universidad 3000, 04510 Mexico City, Colonia CU, Mexico
Phone: + 52 55 56 22 52 90
mafiguer@unam.mx

Supplementary material is available under https://doi.org/10.1055/a-1795-0562

ABSTRACT

The discovery of novel antimicrobials has significantly slowed down over the last three decades. At the same time, humans rely increasingly on antimicrobials because of the progressive antimicrobial resistance in medical practices, human communities, and the environment. Data mining is currently considered a promising option in the discovery of new antibiotics. Some of the advantages of data mining are the ability to predict chemical structures from sequence data, anticipation of the presence of novel metabolites, the understanding of gene evolution, and the corroboration of data from multiple omics technologies. This review analyzes the state-of-the-art for data mining in the fields of bacteria, fungi, and plant genomic data, as well as metabologenomics. It also summarizes some of the most recent research accomplishments in the field, all pinpointing to innovation through uncovering and implementing the next generation of antimicrobials.

Introduction

Antimicrobial resistance (AMR) was considered one of the major public health threats before COVID-19. Then, in 2021, it was declared one of the top ten public health issues by the World Health Organization (WHO). However, AMR has been ignored by global health agencies during the SARS-CoV-2 era, although antibiotics were prescribed indiscriminately in 70–95% of coronavirus disease cases, thus contributing to worsening this pandemic [1–4]. Up to December 9th, 2021, more than 5.8 million people world- wide have died from complications associated with COVID-19, among which a considerable amount included AMR-related issues. This counts for about half of the predicted deaths linked to AMR by 2050 [5]. Unfortunately, the AMR-related death count after COVID-19 is expected to rise significantly, both in the short and long term, mostly due to misuse, abuse, and overuse of antibiotics and sanitizers [3,4].

In other words, humanity will soon be confronted by the consequences of the inappropriate use of antibiotics. The COVID-19 pandemic has shown that preparedness ahead of such life-threat-

---

Dedicated to Professor Dr. A. Douglas Kinghorn on the occasion of his 75th birthday.

* These authors contributed equally to this paper.

ening strains is imperative. There is an urgency to develop new strategies to control the emergence and spread of AMR to avoid further health emergencies from infectious agents [3]. In this scenario, developing new antibiotics is a crucial task that must occur as soon as possible, as recommended by health agencies such as the WHO, the European Centre for Disease Prevention and Control (ECDC), the US Centers for Disease Control and Prevention (CDC), the Infectious Diseases Society of America (IDSA), the Drugs for Neglected Diseases initiative (DNDi), and global non-profit organizations such as the Global Antibiotic Research and Development Partnership (GARDP) [6].

For this purpose, several strategies, including the exploration of natural resources using conventional and state-of-the-art methodologies, as well as synthetic procedures, are currently being applied. Among these cutting-edge approaches, data mining of biological dark matter stands among the most powerful alternatives to contain the health crisis generated by the spread of AMR. Its advantages include the ability to predict the functionalities of the chemical structures starting from their biosynthetic genes, as well as the possibility to improve and diversify these molecules, all prior to production in the laboratory, thus accelerating the discovery process and providing leads for the identification of completely new classes of molecules and functionalities.

Genome mining refers to the prediction of previously uncharacterized natural product biosynthetic gene clusters (BGCs) within sequence data, analysis of the enzymes encoded by these clusters, and the experimental identification of the resulting products [7]. This research strategy was developed after the exploration by the Hopwood [8] and Ōmura [9] groups of the first sequenced *Streptomyces* genomes, which led to the idea that a great number of secondary metabolites (SMs) gene clusters appear not to be expressed under normal or laboratory conditions. The expanse of bacteria, fungi, and plants' bioactive metabolites that remain to be explored has only widened with time and with the accumulation of sequencing data [10].

The continued interest from investigators led to the development of many tools, ever increasing in complexity, to explore this new-found data [11]. The structural, chemical, and biological characterization of the newly discovered molecules and enzymes depends on *in silico* prediction algorithms. From this perspective, two scenarios are possible; first the top-down discovery processes, where the sequence data is initially analyzed, leading to predictions regarding possible functionalities, which afterwards need to be experimentally verified; and second, the bottom-up approaches which involve linking experimental data to *in silico* predictions, and a confirmation step commonly including molecular biology techniques. Currently, the main data mining strategies can be generally divided in phylogeny-based, target or resistance-based, and cultivation independent mining (comprising metagenomic data mining and single cell genomics) (▶ **Fig. 1**). This review summarizes the state-of-the-art for data mining in the fields of bacteria, fungi, and plants genomic data, as well as metabologenomics, for the discovery of new antimicrobial NPs. Without a doubt, these research accomplishments are the foundations for the development of the next generation of drugs to overcome the antimicrobial resistance. The bibliographic investigation was performed on Scopus, Scifinder, Web of Science, and Pubmed, and the main search words and terms applied were "bacterial genomic data mining" (> 350 hits), "fungal genomic data mining" (> 100 hits), and "plants genomic data mining" (> 50 hits), then filtering with the word "antimicrobials" (total 105, 18, and 14 hits, respectively). Also, the word "metabologenomics" (17 hits) was used in the bibliography search. Most of the hits resulted from the time span 2010–2020.

## Discovery of Bioactive Bacterial Products Driven by Genome Mining

Bacteria are the most relevant source of genetic information for the identification of gene clusters, which could produce molecules with antimicrobial and anti-infective effects, as indicated from past discoveries (especially the phylum Actinobacteria and genus *Streptomyces*). The diversity of possible compounds encoded by bacterial DNA is for now difficult to estimate but has been predicted to contain many chemical classes which have not yet been characterized in a laboratory setting. Molecules belonging to these classes could bring some well-needed relief to medical practice and help contain, at least temporarily, the AMR crisis, which is why their exploration is essential [3].

The origin of this biochemical diversity resides in basic natural biological processes. The main driving forces behind the evolution of organisms are those of either collaboration or competition. At genetic level, the evolution processes are defined by mutations, horizontal transfer of genes, gene loss, gene duplication, reshuffling and immunity development (CRISPR-cas systems) [12]. These changes often lead to the biochemical development of clusters of genes codifying the assembly of compounds with antimicrobial effects, improved over millions of years of evolution. Such groups of genes are denominated BGCs. Estimates of such clusters, proteins and genomes evolution, as well as finding the common ancestors, added to sequence similarity studies, have provided extremely efficient tools in uncovering new metabolites [13]. The huge diversity of chemical structures observed in natural products (NPs) research can be thus explained on the basis of the co-evolutionary processes, leading the diversification of biosynthetic mechanisms and enzymes, as needed adaptations to new ecosystems and communities [14, 15]. In this regard, phylogenetic analyses can be highly predictive of protein function and lead to engineering semi-synthetic variants and pathways to achieve improved molecules and greater product yields [16].

The number of studies involving phylogeny-based genome mining have increased exponentially over the last decade, due to the reduction in sequencing costs and worldwide distribution of the technology [17]. Extraordinary accomplishments have been made possible. Some examples are described in the following paragraphs.

Top-down discoveries

Most NPs of microbial origin with a medical application originate from the genera *Streptomyces*, phylum Actinobacteria [18]. For this reason, top-down discoveries are centered on this bacterial group, mainly for two reasons: estimations of gene diversity indicate a yet-unaccountable potential in these genomes; and be-
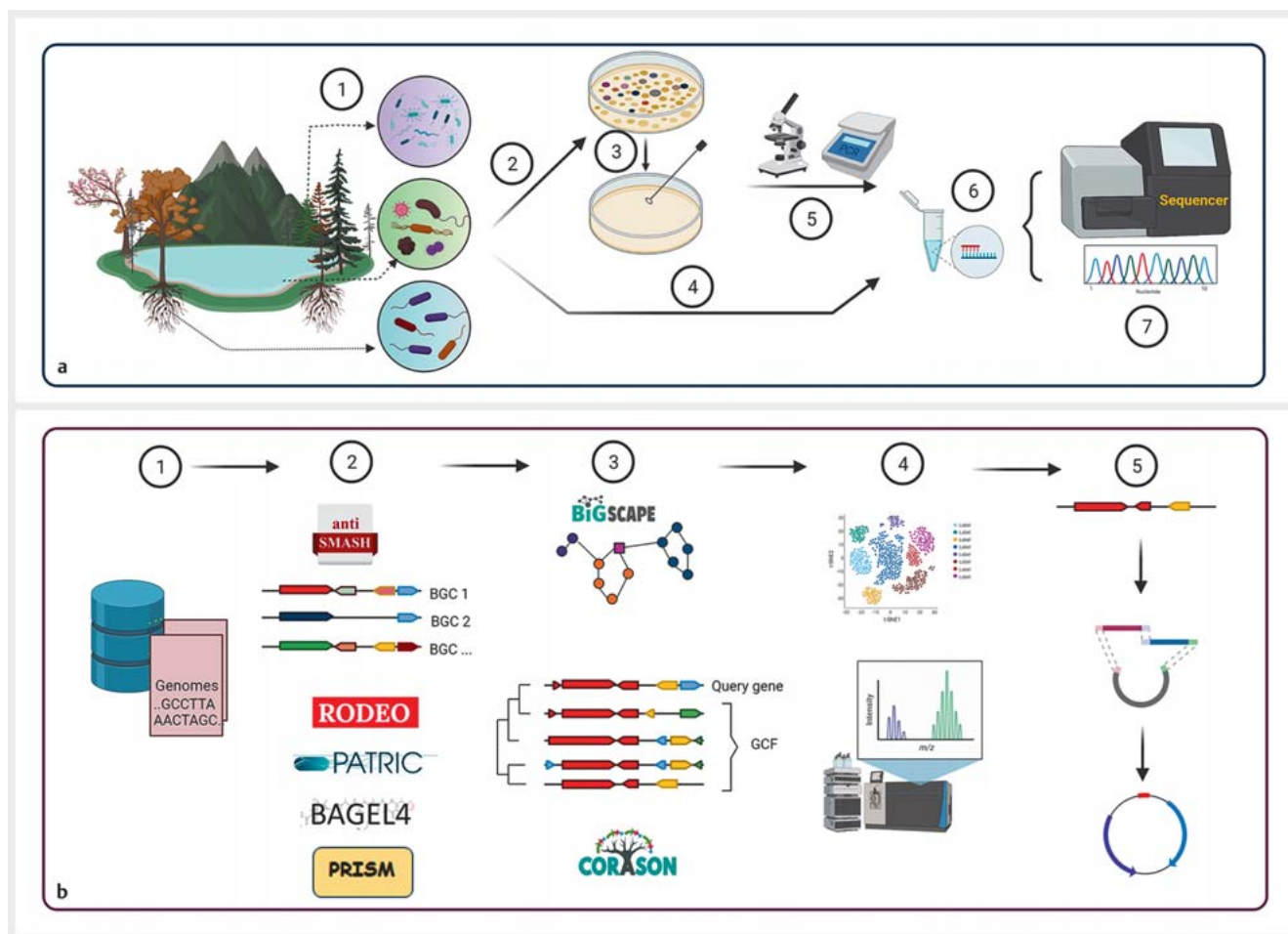
▶ **Fig. 1** Mining genome workflow: **a1** Harvesting samples from different environment and natural sources; **a2** Isolation of colonies for genomic analysis; **a3** Determination of experimental conditions for maintaining viable cultures in the laboratory; **a4** Whole genome for non-cultivable bacteria; **a5** Molecular and microbiological characterization of strain submitted to genomic analysis; **a6** DNA isolation for massive sequencing; **a7** High-throughput sequencing for genomes, metagenomes, pangenomes, single-cell genome, transcriptome. **b1** Preliminary assessment of genomic data; **b2** Identification of BGC in genomes and metagenomes; **b3** Prioritization process for BGCs using networking analysis, phylogenomic tools, and EvoMining strategies; **b4** Correlation of genomics and metabolomics data to pair novel BGCs and metabolites; **b5** Experimental confirmation of NPs production by different methods such as heterologous expression and mutant lines generation. Molecular characterization supporting the flow gene-enzyme-metabolite.

cause most of the databases and tools have started with and are more efficient at investigating Actinobacterial genes and metabolites, compared to recently discovered microorganisms [19]. This allows for a certain degree of superior security in making predictions, and the use of established experimental tools for their validation.

In Actinobacteria, BGCs by both phylogeny-based and chemical structure-based estimations indicate that genomes hold unknown clades of metabologenomic patterns of genes and molecules which remain to be characterized in the laboratory, in some taxonomic groups up to 50% of their BGC content [11, 20, 21]. Among the members of Actinobacteria, the genera *Streptomyces* is the richest in bioactive metabolites such as antibiotics, volatile compounds, siderophores, and other molecules [22]. Recent studies focusing on *Streptomyces* diversity and distribution patterns of BGCs across all public genomes as of 2019 concluded that even strains with identical phylogenetic patterns present signifi-

cantly different numbers of secondary clusters [23, 24]. Thus, strain-level genetic surveys are required to categorize and characterize novel metabolites from this genus. Additionally, it was concluded that 16S-based metagenomics of environmental samples has highly underestimated the number and value of BGCs in *Streptomyces* bacteria, as the data only offer a superficial view of the BGCs encoded in the whole genome, over-representing non-ribosomal peptides (NRPs), polyketides (PKs), terpenes, and lantipeptides.

These findings were further confirmed by Sivakala et al., who concluded that a collection of *Streptomyces* strains (from desert samples), averaging 36 BGCs per genome, displayed a unique biosynthetic potential. The majority of BGCs were predicted as novel and were characterized by an overall decrease in non-ribosomal peptide synthetases (NRPSs) clusters and an increase in the polyketides synthases (PKSs) hybrids and terpenes clusters [25]. Many predicted BGCs are also present in marine *Streptomyces*,

from 16 to 84, according to Xu et. al. [26]. These data show the dynamism of *Streptomyces* as a hyper diverse taxonomic group which has evolved several pathways, importantly increasing the NPs chemical space [27].

