Biotechnology Advances 66 (2023) 108176

Contents lists available at ScienceDirect

Biotechnology Advances

journal homepage: www.elsevier.com/locate/biotechadv

Research review paper

Accessing hidden microbial biosynthetic potential from underexplored sources for novel drug discovery

Lei Li *

State Key Laboratory of Microbial Metabolism and School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, 800 Dongchuan Rd., Shanghai 200240, China

ARTICLE INFO

Keywords:
Natural products
Underexplored sources
Biosynthetic potential
Synthetic biology
Artificial intelligence
Antimicrobial peptides

ABSTRACT

Microbial natural products and their structural analogues have widely used as pharmaceutical agents, especially for infectious diseases and cancer. Despite this success, new structural classes with innovative chemistry and modes of action are urgently needed to be developed to combat the growing antimicrobial resistance and other public health problems. The advances in next-generation sequencing technologies and powerful computational tools open up new opportunities to explore microbial biosynthetic potential from underexplored sources, with millions of secondary metabolites awaiting discovery. The review highlights challenges associated with discovery of new chemical entities, rich reservoirs provided by untapped taxa, ecological niches or host microbiomes, emerging synthetic biotechnologies to unearth the hidden microbial biosynthetic potential for novel drug discovery at scale and speed.

1. History and challenges of microbial natural products for drug development

Microbial natural products (NPs) with intrinsic chemical complexity and intriguing biological profiles, play eminent roles in medicine, agriculture and nutrition (Carroll et al., 2022; Medema et al., 2021; Newman and Cragg, 2020). Historically, microbial NPs have been widely used as antibiotics, anti-tumor compounds and other pharmaceutical agents (Hug et al., 2020; Keller, 2019). In the last five years, a series of NPs or NP-inspired compounds with novel structural scaffolds or modes-of-action (MOA) have been approved for clinical use, crop protection or animal health, such as pleuromutilin-derived lefamulin (antibiotic) and UK-2A-derived fenpicoxamid (fungicide). Undoubtedly, microbial NPs will continue to play important roles in modern medicine, serving as the main source of and inspiration for small molecule drugs to overcome global antimicrobial resistance and other public health problems (Atanasov et al., 2021). Compared with synthetic organic molecules, genetically encoded small molecules offer unique features, including typically higher molecular mass and greater molecular rigidity, and cover a wider area of chemical space for novel drug development (Atanasov et al., 2021). With rapid advances in low-cost and high-throughput DNA sequencing technologies, large-scale microbial (meta) genomic sequencing reveals highly diverse sets of biosynthetic gene clusters (BGCs) in bacteria or fungi, in which characterized metabolites represent just the tip of the iceberg, with millions of novel NPs awaiting discovery (Hemmerling and Piel, 2022; Kalkreuter et al., 2020; Scherlach and Hertweck, 2021).

On the contrast, we are experiencing a substantially increased antimicrobial resistance (AMR) crisis and other public health problems. NPs with innovative chemical scaffolds and MOA are urgently needed to tackle the AMR or other disease-posed problems (Lewis, 2021; Miethke et al., 2021). However, discovery of new chemical entities is uniquely challenging mainly due to high rediscovery rates. For example, a systematic search of silent BGCs revealed that *Streptomyces*, the main NP sources, have been overmined and have low chance to produce new drug leads against Gram-negative pathogens (Lewis, 2021). Except for *Streptomyces*, a number of bacterial or fungal genera, including *Bacillus*, *Pseudomonas*, *Aspergillus*, *Fusarium* and *Penicillium*, have been extensively studied for novel microbial NP discovery over the last few decades (Gavriilidou et al., 2022; Robey et al., 2021). In the recent study, less popular microbial taxa, including Myxococcaceae, Nostocaceae, Pleurocapsa, Weeksellaceae, Basidiomycota and Mucoromycota, hold promise as underexploited sources of novel NPs with applications in the pharmaceutical and agriscience industries (Gavriilidou et al., 2022; Robey et al., 2021). The insight will facilitate BGC prioritization for the discovery of new chemical entities. In this review, the hidden microbial

* Corresponding author at: Laboratory of Synthetic Microbial Drugs, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, China.
E-mail address: lei.li@sjtu.edu.cn.

https://doi.org/10.1016/j.biotechadv.2023.108176
Received 14 November 2022; Received in revised form 15 May 2023; Accepted 16 May 2023
Available online 19 May 2023

biosynthetic potential from underexplored sources is highlighted and the newest developments on emerging biotechnologies to mine and unearth the biosynthetic "dark matter" are discussed.