The group of Metsä-Ketelä has focused on the research of marine symbiotic Actinobacteria, including but not limited to *Streptomyces* [27]. Species from four rare genera (*Actinomadura, Amycolatopsis, Saccharopolyspora*, and *Streptomonospora*) were isolated for the first time from marine sponges harvested in the Persian Gulf. Of these strains, the new bioactive aromatics polyketides denoted persiamycin A and 1-hydroxy-4-methoxy-2-naphthoic acid were discovered and their existence confirmed through experimental approaches (▶ **Fig. 2**).

Rare genera and uncultivable organisms are starting to be studied by a top-down approach as well. The evolution of sequencing technologies also allows, for instance, the assembly of complete or almost-complete genomes from metagenomic data, without the need for cultivating the microorganisms [28–32]. This will be discussed in the cultivation-independent data mining section.

Achievements and latest research examples

Recent advances in the discovery of new antimicrobials by genome mining include the exploration of new species from the Proteobacteria, Firmicutes and Actinobacteria taxa [33, 34], enhancing the ability to predict and prioritize novel BGCs.

For instance, recent genome mining studies have identified the cyanobacterium *Tolypothrix* sp. PCC 7601 as a producer of tolypamide, a cyanobactin with antimicrobial activity structurally differentiated by bearing a prenyl moiety bound to a threonine residue (▶ **Fig. 2**) [35]. Another example for data mining in a new taxon is the genus *Micromonospora*, whose members can produce analogues of ramoplanin and enduracin. In addition, some strains can simultaneously produce the antimicrobial chersinamycin (▶ **Fig. 2**) [36, 37]. Finally, re-visiting bacteria of the genus *Clostridia*, an unconventional biosynthetic gene cluster for closthioamide (CTA) (▶ **Fig. 2**), a symmetric NRP composed of two diaminopropane-linked to polythioamidated monomers, with a mechanism targeting the DNA gyrase, were discovered and experimentally confirmed.

These are some of the most recent discoveries indicating that genome mining is a great alternative to find new SMs with anti-infective action [38]. In the last decade, sequencing capacity and cutting-edge computational methods to analyze these data evolved in parallel, and now there are publicly available platforms for genome mining. Some examples include AntiSMASH (which relies on the highly curated MIBiG database), PRISM, BAGEL, and RiPPMiner [39–43]. Recent discoveries using these tools have been reported for instance, a Ni-Fe hydrogenase gene cluster (hyp cluster) and putative new BGCs in *Streptomyces seoulensis* [44]. In *Staphylococcus aureus* 4181, a strain involved in bovine mastitis, a bacteriocin-denominated aureocin 4181, is involved in resistance against multidrug-resistant (MDR) staphylococci. Interestingly, both BAGEL4 and antiSMASH v.5.0 were implemented in the identification of the gene cluster coding for aureocin 4181 [45, 46]. The findings were confirmed in the wet lab.

Machine learning is also an emerging tool for developing non-supervised computational models for predicting biological outcomes using BGCs as targets [47]. One example is RODEO (Rapid ORF Description and Evaluation Online), an online tool to identify BGCs mainly ribosomally synthesized and post-translationally modified peptides (RiPPs) [48]. In this approach, the mapping of lasso peptide space generated a larger number of putative BGCs, compared to other methods. Moreover, their characterization in the laboratory confirmed the existence of new lasso peptides including citrulassin A, lagmycin and LP2006 (▶ **Fig. 2**) [48].

To prevent re-discovery of the same compounds, other deep learning-based methods have been designed to detect key genes from large-scale expression profiles and reduce redundant information. A highly sought-over edge would be to retain only the relevant biological information from the width of the data and to provide robust predictions of the compounds using only sequence data [49]. This reduces substantially the resources and time needed to confirm the data mining predictions and increases the probability for discovery of bioactive molecules.

Instead of using expression data, other projects of this type used the PRISM discovery platform in tandem with automated LC-MS/MS searches, leading to the identification of aurantizolicin and cyanobactin produced by *Cyanobacteria* (a pathogen associated with gastroenteritis during the intake of contaminated water) (▶ **Fig. 2**) [50] This bacterium appears to biosynthesize a broad group of metabolites with structural variability, generated by post-translational modifications, including prenylation.

For the chemical and biological characterization of newly identified metabolites, cloning and heterologous expression are sometimes necessary when the compound is not normally secreted in the growth medium, or the natural signals for its production are not identified. There is currently a limitation in cloning of large DNA fragments, such as those responsible for producing SMs. Therefore, advances in cloning comprise the development of new heterologous hosts for BGCs expression, e.g., anaerobic bacteria. *S. mutans* UA159 was used as a host for anaerobic BGCs by functionally expressing a known pyrazinone BGC from *Staphylococcus epidermidis* (*Sa. epidermidis*) ATCC 35 984 and two unidentified BGCs from human oral bacteria *S. mutans* 35 and *S. mutans* NMT4863, respectively. Cloning and expression of these BGCs led to the discovery of a (2*E*)-decanoyl dipeptide SNC1-465 (▶ **Fig. 2**) [51].

Target-based genome mining can rely on the interesting perspective that BGCs code also for resistance genes meant to protect organisms from self-intoxication. In the case of the glycopeptide antibiotic complestatin and corbomycin, this strategy led to the elucidation of a novel mode of action by inhibition of autolysins (▶ **Fig. 2**) [52].

Adding to that, CRISPR/Cas9 genome editing has been recently made available for gene manipulation in the widest range of organisms, including Actinobacteria. For example, surfactin production requires key precursors including glutamate (Glu), valine (Val), and *β*-hydroxy fatty acid used in a biosynthetic pathway that initiates with their assembly by SrfAA, building and extension by *srfAB*, extension and cyclization by *srfAC* and which is regulated by the operon srf. When this operon is disrupted by CRISPR/Cas9-editing, interestingly, the yield of piplastin increases. In this
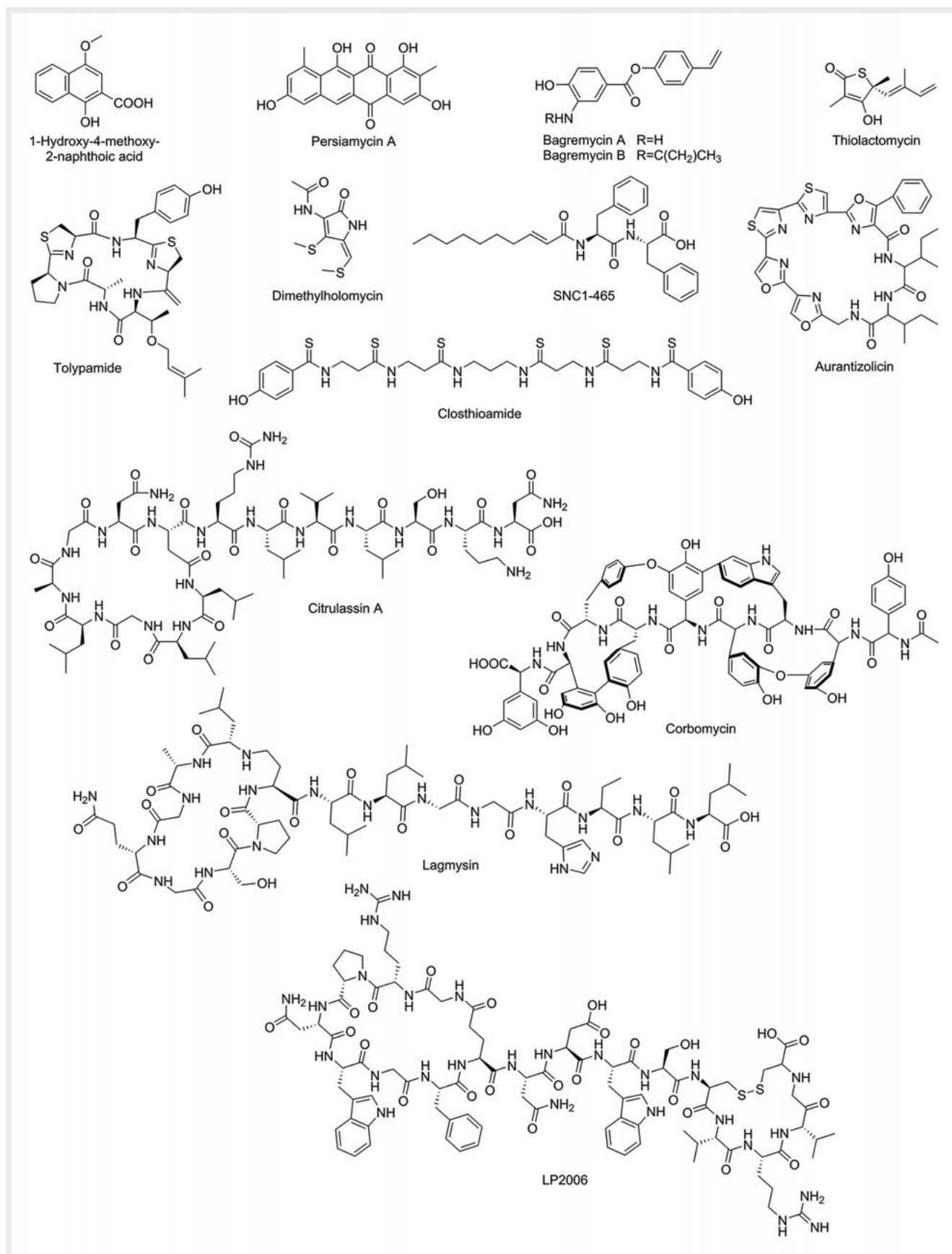
Fig. 2 Chemical structure of compounds identified by genome mining in bacteria.

example, the interruption of surfactin biosynthesis redirected their structural components toward the activation of an alternative pathway, which uses common components from the surfactin biosynthetic pathway. This experiment allowed to identify plipastatin, a cyclic lipopeptide that is synthesized by nonribosomal peptide synthetases (NRPSs) and contains amino acids with a C14–C21(*R*)-hydroxy fatty acid side chain linked by an amide bond [53].

Anticipated limitations in the expression of BGCs due to incompatibility between host and receptor bacterial DNA have been recently overcome with promising results using the CRAGE system [54]. In this technological advancement, a recombinase (RAGE system) can integrate in a single step a DNA fragment between 10 to 48 Kb. The main advantage is that, for large BGCs expression, changes in the regulation or expression control of DNA sequences in the original microorganism is now possible, avoiding the need for heterologous expression. In this study, 10 polyketide synthase and non-ribosomal peptide synthetase (PKS/NRPS) BGCs in the entomopathogenic bacterium *Photorhabdus luminescens* subsp. *laumondii* TT01 were modelled, confirming the enhanced production of 22 metabolites from six of the ten BGCs, the expression of which was modulated successfully by CRAGE-CRISPR in the wet lab [54, 55].

Orphan BGCs have attracted attention since they have yet to reveal their potential as a source of NPs. Due to the increasing number of orphans, the first challenge is centered in their selection and prioritization and connecting them to biosynthesized small molecules. An approach to reduce the number of potential BGCs is to correlate putative antibiotic resistance genes that encode target-modified protein with the orphan BGCs in a method denominated by Tang's Group as "target-directed genome mining" [56]. Co-localization between an orphan PKS-NRPS hybrid BGC (tlm) with a putative fatty acid synthase resistance gene in a *Salinispora* bacterial genome led to laboratory production of several unusual thiotetronic acid NPs, including the well-known fatty acid synthase inhibitor thiolactomycin that was first described over 30 years ago, yet never at the genetic level in regard to biosynthesis and auto-resistance (▶ **Fig. 2**) [56].

Taken together, the impulse placed by new technologies for DNA sequencing, the improvement of computational tools to detect and select the BGCs, the continuous reporting and sharing of genomic data, pan-genome analyses, and a deep understanding about the connection gene-enzyme-metabolite represents an indisputable ally in the new generation of natural product research.

Implementation of artificial intelligence strategies for data mining

The use of artificial intelligence (which in this type of approaches is based on machine learning techniques) has allowed for repurposing medicines. In one brilliant example of repurposing, scientists used machine learning (ML) to teach a neural network what the best antibiotics "look like" from a chemical standpoint and applied it to a library of more than 6000 compounds already used in human medicine [57]. The best hit, halicin, turned out to be active against a wide variety of antibiotic-resistant bacteria, including *E. coli, M. tuberculosis*, and *C. difficile*, with a set of 21 other molecules still under review. Interestingly, the halicin mechanism

of action is also a novelty, as it works by interfering with the electrochemical potential across bacterial membranes [57].

This course of action is followed by the iteration of novel principles and concepts regarding the drug targets, their mode of action and structural characteristics, which can afterwards be applied in the rational design of bioactive compounds. A whole range of ML strategies now focus on predictions related to small-molecule antibiotics and antimicrobial peptides (AMPs) [58]. While significant progress has been made, by, for instance, accelerating the drug progress towards animal trials, these strategies need to be supported by an improvement in data quality and availability, expanding the chemical space, drug repurposing and multidisciplinary collaborations to produce lead molecules adequate for human trials and thoroughly verify the findings through experimentation [58].

Synergy

Synergism is seen as the combination of two or more drugs for medical treatment, aiming to enhance drug efficacy; nevertheless, synergy in bacteria encompasses more elaborate examples. Synergy for the discovery of new antimicrobials can be two-faced: on one side, the existence of several BGCs in the same genome can act together to enhance a biological activity and on the other side, the synergy of different strategies linked to mining genome can improve the ability to discover new antimicrobials. Regarding synergy in nature, recent studies have identified bacteria with the ability to produce more than one antimicrobial at the same time, which are biosynthesized by different clusters of genes located in adjacent genomic regions within their genomes, denominated super-clusters [59]. The simultaneous production of antimicrobials could play an ecological role to enhance the interspecies competition for acquiring nutrients. There is little information available about how the super-clusters are structured and even less information concerning their regulation and simultaneous production [60, 61]. A first approximation revealed that super-clusters act cooperatively with synergistic potential, but cases in the current literature are scarce [62–65]. In nature, simultaneous production of antimicrobials has been identified in *Streptomyces rapamycinicus*, which can produce rapamycin, a compound with a weak antifungal activity that in combination with actinoplanic acid (APL), acts as a strong antifungal agent against *C. albicans*, probably as inhibitor of the enzyme farnesyl protein transferase, and thereby increasing the effect of rapamycin. In this example, BGCs are analyzed by synteny (co-localization of genes) and results about the genetic organization seem to pinpoint different clusters [66]. The phylogenetic approach has been used to explain antimicrobial co-production, allowing for the introduction of novel computational methods based on the probabilities of modules exchange over time. A new property based on the spatial position and order of BGCs could mean a new biological parameter not yet evaluated in the bacterial genome.

Alternatively, synergy can be induced, for example, in response to scarcity of nutrients. Ferroverdins and bagremycins are produced by the same BGCs, but the production of one or the other depends on the availability of iron in the environment. In this situation, using activators or inducers can be a useful strategy in eliciting BGCs and facilitating the identification of simultaneous NPs
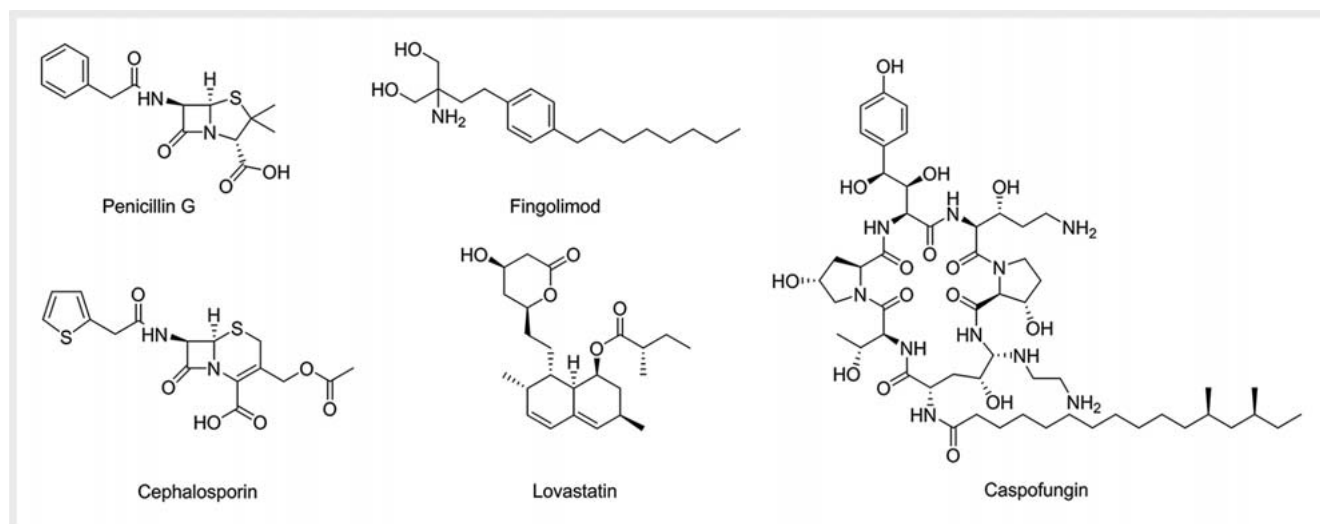
▶ **Fig. 3** Chemical structure of approved drugs inspired on fungal SMs.

(▶ **Fig. 2**) [67]. In the laboratory, production by *P. galatheae* of novel metabolites can be controlled by the addition of inducers, such as andrimid, to increase holomycin production, or trimethropin, to attenuate its production, indicating that activators or repressors can be a useful strategy in eliciting BGCs and facilitating natural product discovery (▶ **Fig. 2**) [61].

Another pathway to induce synergy of antimicrobials is through genetic engineering of modules of various clusters, leading to combinatorial chemistry and semsynthetic compounds with completely novel functions. Such explorations have already been reported; for example, the groups of genes involved in the biosynthesis of aminocoumarin, which is divided into modules, is able to be integrated in new functional BGCs [68]. Finally, Prof. Medema's group discusses some criteria for identifying when BGCs work together and promote synergism, based on genomic analyses: a) conservation across larger evolutionary time scales based on phylogenetic approach, b) proximity between BGCs within genome, c) co-regulated gene expression and d) assessment of the likeness between BGCs for related targets [66]. Without doubt, synergism of BGCs has begun to show new abilities provided by microorganisms that can be used for the discovery of not yet evaluated potent combinations.

Discovery of Bioactive Fungal Products Driven by Genome Mining

The fungal kingdom is the second most phylogenetically diverse and extensive of the tree of life. It is divided into twelve phyla which comprise about 200 orders [69] with an estimated 2.2–3.8 million species [70] and only a short slice (130 000) completely identified [71]. Fungi are crucial for preserving the ecological balance of any ecosystem, as they are, along with bacteria, nature's major decomposers [72]. Furthermore, these microorganisms have a massive economic impact as plant, crop, and animal pathogens, as well as in the food and pharmaceutical industries

[73]. Among them, strains from the Dikarya subkingdom (Phylum Ascomycota and Basidiomycota) are commonly investigated in drug discovery programs, owing to their capacity to produce a myriad of bioactive SMs (~ 20 000) [74] which often represent an amazing starting point for finding new drugs [75, 76]. Some fruitful examples include penicillin, considered the grandfather of antibiotics and the "wonder drug of World War II" [77], cephalosporins [78], the cholesterol lowering lovastatin (compactin) which gave origin to the statins family [79], myriocin, a precursor of fingolimod, the first oral treatment for multiple sclerosis [80], and lately, caspofungin (antifungal), a semisynthetic derivative of pneumocandin (▶ **Fig. 3**) [81]. However, up to the last two decades, the methodologies to identify and characterize NPs from any source relied on bioassay-guided programs, where the bioactivity of an extract and fractions is evaluated during the isolation of the bioactive metabolite [82]. This has led to the discovery of many outstanding drugs, principally antibacterial (penicillins, cephalosporins, quinolones, tetracylins, macrolides, aminoglycosides, among others) [83] and anticancer (taxol, camptothecin, vincristine, vinblastine, podophilotoxin, to mention a few) [84, 85]; however, the re-isolation of known compounds via this process is a bottleneck. This is demonstrated by the lack of new kinds of antibiotics in the market over the last 35 years. To overcome this, dereplication procedures [86] using LC-UV-MS/MS [87] and NMR [88, 89] have been developed. Importantly, the emergence of the omics era, the fast-paced evolution of bio and chemoinformatics tools to mine huge amounts of data, and the expansion of affordable whole-genome sequencing technologies have shifted the way of NPs research, by exploring and comparing DNA sequences with annotated repositories of BGCs [90].

In fungi, the mechanisms that activate their extraordinary metabolic capacity are associated with environmental changes and complex ecological networks, turning on several metabolic pathways, and consequently the biosynthesis of exciting and unusual backbones [91]. In fungal genomes, the genes involved in the biosynthesis of NPs are concatenated in BGCs [92]. These clus-
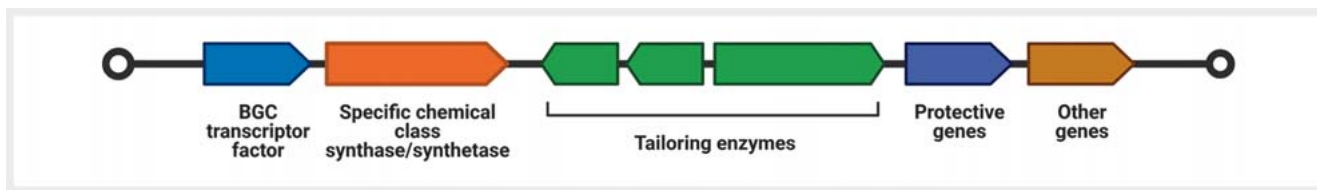
▶ **Fig. 4** Outline of a typical fungal BGC constructed by a transcription factor, a scaffold defining synthase or synthetase, decorative enzymes, detoxifying enzymes and other genes. Adapted from [93].
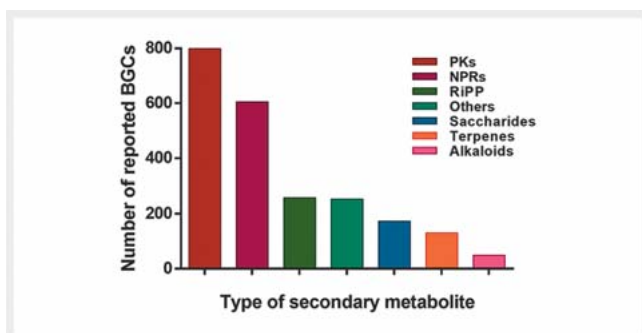
▶ **Fig. 5** Type of SMs with annotated BGCs. PKs: polyketides, NRPs: nonribosomal peptides, RiPPs: ribosomally synthesized and post-translationally modified peptides.

ters are mainly constituted by: i) a scaffold-defining synthase or synthetase (PK, terpene, NRP, etc.), that use building blocks to generate the main structure; ii) modifying enzymes such as oxidases, transferases, reductases, and acyltransferases, among others, that incorporate or transform functional groups; iii) specific or global transcription factors that positively or negatively regulate the expression of the BGC; iv) DNA segments encoding proteins to ameliorate the SM toxicity, for example, efflux pumps, transporters, detoxifying enzymes or double copies of the target enzyme; and v) other hypothetical genes with unknown activity (▶ **Fig. 4**) [82, 92, 93].

As described elsewhere, fungal and bacterial BGCs are quite similar; however, there are some differences that make the characterization of the former ever-challenging [94]. For example, bacterial BGCs are normally congregated in operons and their transcription occurs in a single mRNA, whereas in fungi, genes are separated by non-codifying DNA sequences, and are transcribed individually. According to MIBiG (Minimum Information about a Biosynthetic Gene cluster) database, up to May 2021 more than 1900 BGCs of several chemical classes have been annotated (▶ **Fig. 5** and **Table S1**, Supporting information) [95]. Among them, only 14.5% are of fungal origin, with a major representation of fungi from the Pezizomycotina subphylum (*Aspergillus, Penicillium, Fusarium, Alternaria, Phoma, Talaromyces, Beauveria*, and *Claviceps*; 180 BGCs, ▶ **Table 1**).

Robey et al. [96] recently demonstrated that fungi represent a huge reservoir of SM; however, less than 3% of the BGCs have been linked to a chemical product. This conclusion was based on the analysis of 1037 genomes from Dikarya and non-Dikarya taxa,

using antiSMASH, and exploration of the fungal chemical space. Results from this study predicted 36399 BGCs ranging from 5 to 220 kb, organized in 12067 gene cluster families (GCF), which are mainly species specific [96]. This finding could be supported by the fact that fungal chemodiversity is associated with three different evolutionary BGCs processes: i) heterologous and paralogous BGCs functional divergence (accumulation of molecular differences that lead to changes in function or phenotype), ii) BGC horizontal transfer and iii) *de novo* BGC assembly; contributing to the immense repertoire of fungal SM [92].

Some of the best achievements and latest research examples

Throughout history, fungi have only contributed with five antibiotics against bacteria (including mycophenolic acid, the previously cited *β*-lactams, fusidic acid, and retapamulin, a semisynthetic derivative of pleuromutilin), and two antifungals (griseofulvin and caspofungin) (▶ **Fig. 6**). Among them, fusidic acid (isolated in 1962 from *Acremonium fusidioides*, previously known as *Fusidium coccineum* [97], by Leo Pharma, Denmark) has been used over the past 45 years in Europe to treat skin and chronic bone and joint infections [98]. It inhibits protein synthesis by a new mechanism that involves binding to EF-G-GPD. This molecule is unique in its class; due to its novel structure and mode of action, microorganisms have not yet developed cross-resistance against it [98]. The most recently approved antibiotic from fungi is retapamulin [99]. This broad-spectrum antibiotic is a semisynthetic derivative of pleuromutilin, a diterpenoid isolated from *Pleurotus mutilus* in 1951 [100, 101] that displayed moderate activity towards Gram-positive bacteria. Retapamulin was approved by the FDA in 2007 to topically treat impetigo, a skin infection caused by *Staphylococcus aureus* or *Streptococcus pyogenes*. Mechanistically, it selectively inhibits protein synthesis by binding to the 50S subunit of the bacterial ribosome in a different manner than macrolides [102]. Like fusidic acid, microorganisms have not developed cross resistance against retapamulin.

Despite the potential of fungi as a source of chemical entities for drug development, and the crucial need for new antimicrobials for human use against multidrug resistant infections, most pharmaceutical companies have dropped their antimicrobial drug discovery programs, due to the low income and the severe regulatory frameworks imposed by regulatory agencies [103]. Nevertheless, since 2012, when IDSA requested the FDA and EMA (European Medicines Agency) to set more realistic and economic regulatory pathways for novel antibiotics registration and authorization, research attention has shifted back to microorganisms-de-

▶ **Table 1** Fungal genus from which a BGC has been annotated.