2. Underexplored sources for the discovery of novel microbial drugs

Underexplored or less-well-studied sources, including untapped bacterial or fungal taxa, microorganisms from global ecological niches as well as complex microbial communities, have been found to encode highly diverse and chemically distinct NPs, which provide rich and unique reservoirs for novel microbial drug discovery (Hemmerling and Piel, 2022; Scherlach and Hertweck, 2021).

2.1. Untapped bacterial or fungal taxa

By accessing ~170,000 bacterial genomes, a compendium of NP biosynthetic diversity (numbers and types of BGCs and GCFs) in genome-sequenced bacteria was systematically evaluated (Fig. 1a) (Gavriilidou et al., 2022). The study clearly indicated that only ~3% NPs have been characterized and several less-well-studied taxa show high biosynthetic diversity, including Myxococcaceae, Nostocaceae, Pleurocapsa and Weeksellaceae (Gavriilidou et al., 2022). These findings will greatly help overcome one of the main bottlenecks of NP discovery: BGC prioritization for research. Furthermore, genomic and cheminformatics analyses of 1037 fungi revealed that bacterial and fungal sources represent significantly distinct chemical space and biosynthetic logic (Robey et al., 2021). The interpreted atlas of fungal BGCs also revealed that a large variety of chemical entities in a majority of underexplored fungal taxa awaits discovery (Robey et al., 2021).

2.2. Microorganisms from global ecological niches

With advances in sequencing technologies and computational tools, reconstitution of metagenome-assembled genomes (MAGs) has enabled insights into the biosynthetic potential of global ecological niches, including ocean and glacier (Fig. 1a). In a joint study, Nayfach et al., accessed a genomic catalog of Earth's microbiomes, including >10,000 metagenomes from diverse habitats covering all of Earth's continents and oceans (Nayfach et al., 2021). The authors identified 87,187 putatively novel BGCs and found that the phylum Acidobacteria has an underestimated biosynthetic potential. Moreover, Paoli et al., investigated the biosynthetic diversity of microorganisms in the global ocean (>1000 seawaters samples) and unearthed ~40,000 putatively new BGCs encoded by ~10,000 microbial genomes (Paoli et al., 2022). As a representative case, turbinmicin was identified from marine animal microbiomes, which showed potent activity against urgent-threat multidrug-resistant (MDR) fungal pathogens in mice by a fungal-