Genus	Annotated BGCs	Genus	Annotated BGCs	Genus	Annotated BGCs
Aspergillus	88	*Ustilago*	3	*Cordyceps*	1
Penicillium	28	*Cercospora*	2	*Dothistroma*	1
Fusarium	25	*Cladonia*	2	*Gelatoporia*	1
Alternaria	12	*Daldinia*	2	*Gliocladium*	1
Phoma	9	*Diaporthe*	2	*Hypholoma*	1
Talaromyces	7	*Hypomyces*	2	*Kluyveromyces*	1
Beauveria	6	*Hypoxylon*	2	*Lasiodiplodia*	1
Claviceps	5	*Leptosphaeria*	2	*Malbranchea*	1
Clostridium	4	*Nodulisporium*	2	*Neonectria*	1
Epichloe	4	*Paecilomyces*	2	*Neurospora*	1
Glarea	4	*Pestalotiopsis*	2	*Phyllosticta*	1
Monascus	4	*Sarocladium*	2	*Pochonia*	1
Parastagonospora	4	*Stachybotrys*	2	*Purpureocillium*	1
Acremonium	3	*Tolypocladium*	2	*Pyrenochaetopsis*	1
Botrytis	3	*Aureobasidium*	1	*Shiraia*	1
Byssochlamys	3	*Bipolaris*	1	*Sordaria*	1
Chaetomium	3	*Cochliobolus*	1	*Stereum*	1
Escovopsis	3	*Collariella*	1	*Strobilurus*	1
Metarhizium	3	*Cordana*	1	*Thermothelomyces*	1
Pyricularia	3	*Ustilago*	3	*Trichophyton*	1
Trichoderma	3	*Curvularia*	1		

rived NPs [74]. In this regard, academia has taken the lead and many fungal NPs with antibacterial activity, or even targeting several other bacterial survival mechanisms, have been recently discovered. The methodologies involved on such programs are diverse, e.g., bioassay-guided isolation, taxonomy-based isolation, targeted and untargeted metabolomics approaches applied to fungi grown under standard and modified conditions (One Strain Many Compounds, OSMAC), and processes that combine all those schemes have been successfully applied (Supporting Information).
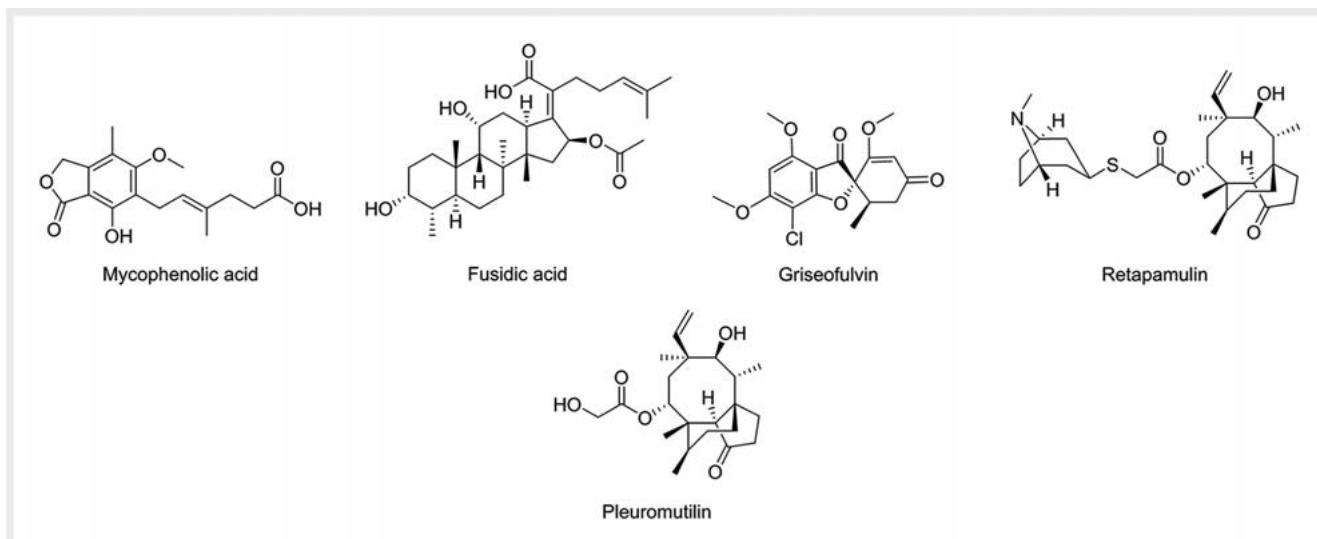
Fungal macrolides through genome mining coupled with heterologous expression

Recently, the Asai's group described a tri-partite methodology that includes synthetic biology, genome mining and heterologous expression to uncover the fungal machinery involved in the biosynthesis of macrolides. This approach led to the identification and validation of BGCs comprising a highly reducing polyketide synthase (HR-PKS) and a thioesterase in the genome of *Arthrinium phaeospermum* (*apmlA* and *apmlb*, respectively) [104]. Correspondingly, the heterologous expression of the built BGCs in *Aspergillus oryzae* led to the isolation of phaeospelides A and B, two macrolides of 34 and 32 members (▶ **Fig. 7**).

In a subsequent work, phylogenetic and network analysis of the HR-PKS gene exposed a family of macrolide BGCs, including a glycosylphosphatidylinositol–ethanolamine phosphate transferase (GPI-EPT) homolog. Genome mining of the GPI-EPT gene led to the identification of two BGCs, one from *Aspergillus kawachii* IFO 4308 (*akml* cluster) and one from *Colletotrichum incanum* MAFF 238 704 (*ciml* cluster). Heterologous expression of several combinations of genes that make up the *akml* and *ciml* clusters resulted in the production and isolation of several 24 (AKML A–D) and 22 (CIML A–D) member macrolides containing phosphoethanolamine or phosphocholine moieties (▶ **Fig. 7**). These macrolides were evaluated as antibacterial and antifungal agents in bioassay-guided programs, showing moderate activities [105].

Discovery of Bioactive NPs Ducts from Microbes by Metabologenomics

Microbial NPs have demonstrated their importance as drugs, in particular to treat cancer and infectious diseases. As mentioned above, high-throughput screening (HTS) processes and the creation of synthetic compounds libraries were not as successful as expected by the pharmaceutical industry to reach a great number of

▶ **Fig. 6** Chemical structure of antibiotics derived from fungi.
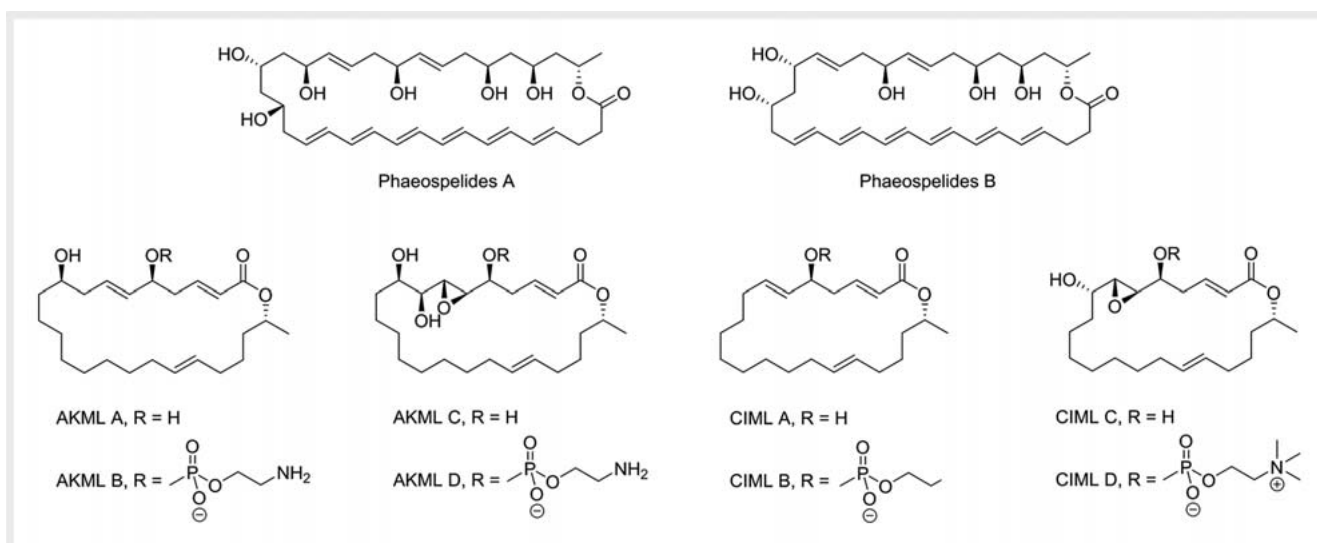
▶ **Fig. 7** Fungal macrolides identified by genome mining coupled with heterologous expression.

developmental drug candidates. Contrastingly, methods that rely on genomics, bioinformatics, and chemical analysis (metabolomics), or the combination of them (metabologenomics) are now recognized as better alternatives for the discovery of new drugs from nature (▶ **Fig. 8**). The term metabologenomics was officially introduced in 2016 and has one main advantage over the traditional search of potential drugs: it does not require prior knowledge of the molecular structure or bioactivity of the compounds, their detection by high-resolution mass spectrometry (HRMS) constitutes the basis for their experimental confirmation, thus avoiding silent BGCs [106].

Genome sequencing has become one of the most valuable tools to understand the potential of microbes to produce SMs.

Since the increase in knowledge of microbial BGCs and other important attributes within their genomes, efforts for large-scale sequencing have been done to accurately predict novel NPs BGCs, to improve the large-scale assembly, cloning, and expression of BGCs, and to increase the diversity of new compounds [107]. On the other hand, metabolomics data obtained by using liquid chromatography coupled with high-resolution mass spectrometry (LC-HRMS) has allowed the identification of hundreds of metabolites for each strain, with high sensitivity and selectivity. Furthermore, MS data analysis tools, such as principal component and molecular networking (MN) analysis, have been successfully applied to demonstrate the chemical space and diversity of metabolites in samples, which could be correlated to the functional phenotype
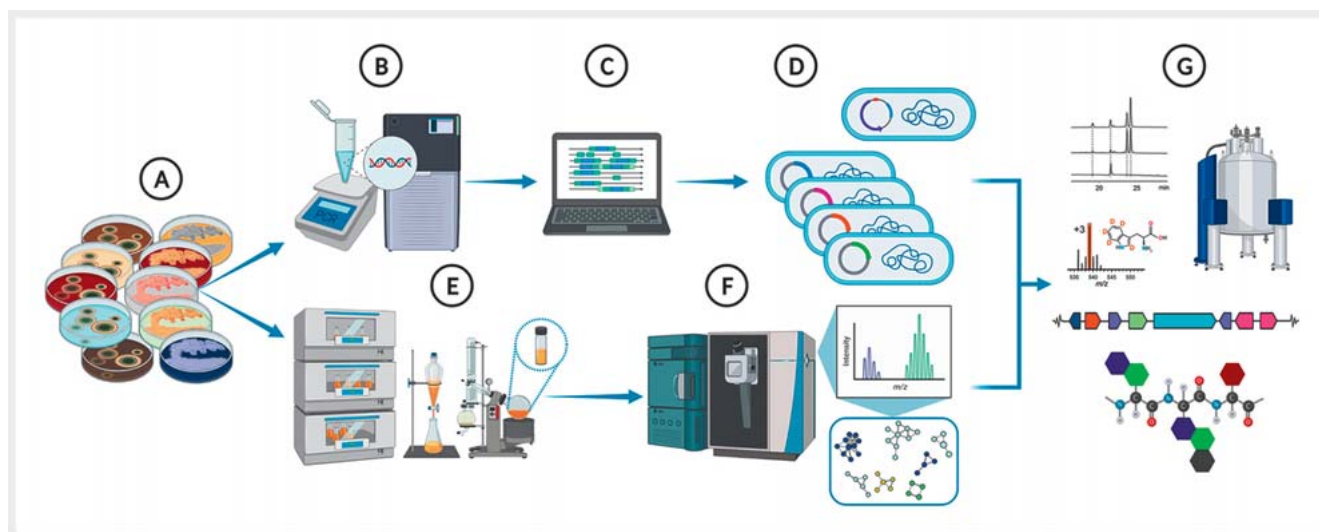
▶ **Fig. 8** Metabologenomics workflow: A. Microbes (fungi or bacteria) isolated from natural sources; B. DNA extraction and sequencing of specific genes or whole genome; C. Automated gene clusters prediction; D. Heterologous expression using CRISPR-Cas9 or artificial chromosomes technologies; E. Microbes are cultivated under different conditions and extracts are obtained; F. Metabolomic and molecular networking analysis using MS² data of the extracts; G. Correlation of genomics and metabolomics data to pair novel BGCs and metabolites. New NPs are usually confirmed using LC-MS data, stable isotope feeding experiments or by traditional isolation and structure elucidation (NMR, MS, CD, X-ray, etc.).

of the natural source [108]. When the genomic and metabolomic data are combined, simple correlation scores reveal the probability that observed NPs and a specific gene cluster family (GCF) are associated if they co-occur across multiple strains.

Certainly, the discovery of new specialized metabolites through the expression of their BGCs in heterologous hosts offers some advantages that complement traditional chemical studies in cultivable organisms [109, 110]. However, there are many limitations for the expression of BGCs in heterologous hosts, such as the use of unstable and difficult to manipulate artificial chromosomes, the incorporation of incomplete BGCs in cosmids and fosmids, and the lack of specific analytical methods to differentiate the heterologous expressed products form host's endogenous NPs. To solve the latter, mass-spectrometry (MS) metabolomics plays an important role to detect such metabolites and correlate them with natural or induced BGCs.

In the next paragraphs, we review a series of examples on metabologenomics studies for the discovery of bioactive NPs. Without a doubt, these comprehensive studies between microbiologists, structural and molecular biologists, chemists, and bioinformatic analysts are now part of the new era in drug discovery. In fact, a community initiative to systematically document links between metabolome and (meta)genome data was recently developed in the form of the Paired Omics Data Platform (PoDP; https://pairedomicsdata.bioinformatics.nl/).

## Some of the best achievements and latest research examples

Rimosamides

McClure and collaborators [111] developed a metabolite/gene cluster correlation platform to elucidate the biosynthetic origins of a new family of NPs produced by *Streptomyces rimosus* NRRL B-2659, the rimosamides (▶ **Fig. 9**). The core of these compounds

contains a depsipeptide bond at the point of bifurcation in their unusual, branched structures and are closely related to the detoxins. These groups of molecules are biosynthesized by a NRPS and a PKS. The study involved the identification of the co-occurrence of an unknown compound (negative dereplication in NPs databases) by LC-MS analysis of 178 extracts of actinobacteria, and specific gene cluster families on their annotated genomes. By doing so, 14 strains containing the compound matched six families of genes. Remarkably, almost all strains that encode for these genes were found within the *S. rimosus* clade. The proposed BGC for these compounds was established by detailed bioinformatic analysis. Then, heterologous expression in *S. lividans* 66 and LC-MS analyses of its organic extract were performed to demonstrate the production of the rimosamides. Finally, authors found that these compounds antagonize the antibiotic activity of blasticidin S in *Bacillus cereus* due to their structural similarity with the detoxines. Since blasticidin S has not been found in rimosamide-producing strains, these compounds could help them in the defense against other *Streptomycetes*.

Tambromycins

Following the same correlation approach described for the rimosamides, Goering and colleagues [112] uncovered a series of NPs and its BGCs in different actinomycetes, the tambromycins (▶ **Fig. 9**). These compounds are closely related to JBIR-34/35 and possess substructures of a hydroxylated and chlorinated indole, a methyl-oxazoline, and 2-methyl-serine. Successful correlation in six different strains of *Streptomyces* was demonstrated by extensive BGC analysis and LC-MS identification of the matching ions. The strain F-4474 was selected for fermentation and purification of the metabolite because it showed the highest intensity of the tambromycin ion.
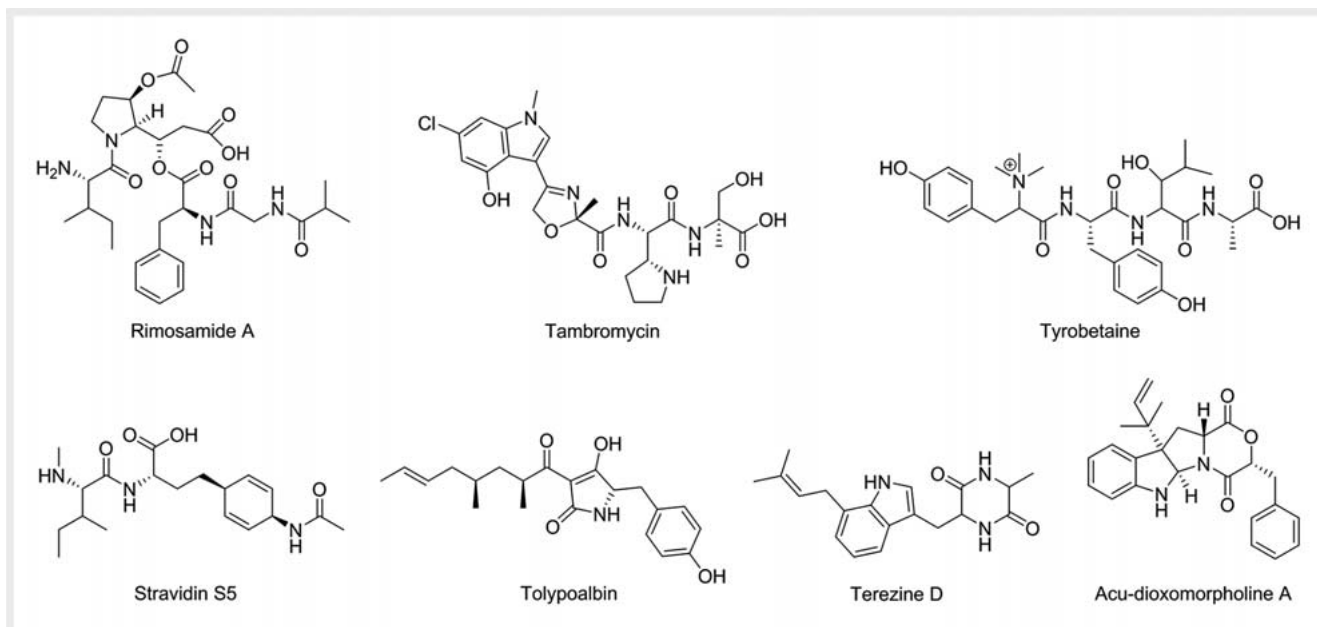
▶ **Fig. 9** Compounds identified by metabologenomics.

Functional annotation of genes in the tambromycins BGC revealed the presence of a tryptophan halogenase, a tryptophan aldolase, an aldehyde dehydrogenase, a methylation domain, and an alanine-hydroxymethyl transferase, consistent with the proposed biosynthesis for these compounds. In addition, feeding experiments with stable isotope labeled amino acids were carried out to confirm the biosynthesis of tambromycins. Unfortunately, the main compound (tambromycin) was inactive against a set of clinically important bacteria but showed antiproliferative activity against different cancerous B- and T-cell lines.

Tyrobetaines

Parkinson and colleagues [106] discovered a novel class of non-ribosomal peptides (NRPs) with an unusual trimethylammonium tyrosine residue, the tyrobetaines, produced by *Streptomyces* sp. NRRL WC-3773 (▶ **Fig. 9**). Then, MN analysis (clustering of molecular features based on MS and MS² fragmentation patterns) was used to improve the metabologenomics correlation results. The BGC of the tyrobetaines was predicted using BLAST, and the boundaries for the BGC were estimated by comparing the BGCs from multiple tyrobetaine producing strains. For this, domains of an *N*-methyltransferase, a phosphopantetheinyl transferase, a P450 monooxygenase, among others, were consistent with the proposed biosynthesis of these NPs. As for the rimosamides, *S. lividans* 66 was used for heterologous expression of these NRPs. Feeding experiments with stable isotope labeled L-tyrosine, L-leucine, L-alanine, methionine, 3-chlorotyrosine coupled with MS analysis confirmed the biosynthesis of the tyrobetaines. Finally, tyrobetaine and its chlorinated derivative were tested against ESKAPE pathogens (*Enterococcus faecalis* ATCC 19433, *S. aureus* ATCC 29213, *Klebsiella pneumoniae* ATCC 27736, *Acinetobacter baumannii* ATCC 19606, *Pseudomonas aeruginosa* PAO1, and

Escherichia coli ATCC 25922) and human lung cancer cell line A549, with no detectable activity at the assayed concentrations.

Stravidins

Stravidins were discovered from *Streptomyces avidinii* as components of an interesting antibiotic complex [113]. These peptide antibiotics synergistically achieve biotin deficiency in target microbes. The unique amino acid monomer amiclenomycin (Acm) inhibits the 7,8-diaminopelargonic acid (DAPA) aminotransferase (BioA),which is essential for survival and persistence of mycobacteria. Interestingly, other Acm-containing peptides have shown activity against Gram-negative bacteria.

Using the metabologenomics data set from a collection of *Streptomyces* strains, Montaser and Kelleher identified the NP/BGC correlation in nine strains. In particular, the strain NRRL S-98 showed a specific ion and fragments characteristic of the stravidins and correlated to an orphan BGC. This compound was isolated and characterized as stravidin S5 (▶ **Fig. 9**). BLAST-P analysis revealed additional members of this GCF and allowed to propose the biosynthetic pathway for these compounds. Finally, heterologous expression of the stravidins BGC in *S. lividans* 66 using the CRISPR-Cas9 technology, was successfully accomplished and confirmed by UPLC-HRMS/MS analysis.

4-Hydroxy pyridones

Fungal 4-hydroxy pyridones are polyketide-NRP compounds with diverse structures and biological activities [114]. The biosynthesis of these compounds is initiated by the PKS-NRPS and a *trans*-acting enoyl reductase (ER) to yield a pentacyclic intermediate, acryltetramic acid. This compound undergoes an oxidative ring expansion by a cytochrome P450 enzyme to give the six-membered skeleton of 4-hydroxy-2-pyridone, followed by a variety of modifications by tailoring enzymes.

Genome mining of in-house strains led to the identification of the 4-hydroxy-pyridones BGC in the endophyte *Tolypocladium* sp. 49Y isolated from the leaf of *Acorus tatarinowii*. Then, the genes were introduced into a heterologous expression host, *Aspergillus oryzae* NSAR1 (AO1 strain). As predicted, a new compound, toly-poalbin (▶ **Fig. 9**), was detected by HPLC analysis in the host culture extract. A homologous gene, also present in the BGC, was then introduced into the AO1 strain to establish a three-gene strain (AO2). The latter produced additional new 4-hydroxy-pyridones. Some of these compounds showed significant antifungal effects against *Cryptococcus neoformans* in bioassay-guided experiments.

Terezine D

As mentioned, there are technological challenges in the insertion of large DNA fragments in homologous or heterologous systems for the expression of microbial NPs. Specifically, for the expression of fungal BGC, fungal artificial chromosome (FAC) expression vectors have been used with some success [115] In this context, the unbiased Random Shear BAC technology coupled with an autonomous fungal replicating element, AMA1, was applied to increase the transformation efficiency for the expression of FACs in *A. nidulans* [115]. The construction of the unbiased shuttle BAC library of *A. terreus* was possible since the genome of this fungus was fully sequenced and extensively annotated (56 BGC). The confirmation of the successful transformation of *A. nidulans* was established by LC-HRMS-MS/MS analysis of the transformant extracts, followed by the production of terezine D, a stable intermediate in the astechrome biosynthesis (▶ **Fig. 9**).

Acu-dioxomorpholines

Robey and collaborators used FAC-MS technology to elucidate the biosynthetic pathway for a rare group of metabolites, the acu-dioxomorpholines (▶ **Fig. 9**) produced by *Aspergillus aculeatus* [116]. The fungal genome was randomly sheared and ~ 100 kb fragments were inserted into FACs, *A. nidulans*/*E. coli* shuttle vectors. The FACs with the BGCs were then transformed into the heterologous host *A. nidulans*. MS-based metabolomics allowed the detection of SMs and associated them with each FAC, assessing the effect of BGC mutants. The biosynthesis of the acu-dioxomorpholines was proved by stable isotope feeding experiments. As expected, [D$_5$-indole]-Trp and [D$_5$-phenyl]-Phe were incorporated in the biosynthetic products.

Discovery of Bioactive Plant Products Driven by Genome Mining

Plants have played a crucial role in the establishment and development of human societies, as some have been used empirically for thousands of years to treat a plethora of diseases [117]. Historically, plants were used as crude medicine or in preparations, [117–119] and later on (19th century) they were implemented as starting material to isolate bioactive compounds, commencing with the separation of morphine from *Papaver sofmniferum* in 1817 [117,120]. Since then, more than 133 000 bioactive NPs from higher plants have been discovered [74]. Several ap-
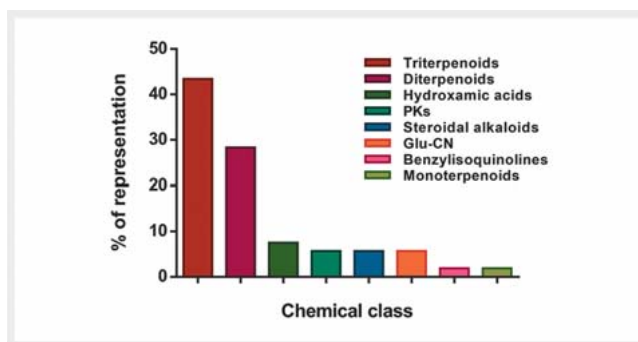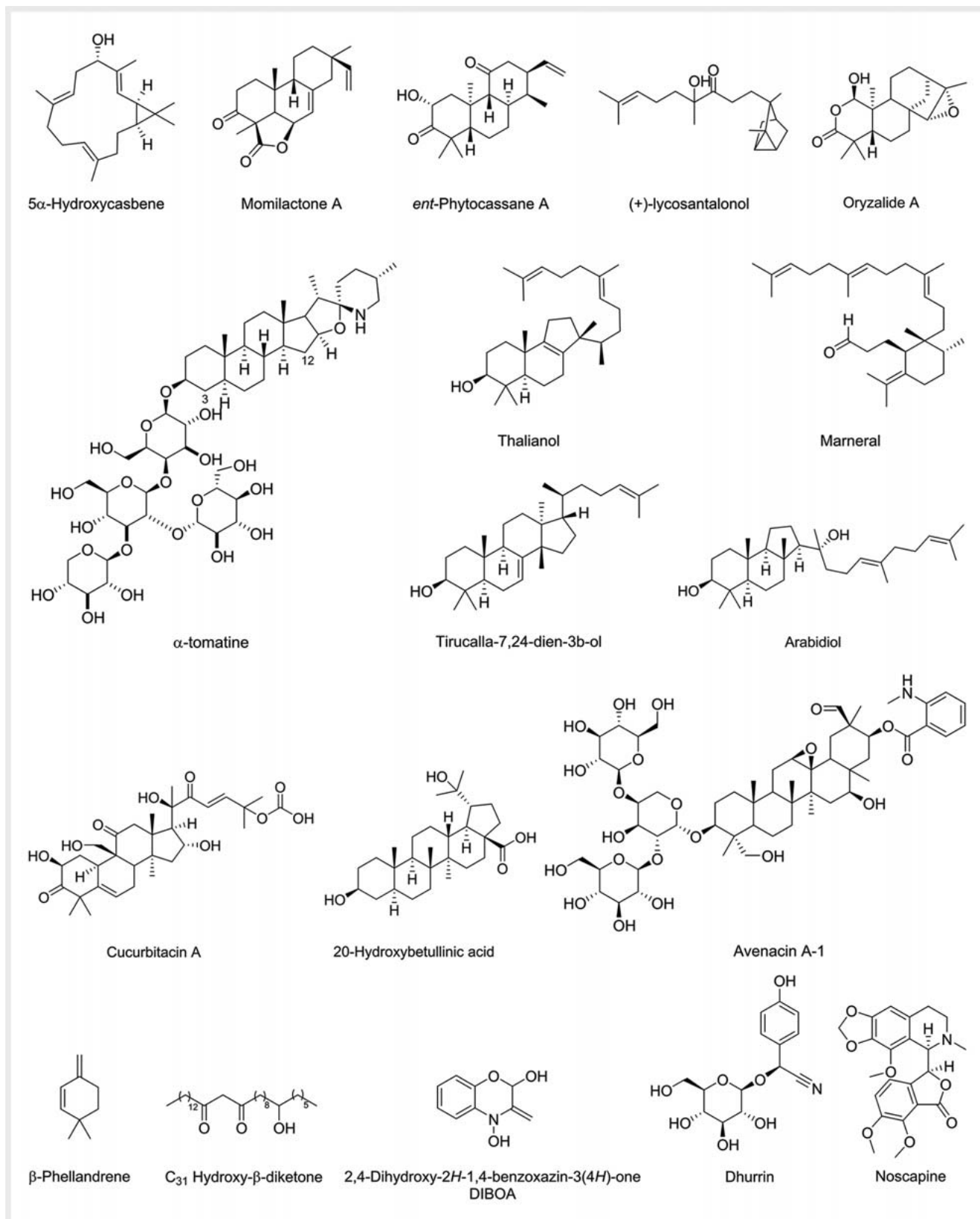

▶ **Fig. 10** Percentage of annotated plant BGCs by chemical class.

proaches, including conventional chromatographic techniques, [121] bioassay-guided isolation, MS and NMR metabolomics-guided isolation, [122] and recently, multi-informative methodologies, [123] have been followed. Unfortunately, as of now, there is a major gap in the application of genome mining methodologies to understand the biosynthesis of plant derived NPs, [124] and an even greater one to integrate such techniques in the discovery of new plant derived bioactive NPs. The details contributing to this lag are linked to the biosynthetic pathway through which a plant SM is biologically synthesized (sequence of enzymatic reactions), and the fact that genes participating in the assembly of a molecule are commonly scattered along the genome. Conversely, those found in microorganisms are organized in BGCs and much easier to detect. Plant BGCs range from ~ 35 to hundreds of kb, and congregate from three to 10 genes, [125,126] including those responsible for the first step in the biosynthesis of a SM (generally, but not always), and at least two coding peripheral enzymes [127]. According to Nützmann et al., BGCs can be grouped in four types. a) Archetypal cluster: it is compact and contains all genes required for the biosynthesis of SMs; b) core cluster + peripheral genes: most of the genes are contiguous but some are located in diffident locus (*cis* or *trans*, linked or unlinked with the core cluster, respectively); c) core clusters + satellite subgrups: the majority of genes for the first committed step are organized in a core cluster, but some genes (two or three) are present as satellites elsewhere in the genome; and d) core cluster + peripheral genes: the main cluster encodes for some tailoring enzymes and the gene encoding the scaffold of the NP is elsewhere in the genome (*cis* or *trans*) [127].

The boom of plant BGCs started about 24 years ago with the discovery of the cyclic hydroxamic acid 2,4-dihydroxy-1,4-benzoxazin-3-one (DIBOA) BGC, in Graminaea, which is assembled by five genes, Bx1 through Bx5, [128] and the antibacterial triterpenoids avenacins in oat (*Avena* spp.) [127,129]. Since then, more than fifty BGCs have been revealed in several species across the phylogenetic three of the Plantae kingdom [130].

Among the plant BGCs reported so far, triterpenoids are the most exploited, with about 43.4% (23 SM) of the universe with structures covering a limited but informative portion of the chemical space. Next, diterpenoids contribute with 28.3% (15 SM), hydroxamic acids (7.5%, four SM), polyketides, steroidal alkaloids and cyanogenic glycosides contribute with 5.7% each (three SM

▶ **Fig. 11** Plant-derived compounds for which a BGC has been identified.

The figure shows the following labeled chemical structures:
- 5α-Hydroxycasbene
- Momilactone A
- *ent*-Phytocassane A
- (+)-lycosantalonol
- Oryzalide A
- α-tomatine
- Thalianol
- Marneral
- Tirucalla-7,24-dien-3b-ol
- Arabidiol
- Cucurbitacin A
- 20-Hydroxybetullinic acid
- Avenacin A-1
- β-Phellandrene
- C_{31} Hydroxy-β-diketone
- 2,4-Dihydroxy-2*H*-1,4-benzoxazin-3(4*H*)-one DIBOA
- Dhurrin
- Noscapine

individually), and benzylisokilolines alkaloids and monoterpenoids with 1.9% (one CE representing each chemical class) (▶ **Fig. 10**) [125, 127, 131]. ▶ **Fig. 11** shows the most relevant structure representative of every chemical class; for a deeper view of the structures reported, readers are referred to Nützmann et al. [127] and Polturak et al. [131].

Biologically, plant SM encoded in a BGC play essential roles in chemical ecology, especially in interactions with competing plants, herbivores, insects, bacteria, and fungi, demonstrating that they are only biosynthesized under biotic stress conditions, as defense mechanisms[113]. Thus, the study of BGCs in plants opens the door to better understand and elucidate complex biosynthetic pathways, as well as the discovery and exploitation of novel SMs that could be of great interest for the agrochemical and pharmaceutical industries, as they could be potentially developed in new biocides or candidate drugs to treat infectious diseases caused by human pathogens.

Perspectives

The limitations of genome mining are due to information and database availability on gene functions and biosynthetic pathways. Recent approaches suggest pathways to circumvent or alleviate these hurdles. Omics studies to obtain data on the genomes, transcriptomes, and metabolomes of the same samples can improve the coverage and bridge the gaps in knowledge by connecting molecules with their BGC of origin. Such explorative studies of biologically hyper diverse geographical regions or of those with extreme conditions can complete our knowledge of the pre-existing chemical diversity of the tree of life. In addition, genomic data analyses provide possibilities to design suitable growth conditions in the laboratory for formerly not cultivable organisms, and thus allow for their metabolites to be studied. Finally, genome editing capabilities have been greatly developed and have simplified the exploration of cryptic BSCs, especially with the advent of CRISPR-Cas technologies [55]. While the number of molecules uncovered by genome mining has increased, with some even indicating the presence of completely novel chemical classes or genetic entities, there is very little funding and public interest to take these discoveries to the clinical stage and confirm their suitability in the clinical/medical/industrial practices. Screening these by cheaper/ faster means, such as artificial intelligence/machine learning strategies, high throughput screenings by *in vitro* testing in organoids, has been proposed as ways forward to decrease some of the risks of failure further in their validation.

Supporting Information

Traditional approaches for antimicrobials discovery from fungi and fungal natural products with minimum information about biosynthetic gene cluster (**Table S1**) are available in the Supporting Information.

Contributors' Statement

Drafting the manuscript, design of the review, and critical revision of the manuscript: J. Rivera-Chávez, C. D. Ceapa, and M. Figueroa. All authors read and approved the final version of the manuscript.

Acknowledgements

Conflict of Interest

The authors declare that they have no conflict of interest.

References

[1] Nieuwlaat R, Mbuagbaw L, Mertz D, Burrows LL, Bowdish DME, Moja L, Wright GD, Schunemann HJ. Coronavirus disease 2019 and antimicrobial resistance: Parallel and interacting health emergencies. Clin Infect Dis 2021; 72: 1657–1659

[2] Rawson TM, Ming D, Ahmad R, Moore LSP, Holmes AH. Antimicrobial use, drug-resistant infections and COVID-19. Nat Rev Microbiol 2020; 18: 409–410

[3] Lobie TA, Roba AA, Booth JA, Kristiansen KI, Aseffa A, Skarstad K, Bjoras M. Antimicrobial resistance: A challenge awaiting the post-COVID-19 era. Int J Infect Dis 2021; 111: 322–325

[4] Rizvi SG, Ahammad SZ. COVID-19 and antimicrobial resistance: A cross-study. Sci Total Environ 2021; 807: 150873

[5] O'Neill J. Antimicrobial Resistance: Tackling a Crisis for the Health and Wealth of Nations. London, UK: Wellcome Collection. Attribution 4.0 International (CC BY 4.0); 2014

[6] Piddock LJV. The global antibiotic research and development partnership (GARDP): Researching and developing new antibiotics to meet global public health needs. MedChemComm 2019; 10: 1227–1230

[7] Williams AN, Sorout N, Cameron AJ, Stavrinides J. The integration of genome mining, comparative genomics, and functional genetics for biosynthetic gene cluster identification. Front Genet 2020; 11: 600116

[8] Bentley SD, Chater KF, Cerdeño-Tárraga AM, Challis GL, Thomson N, James KD, Harris DE, Quail MA, Kieser H, Harper D. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3 (2). Nature 2002; 417: 141–147

[9] Ikeda H, Ishikawa J, Hanamoto A, Shinose M, Kikuchi H, Shiba T, Sakaki Y, Hattori M, Ōmura S. Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. Nat Biotechnol 2003; 21: 526–531

[10] Baltz RH. Molecular beacons to identify gifted microbes for genome mining. J Antibiot 2017; 70: 639–646

[11] Navarro-Muñoz JC, Selem-Mojica N, Mullowney MW, Kautsar SA, Tryon JH, Parkinson EI, De Los Santos EL, Yeong M, Cruz-Morales P, Abubucker S. A computational framework to explore large-scale biosynthetic diversity. Nat Chem Biol 2020; 16: 60–68

[12] Adamek M, Alanjary M, Ziemert N. Applied evolution: Phylogeny-based approaches in natural products research. Nat Prod Rep 2019; 36: 1295–1312

[13] Yan X, Ge H, Huang T, Hindra, Yang D, Teng Q, Crnovčić I, Li X, Rudolf JD, Lohman JR. Strain prioritization and genome mining for enediyne natural products. mBio 2016; 7: e02104–02116

[14] Caldera EJ, Chevrette MG, McDonald BR, Currie CR. Local adaptation of bacterial symbionts within a geographic mosaic of antibiotic coevolution. Appl Environ Microbiol 2019; 85: e01580–e01619

[15] Mullis MM, Rambo IM, Baker BJ, Reese BK. Diversity, ecology, and prevalence of antimicrobials in nature. Front Microbiol 2019; 10: 2518

[16] Hindra, Yang D, Teng Q, Dong LB, Crnovcic I, Huang T, Ge H, Shen B. Genome mining of *Streptomyces mobaraensis* DSM40847 as a bleomycin

producer providing a biotechnology platform to engineer designer bleo-mycin analogues. Org Lett 2017; 19: 1386–1389

[17] Scherlach K, Hertweck C. Mining and unearthing hidden biosynthetic potential. Nat Commun 2021; 12: 3864

[18] Quinn GA, Banat AM, Abdelhameed AM, Banat IM. Streptomyces from traditional medicine: Sources of new innovations in antibiotic discovery. J Med Microbiol 2020; 69: 1040–1048

[19] Kautsar SA, Blin K, Shaw S, Navarro-Muñoz JC, Terlouw BR, van der Hooft JJJ, van Santen JA, Tracanna V, Suarez Duran HG, Pascal Andreu V, Selem-Mojica N, Alanjary M, Robinson SL, Lund G, Epstein SC, Sisto AC, Charkoudian LK, Collemare J, Linington RG, Weber T, Medema MH. MIBiG 2.0: A repository for biosynthetic gene clusters of known function. Nucleic Acids Res 2020; 48: D454–D458

[20] Lasch C, Gummerlich N, Myronovskyi M, Palusczak A, Zapp J, Luzhetskyy A. Loseolamycins: A group of new bioactive alkylresorcinols produced after heterologous expression of a type III PKS from *Micromonospora endolithica*. Molecules 2020; 25: 4594

[21] Styczynski M, Rogowska A, Gieczewska K, Garstka M, Szakiel A, Dziewit L. Genome-based insights into the production of carotenoids by Antarctic bacteria, *Planococcus* sp. ANT_H30 and *Rhodococcus* sp. ANT_H53B. Molecules 2020; 25: 4357

[22] Salwan R, Sharma V. Molecular and biotechnological aspects of second-ary metabolites in actinobacteria. Microbiol Res 2020; 231: 126374

[23] Belknap KC, Park CJ, Barth BM, Andam CP. Genome mining of biosyn-thetic and chemotherapeutic gene clusters in *Streptomyces* bacteria. Sci Rep 2020; 10: 2003

[24] Vicente CM, Thibessard A, Lorenzi JN, Benhadj M, Hôtel L, Gacemi-Kirane D, Lespinet O, Leblond P, Aigle B. Comparative genomics among closely related *Streptomyces* strains revealed specialized metabolite biosynthetic gene cluster diversity. Antibiotics 2018; 7: 86

[25] Sivakala KK, Gutiérrez-García K, Arul Jose P, Thinesh T, Anandham R, Barona-Gómez F, Sivakumar N. Desert environments facilitate unique evolution of biosynthetic potential in *Streptomyces*. Molecules 2021; 26: 588

[26] Xu L, Ye KX, Dai WH, Sun C, Xu LH, Han BN. Comparative genomic in-sights into secondary metabolism biosynthetic gene cluster distribu-tions of marine *Streptomyces*. Mar Drugs 2019; 17: 498

[27] Matroodi S, Siitonen V, Baral B, Yamada K, Akhgari A, Metsä-Ketelä M. Genotyping-guided discovery of persiamycin A from sponge-associated halophilic *Streptomonospora* sp. PA3. Front Microbiol 2020; 11: 1237

[28] Wang X, Zhou H, Chen H, Jing X, Zheng W, Li R, Sun T, Liu J, Fu J, Huo L. Discovery of recombinases enables genome mining of cryptic biosyn-thetic gene clusters in *Burkholderiales* species. Proc Natl Acad Sci U S A 2018; 115: E4255–E4263

[29] Zhang W, Ji R, Liu J, Pan Y, Wu LF, Lin W. Two metagenome-assembled genome sequences of magnetotactic bacteria in the order *Magnetococ-cales*. Microbiol Resour Announc 2020; 9: e00363-00420

[30] Dose B, Niehs SP, Scherlach K, Shahda S, Flórez LV, Kaltenpoth M, Hert-weck C. Biosynthesis of sinapigladioside, an antifungal isothiocyanate from *Burkholderia* symbionts. ChemBioChem 2021; 22: 1920–1924

[31] Marter P, Huang S, Brinkmann H, Pradella S, Jarek M, Rohde M, Bunk B, Petersen J. Filling the gaps in the cyanobacterial tree of life-metagenome analysis of *Stigonema ocellatum* DSM 106950, *Chlorogloea purpurea* SAG 13.99 and *Gomphosphaeria aponina* DSM 107014. Genes 2021; 12: 389

[32] Romero MF, Gallego D, Lechuga-Jiménez A, Martínez JF, Barajas HR, Hayano-Kanashiro C, Peimbert M, Cruz-Ortega R, Molina-Freaner FE, Alcaraz LD. Metagenomics of mine tailing rhizospheric communities and its selection for plant establishment towards bioremediation. Micro-biol Res 2021; 247: 126732

[33] Letzel A–C, Pidot SJ, Hertweck C. Genome mining for ribosomally syn-thesized and post-translationally modified peptides (RiPPs) in anaerobic bacteria. BMC Genomics 2014; 15: 983

[34] Travin DY, Bikmetov D, Severinov K. Translation-targeting RiPPs and where to find them. Front Genet 2020; 11: 226

[35] Purushothaman M, Sarkar S, Morita M, Gugger M, Schmidt EW, Morina-ka BI. Genome-mining-based discovery of the cyclic peptide tolypamide and TolF, a Ser/Thr forward *O*-prenyltransferase. Angew Chem Int Ed 2021; 60: 8460–8465

[36] Carro L, Castro JF, Razmilic V, Nouioui I, Pan C, Igual JM, Jaspars M, Goodfellow M, Bull AT, Asenjo JA. Uncovering the potential of novel mi-cromonosporae isolated from an extreme hyper-arid Atacama Desert soil. Sci Rep 2019; 9: 4678

[37] Morgan KT, Zheng J, McCafferty DG. Discovery of six ramoplanin family gene clusters and the lipoglycodepsipeptide chersinamycin. ChemBio-Chem 2021; 22: 176–185

[38] Dunbar KL, Büttner H, Molloy EM, Dell M, Kumpfmüller J, Hertweck C. Genome editing reveals novel thiotemplated assembly of polythioamide antibiotics in anaerobic bacteria. Angew Chem Int Ed 2018; 57: 14080–14084

[39] Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema MH, Weber T. AntiSMASH 6.0: Improving cluster detection and comparison capabilities. Nucleic Acids Res 2021; 49: W29–W35

[40] Skinnider MA, Johnston CW, Gunabalasingam M, Merwin NJ, Kieliszek AM, MacLellan RJ, Li H, Ranieri MRM, Webster ALH, Cao MPT, Pfeifle A, Spencer N, To QH, Wallace DP, Dejong CA, Magarvey NA. Comprehen-sive prediction of secondary metabolite structure and biological activity from microbial genome sequences. Nat Commun 2020; 11: 6058

[41] van Heel AJ, de Jong A, Montalban-Lopez M, Kok J, Kuipers OP. BAGEL3: Automated identification of genes encoding bacteriocins and (non-)bac-tericidal posttranslationally modified peptides. Nucleic Acids Res 2013; 41: W448–W453

[42] Agrawal P, Khater S, Gupta M, Sain N, Mohanty D. RiPPMiner: A bioinfor-matics resource for deciphering chemical structures of RiPPs based on prediction of cleavage and cross-links. Nucleic Acids Res 2017; 45: W80–W88

[43] Medema MH, Kottmann R, Yilmaz P, Cummings M, Biggins JB, Blin K, de Bruijn I, Chooi YH, Claesen J, Coates RC, Cruz-Morales P, Duddela S, Dusterhus S, Edwards DJ, Fewer DP, Garg N, Geiger C, Gomez-Escribano JP, Greule A, Hadjithomas M, Haines AS, Helfrich EJ, Hillwig ML, Ishida K, Jones AC, Jones CS, Jungmann K, Kegler C, Kim HU, Kotter P, Krug D, Masschelein J, Melnik AV, Mantovani SM, Monroe EA, Moore M, Moss N, Nutzmann HW, Pan G, Pati A, Petras D, Reen FJ, Rosconi F, Rui Z, Tian Z, Tobias NJ, Tsunematsu Y, Wiemann P, Wyckoff E, Yan X, Yim G, Yu F, Xie Y, Aigle B, Apel AK, Balibar CJ, Balskus EP, Barona-Gomez F, Bechthold A, Bode HB, Borriss R, Brady SF, Brakhage AA, Caffrey P, Cheng YQ, Clardy J, Cox RJ, De Mot R, Donadio S, Donia MS, van der Donk WA, Dorrestein PC, Doyle S, Driessen AJ, Ehling-Schulz M, Entian KD, Fischbach MA, Gerwick L, Gerwick WH, Gross H, Gust B, Hertweck C, Hofte M, Jensen SE, Ju J, Katz L, Kaysser L, Klassen JL, Keller NP, Kormanec J, Kuipers OP, Kuzuyama T, Kyrpides NC, Kwon HJ, Lautru S, Lavigne R, Lee CY, Linquan B, Liu X, Liu W, Luzhetskyy A, Mahmud T, Mast Y, Mendez C, Metsa-Ketela M, Micklefield J, Mitchell DA, Moore BS, Moreira LM, Muller R, Neilan BA, Nett M, Nielsen J, O'Gara F, Oikawa H, Osbourn A, Osburne MS, Ostash B, Payne SM, Pernodet JL, Petricek M, Piel J, Ploux O, Raaijmakers JM, Salas JA, Schmitt EK, Scott B, Seipke RF, Shen B, Sherman DH, Sivonen K, Smanski MJ, Sosio M, Stegmann E, Sussmuth RD, Tahlan K, Thomas CM, Tang Y, Truman AW, Viaud M, Walton JD, Walsh CT, Weber T, van Wezel GP, Wilkinson B, Willey JM, Wohlleben W, Wright GD, Ziemert N, Zhang C, Zotchev SB, Breitling R, Takano E, Glockner FO. Minimum information about a biosynthetic gene cluster. Nat Chem Biol 2015; 11: 625–631

[44] Shin J, Park S, Lee JS, Lee EJ, Youn HD. Complete genome sequence and comparative analysis of *Streptomyces seoulensis*, a pioneer strain of nickel superoxide dismutase. Genes Genomics 2020; 42: 273–281

[45] Marques-Bastos SLS, Coelho MLV, de Sousa Santos IN, Farias FM, Francisco MS, Albano RM, Alviano CS, de Freire Bastos MdC. Draft ge-nome sequence of the producer strain of aureocin 4181, an antimicro-

bial peptide with antagonistic activity against multidrug-resistant staphylococci. J Glob Antimicrob Resist 2020; 23: 331–333

[46] Marques-Bastos SLS, Coelho MLV, Ceotto-Vigoder H, Fagundes PC, Almeida GS, Brede DA, Nes IF, de Paiva Brito MAV, de Freire Bastos MdC. Molecular characterization of aureocin 4181: A natural N-formylated aureocin A70 variant with a broad spectrum of activity. Braz J Microbiol 2020; 51: 1527–1538

[47] Vougas K, Sakellaropoulos T, Kotsinas A, Foukas GRP, Ntargaras A, Koinis F, Polyzos A, Myrianthopoulos V, Zhou H, Narang S. Machine learning and data mining frameworks for predicting drug response in cancer: An overview and a novel in silico screening process based on association rule mining. Pharmacol Ther 2019; 203: 107395

[48] Tietz JI, Schwalen CJ, Patel PS, Maxson T, Blair PM, Tai H-C, Zakai UI, Mitchell DA. A new genome-mining tool redefines the lasso peptide biosynthetic landscape. Nat Chem Biol 2017; 13: 470–478

[49] Kong L, Chen Y, Xu F, Xu M, Li Z, Fang J, Zhang L, Pian C. Mining influential genes based on deep learning. BMC Bioinform 2021; 22: 1–12

[50] Skinnider MA, Johnston CW, Edgar RE, Dejong CA, Merwin NJ, Rees PN, Magarvey NA. Genomic charting of ribosomally synthesized natural product chemical space facilitates targeted mining. PNAS 2016; 113: E6343–E6351

[51] Hao T, Xie Z, Wang M, Liu L, Zhang Y, Wang W, Zhang Z, Zhao X, Li P, Guo Z. An anaerobic bacterium host system for heterologous expression of natural product biosynthetic gene clusters. Nat Commun 2019; 10: 3665

[52] Culp EJ, Waglechner N, Wang W, Fiebig-Comyn AA, Hsu YP, Koteva K, Sychantha D, Coombes BK, Van Nieuwenhze MS, Brun YV. Evolution-guided discovery of antibiotics that inhibit peptidoglycan remodelling. Nature 2020; 578: 582–587

[53] Zou D, Maina SW, Zhang F, Yan Z, Ding L, Shao Y, Xin Z. Mining new plipastatins and increasing the total yield using CRISPR/Cas9 in genome-modified Bacillus subtilis 1A751. J Agric Food Chem 2020; 68: 11358–11367

[54] Wang G, Zhao Z, Ke J, Engel Y, Shi YM, Robinson D, Bingol K, Zhang Z, Bowen B, Louie K, Wang B, Evans R, Miyamoto Y, Cheng K, Kosina S, De Raad M, Silva L, Luhrs A, Lubbe A, Hoyt DW, Francavilla C, Otani H, Deutsch S, Washton NM, Rubin EM, Mouncey NJ, Visel A, Northen T, Cheng JF, Bode HB, Yoshikuni Y. CRAGE enables rapid activation of biosynthetic gene clusters in undomesticated bacteria. Nat Microbiol 2019; 4: 2498–2510

[55] Ke J, Robinson D, Wu ZY, Kuftin A, Louie K, Kosina S, Northen T, Cheng JF, Yoshikuni Y. CRAGE-CRISPR facilitates rapid activation of secondary metabolite biosynthetic gene clusters in bacteria. Cell Chem Biol 2021; 29: 1–15

[56] Tang X, Li J, Millán-Aguiñaga N, Zhang JJ, O'Neill EC, Ugalde JA, Jensen PR, Mantovani SM, Moore BS. Identification of thiotetronic acid antibiotic biosynthetic pathways by target-directed genome mining. ACS Chem Biol 2015; 10: 2841–2849

[57] Stokes JM, Yang K, Swanson K, Jin W, Cubillos-Ruiz A, Donghia NM, MacNair CR, French S, Carfrae LA, Bloom-Ackermann Z, Tran VM, Chiappino-Pepe A, Badran AH, Andrews IW, Chory EJ, Church GM, Brown ED, Jaakkola TS, Barzilay R, Collins JJ. A deep learning approach to antibiotic discovery. Cell 2020; 180: 688–702

[58] Melo MCR, Maasch J, de la Fuente-Nunez C. Accelerating antibiotic discovery through artificial intelligence. Commun Biol 2021; 4: 1050

[59] Alanjary M, Medema MH. Mining bacterial genomes to reveal secret synergy. J Biol Chem 2018; 293: 19996–19997

[60] Sengupta S, Chattopadhyay MK, Grossart H–P. The multifaceted roles of antibiotics and antibiotic resistance in nature. Front Microbiol 2013; 4: 47

[61] Buijs Y, Zhang SD, Jørgensen KM, Isbrandt T, Larsen TO, Gram L. Enhancement of antibiotic production by co-cultivation of two antibiotic producing marine Vibrionaceae strains. FEMS Microbiol Ecol 2021; 97: fiab041

[62] Ward JM, Hodgson JE. The biosynthetic genes for clavulanic acid and cephamycin production occur as a 'super-cluster'in three Streptomyces. FEMS Microbiol Lett 1993; 110: 239–242

[63] Koumoutsi A, Chen XH, Henne A, Liesegang H, Hitzeroth G, Franke P, Vater J, Borriss R. Structural and functional characterization of gene clusters directing nonribosomal synthesis of bioactive cyclic lipopeptides in Bacillus amyloliquefaciens strain FZB42. J Bacteriol Res 2004; 186: 1084–1096

[64] Mast Y, Weber T, Gölz M, Ort-Winklbauer R, Gondran A, Wohlleben W, Schinko E. Characterization of the 'pristinamycin supercluster' of Streptomyces pristinaespiralis. Microb Biotechnol 2011; 4: 192–206

[65] Arp J, Götze S, Mukherji R, Mattern DJ, García-Altares M, Klapper M, Brock DA, Brakhage AA, Strassmann JE, Queller DC. Synergistic activity of cosecreted natural products from amoebae-associated bacteria. PNAS 2018; 115: 3758–3763

[66] Mrak P, Krastel P, Lukančič PP, Tao J, Pistorius D, Moore CM. Discovery of the actinoplanic acid pathway in Streptomyces rapamycinicus reveals a genetically conserved synergism with rapamycin. J Biol Chem 2018; 293: 19982–19995

[67] Martinet L, Naômé A, Baiwir D, De Pauw E, Mazzucchelli G, Rigali S. On the risks of phylogeny-based strain prioritization for drug discovery: Streptomyces lunaelactis as a case study. Biomolecules 2020; 10: 1027

[68] Del Carratore F, Zych K, Cummings M, Takano E, Medema MH, Breitling R. Computational identification of co-evolving multi-gene modules in microbial biosynthetic gene clusters. Commun Biol 2019; 2: 83

[69] Li Y, Steenwyk JL, Chang Y, Wang Y, James TY, Stajich JE, Spatafora JW, Groenewald M, Dunn CW, Hittinger CT. A genome-scale phylogeny of the kingdom Fungi. Curr Biol 2021; 31: 1653–1665

[70] Hawksworth DL, Lücking R. Fungal diversity revisited: 2.2 to 3.8 million species. Microbiol Spectr 2017; 5: FUNK-0052-2016. doi:10.1128/microbiolspec.FUNK-0052-2016

[71] Hibbett D, Abarenkov K, Kõljalg U, Öpik M, Chai B, Cole J, Wang Q, Crous P, Robert V, Helgason T. Sequence-based classification and identification of Fungi. Mycologia 2016; 108: 1049–1068

[72] Frąc M, Hannula SE, Bełka M, Jędryczka M. Fungal biodiversity and their role in soil health. Front Microbiol 2018; 9: 707

[73] Lücking R, Aime MC, Robbertse B, Miller AN, Ariyawansa HA, Aoki T, Cardinali G, Crous PW, Druzhinina IS, Geiser DM. Unambiguous identification of fungi: Where do we stand and how accurate and precise is fungal DNA barcoding? IMA Fungus 2020; 11: 14

[74] Chassagne F, Cabanac G, Hubert G, David B, Marti G. The landscape of natural product diversity and their pharmacological relevance from a focus on the Dictionary of Natural Products®. Phytochem Rev 2019; 18: 601–622

[75] Lind AL, Wisecaver JH, Lameiras C, Wiemann P, Palmer JM, Keller NP, Rodrigues F, Goldman GH, Rokas A. Drivers of genetic diversity in secondary metabolic gene clusters within a fungal species. PLoS Biol 2017; 15: e2003583

[76] Pfannenstiel BT, Keller NP. On top of biosynthetic gene clusters: How epigenetic machinery influences secondary metabolism in fungi. Biotechnol Adv 2019; 37: 107345

[77] Quinn R. Rethinking antibiotic research and development: World War II and the penicillin collaborative. Am J Public Health 2013; 103: 426–434

[78] Bo G. Giuseppe Brotzu and the discovery of cephalosporins. Clin Microbiol Infect 2000; 6: 6–8

[79] Endo A. A historical perspective on the discovery of statins. Proc Jpn Acad Ser B Phys Biol Sci 2010; 86: 484–493

[80] Chun J, Brinkmann V. A mechanistically novel, first oral therapy for multiple sclerosis: The development of fingolimod (FTY720, Gilenya). Discov Med 2011; 12: 213–228

[81] Balkovec JM, Hughes DL, Masurekar PS, Sable CA, Schwartz RE, Singh SB. Discovery and development of first in class antifungal caspofungin (CANCIDAS®)-A case study. Nat Prod Rep 2014; 31: 15–34

[82] Tran PN, Yen MR, Chiang CY, Lin HC, Chen PY. Detecting and prioritizing biosynthetic gene clusters for bioactive compounds in bacteria and fungi. Appl Microbiol Biotechnol 2019; 103: 3277–3287

[83] Hutchings MI, Truman AW, Wilkinson B. Antibiotics: Past, present and future. Curr Opin Microbiol 2019; 51: 72–80

[84] Noble RL. The discovery of the vinca alkaloids-chemotherapeutic agents against cancer. Biochem Cell Biol 1990; 68: 1344–1351

[85] Oberlies NH, Kroll DJ. Camptothecin and taxol: Historic achievements in natural products research. J Nat Prod 2004; 67: 129–135

[86] Hubert J, Nuzillard JM, Renault JH. Dereplication strategies in natural product research: How many tools and methodologies behind the same concept? Phytochem Rev 2017; 16: 55–95

[87] El-Elimat T, Figueroa M, Ehrmann BM, Cech NB, Pearce CJ, Oberlies NH. High-resolution MS, MS/MS, and UV database of fungal secondary metabolites as a dereplication protocol for bioactive natural products. J Nat Prod 2013; 76: 1709–1716

[88] Halabalaki M, Vougogiannopoulou K, Mikros E, Skaltsounis AL. Recent advances and new strategies in the NMR-based identification of natural products. Curr Opin Biotechnol 2014; 25: 1–7

[89] Egan JM, van Santen JA, Liu DY, Linington RG. Development of an NMR-based platform for the direct structural annotation of complex natural products mixtures. J Nat Prod 2021; 84: 1044–1055

[90] van der Lee TA, Medema MH. Computational strategies for genome-based natural product discovery and engineering in fungi. Fungal Genet Biol 2016; 89: 29–36

[91] Sarma V, Hyde KD. A review on frequently occurring fungi in mangroves. Fungal Divers 2001; 8: 1–34

[92] Rokas A, Mead ME, Steenwyk JL, Raja HA, Oberlies NH. Biosynthetic gene clusters and the evolution of fungal chemodiversity. Nat Prod Rep 2020; 37: 868–878

[93] Keller NP. Fungal secondary metabolism: Regulation, function and drug discovery. Nat Rev Microbiol 2019; 17: 167–180

[94] Osbourn A. Secondary metabolic gene clusters: Evolutionary toolkits for chemical innovation. Trends Genet 2010; 26: 449–457

[95] Kautsar SA, Blin K, Shaw S, Navarro-Muñoz JC, Terlouw BR, van der Hooft JJ, Van Santen JA, Tracanna V, Suarez Duran HG, Pascal Andreu V. MIBiG 2.0: A repository for biosynthetic gene clusters of known function. Nucleic Acids Res 2020; 48: D454–D458

[96] Robey MT, Caesar LK, Drott MT, Keller NP, Kelleher NL. An interpreted atlas of biosynthetic gene clusters from 1000 fungal genomes. PNAS 2021; 118: e2020230118

[97] Godtfredsen W, Roholt K, Tybring L. Fucidin: A new orally active antibiotic. Lancet 1962; 279: 928–931

[98] Fernandes P. Fusidic acid: A bacterial elongation factor inhibitor for the oral treatment of acute and chronic staphylococcal infections. Cold Spring Harb Perspect Med 2016; 6: a025437

[99] Novak R, Shlaes DM. The pleuromutilin antibiotics: A new class for human use. Curr Opin Investig Drugs 2010; 11: 182–191

[100] Kavanagh F, Hervey A, Robbins WJ. Antibiotic substances from basidiomycetes: VIII. *Pleurotus multilus* (Fr.) Sacc. and *Pleurotus passeckerianus* Pilat. Proc Natl Acad Sci U S A 1951; 37: 570–574

[101] Yang LP, Keam SJ. Spotlight on retapamulin in impetigo and other uncomplicated superficial skin infections. Am J Clin Dermatol 2008; 9: 411–413

[102] Yan K, Madden L, Choudhry AE, Voigt CS, Copeland RA, Gontarek RR. Biochemical characterization of the interactions of the novel pleuromutilin derivative retapamulin with bacterial ribosomes. Antimicrob Agents Chemother 2006; 50: 3875–3881

[103] Hesterkamp T. Antibiotics clinical development and pipeline. Curr Top Microbiol Immunol 2016; 398: 447–474

[104] Morishita Y, Zhang H, Taniguchi T, Mori K, Asai T. The discovery of fungal polyene macrolides via a postgenomic approach reveals a poly-ketide macrocyclization by *trans*-acting thioesterase in fungi. Org Lett 2019; 21: 4788–4792

[105] Morishita Y, Aoki Y, Ito M, Hagiwara D, Torimaru K, Morita D, Kuroda T, Fukano H, Hoshino Y, Suzuki M, Taniguchi T, Mori K, Asai T. Genome mining-based discovery of fungal macrolides modified by glycosyl-phosphatidylinositol (GPI)-ethanolamine phosphate transferase homologues. Org Lett 2020; 22: 5876–5879

[106] Parkinson EI, Tryon JH, Goering AW, Ju KS, McClure RA, Kemball JD, Zhukovsky S, Labeda DP, Thomson RJ, Kelleher NL, Metcalf WW. Discovery of the tyrobetaine natural products and their biosynthetic gene cluster via metabologenomics. ACS Chem Biol 2018; 13: 1029–1037

[107] Grigoriev IV, Cullen D, Goodwin SB, Hibbett D, Jeffries TW, Kubicek CP, Kuske C, Magnuson JK, Martin F, Spatafora JW, Tsang A, Baker SE. Fueling the future with fungal genomics. Mycology 2011; 2: 192–209

[108] Martínez-Cárdenas A, Cruz-Zamora Y, Fajardo-Hernández CA, Villanueva-Silva R, Cruz-García F, Raja HA, Figueroa M. Genome mining and molecular networking-based metabolomics of the marine facultative *Aspergillus* sp. MEXU 27854. Molecules 2021; 26: 5362

[109] Clevenger KD, Bok JW, Ye R, Miley GP, Verdan MH, Velk T, Chen C, Yang K, Robey MT, Gao P, Lamprecht M, Thomas PM, Islam MN, Palmer JM, Wu CC, Keller NP, Kelleher NL. A scalable platform to identify fungal secondary metabolites and their gene clusters. Nat Chem Biol 2017; 13: 895–901

[110] Sukmarini L. Recent advances in discovery of lead structures from microbial natural products: Genomics- and metabolomics-guided acceleration. Molecules 2021; 26: 2542

[111] McClure RA, Goering AW, Ju KS, Baccile JA, Schroeder FC, Metcalf WW, Thomson RJ, Kelleher NL. Elucidating the rimosamide-detoxin natural product families and their biosynthesis using metabolite/gene cluster correlations. ACS Chem Biol 2016; 11: 3452–3460

[112] Goering AW, McClure RA, Doroghazi JR, Albright JC, Haverland NA, Zhang Y, Ju KS, Thomson RJ, Metcalf WW, Kelleher NL. Metabologenomics: Correlation of microbial gene clusters with metabolites drives discovery of a nonribosomal peptide with an unusual amino acid monomer. ACS Cent Sci 2016; 2: 99–108

[113] Montaser R, Kelleher NL. Discovery of the biosynthetic machinery for stravidins, biotin antimetabolites. ACS Chem Biol 2020; 15: 1134–1140

[114] Zhang WY, Zhong Y, Yu Y, Shi DF, Huang HY, Tang XL, Wang YH, Chen GD, Zhang HP, Liu CL, Hu D, Gao H, Yao XS. 4-Hydroxy pyridones from heterologous expression and cultivation of the native host. J Nat Prod 2020; 83: 3338–3346

[115] Bok JW, Ye R, Clevenger KD, Mead D, Wagner M, Krerowicz A, Albright JC, Goering AW, Thomas PM, Kelleher NL, Keller NP, Wu CC. Fungal artificial chromosomes for mining of the fungal secondary metabolome. BMC Genom 2015; 16: 343

[116] Robey MT, Ye R, Bok JW, Clevenger KD, Islam MN, Chen C, Gupta R, Swyers M, Wu E, Gao P, Thomas PM, Wu CC, Keller NP, Kelleher NL. Identification of the first diketomorpholine biosynthetic pathway using FAC-MS technology. ACS Chem Biol 2018; 13: 1142–1147

[117] Balunas MJ, Kinghorn AD. Drug discovery from medicinal plants. Life Sci 2005; 78: 431–441

[118] Kinghorn AD. Pharmacognosy in the 21st century. J Pharm Pharmacol 2001; 53: 135–148

[119] Hardy K, Buckley S, Collins MJ, Estalrrich A, Brothwell D, Copeland L, García-Tabernero A, García-Vargas S, De La Rasilla M, Lalueza-Fox C. Neanderthal medics? Evidence for food, cooking, and medicinal plants entrapped in dental calculus. Naturwissenschaften 2012; 99: 617–626

[120] Serturner F. Ueber das Morphium, eine neue salifahige Grundlage, und die Mekonsaure als Hauptbestandtheile des Opiums. Ann Phys 1817; 55: 56–89

[121] Bucar F, Wube A, Schmid M. Natural product isolation-how to get from biological material to pure compounds. Nat Prod Rep 2013; 30: 525–545

[122] Tsugawa H, Rai A, Saito K, Nakabayashi R. Metabolomics and complementary techniques to investigate the plant phytochemical cosmos. Nat Prod Rep 2021; 38: 1729–1759

[123] Olivon F, Allard PM, Koval A, Righi D, Genta-Jouve G, Neyts J, Apel C, Pannecouque C, Nothias LF, Cachet X. Bioactive natural products prioritization using massive multi-informational molecular networks. ACS Chem Biol 2017; 12: 2644–2651

[124] Wu F. Mining plant genomes for rapid discovery of nature products. Mol Plant 2019; 12: 13

[125] Nützmann HW, Osbourn A. Gene clustering in plant specialized metabolism. Curr Opin Biotechnol 2014; 26: 91–99

[126] Medema MH, Osbourn A. Computational genomic identification and functional reconstitution of plant natural product biosynthetic pathways. Nat Prod Rep 2016; 33: 951–962

[127] Nützmann HW, Huang A, Osbourn A. Plant metabolic clusters–from genetics to genomics. New Phytol 2016; 211: 771–789

[128] Frey M, Chomet P, Glawischnig E, Stettner C, Grün S, Winklmair A, Eisenreich W, Bacher A, Meeley RB, Briggs SP. Analysis of a chemical plant defense mechanism in grasses. Science 1997; 277: 696–699

[129] Qi X, Bakht S, Leggett M, Maxwell C, Melton R, Osbourn A. A gene cluster for secondary metabolism in oat: Implications for the evolution of metabolic diversity in plants. PNAS 2004; 101: 8233–8238

[130] Kautsar SA, Suarez Duran HG, Blin K, Osbourn A, Medema MH. PlantiSMASH: Automated identification, annotation and expression analysis of plant biosynthetic gene clusters. Nucleic Acids Res 2017; 45: W55–W63

[131] Polturak G, Osbourn A. The emerging role of biosynthetic gene clusters in plant defense and plant interactions. PLoS Pathog 2021; 17: e1009698