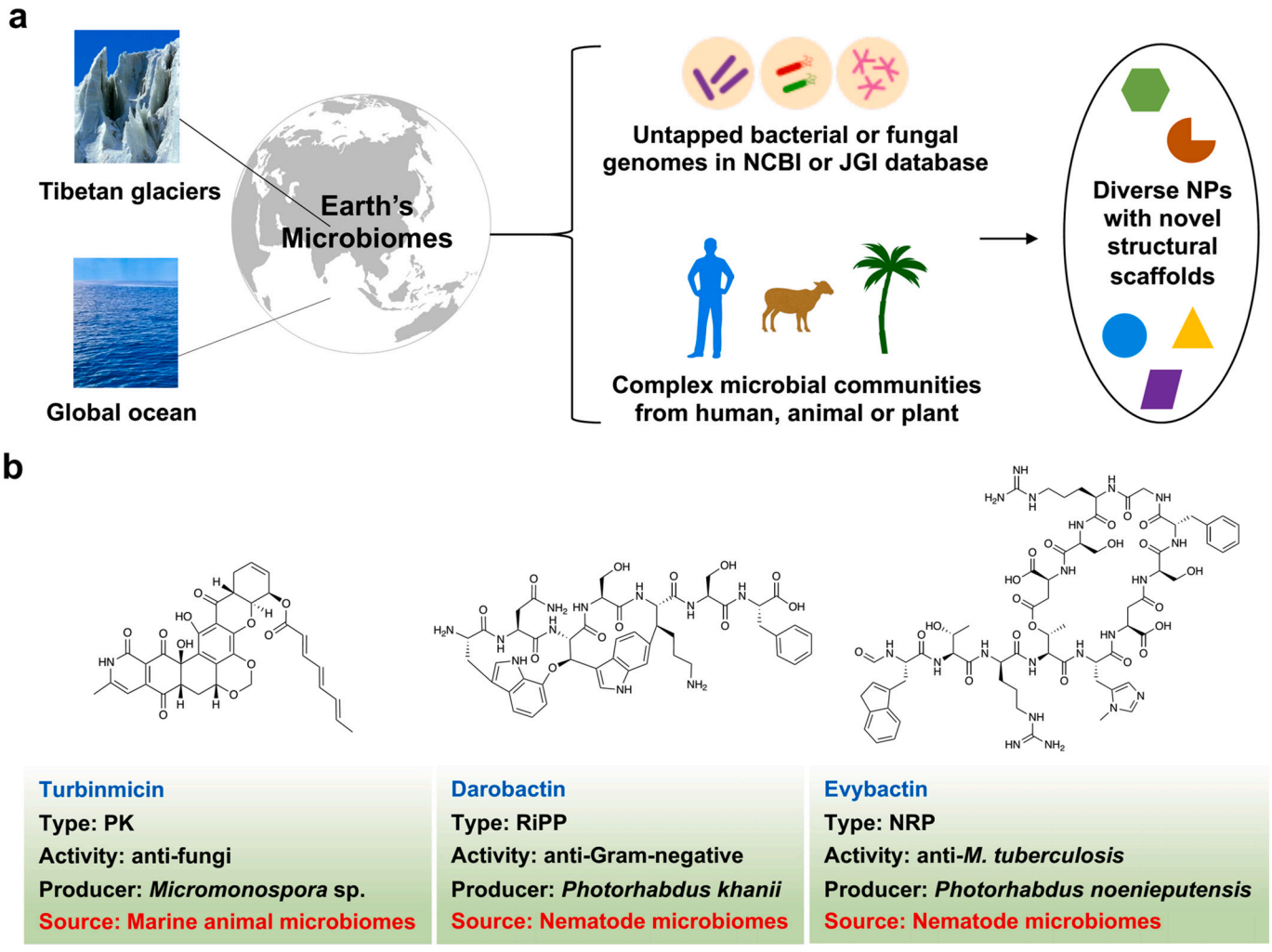

Fig. 1. Underexplored Sources for Microbial Drug Discovery. (a) Unearthing hidden biosynthetic potential from the Earth's microbiomes. Underexplored sources, including untapped microbial taxa, global ecological niches and human, animal or plant microbiomes, open up new opportunities for the discovery of new chemical entities for novel drug discovery. (b) Three structurally diverse NPs with distinct antimicrobial activities from animal microbiomes. Abbreviations: PK, polyketide; RiPP, ribosomally synthesized and post-translationally modified peptide; NRP, non-ribosomal peptide.

specific MOA (Fig. 1b) (Zhang et al., 2022). In another study, Tibetan glaciers, including cryoconite, ice and snow habitats, were found to harbor a diverse range of novel microbial species and provide a unique source to explore BGC novelty and evolution (Liu et al., 2022a, 2022b). Interestingly, although soil ecosystems have been overmined for drug discovery, a number of abundant and phylogenetically bacterial phylums, including Acidobacteria, Gemmatimonadetes, Rokubacteria and Verrucomicobia, has been underestimated, which has a large potential to produce highly diverse chemical scaffolds that warrant further investigation (Crits-Christoph et al., 2018). The studies indicated that metagenomic approaches can overcome the limitation of cultivation techniques and open up opportunities to assess bioactive NPs encoded by environmental DNA.

2.3. Complex microbial communities

Small molecules from soil, ocean or glacier microorganisms can not evolve to function in hosts, such as animals and plants. The symbionts of animal or plant microbiomes have potential to produce new molecules that will function systemically in the inhabited hosts. In fact, complex microbial communities, such as human, animal or plant microbiomes, have been found to encode rich BGCs for new chemistry (Fig. 1a) (Almeida et al., 2021; Bickhart et al., 2022; Scott and Piel, 2019). For example, Bickhart et al., accessed the sleep fecal metagenomes and identified 1400 complete BGCs, most of which are novel (Bickhart et al., 2022). As representative cases, two chemically and mechanistically diverse anti-microbial peptides, including darobactin and evybactin, were isolated from Nematode microbiomes, which exhibited selective activities against Gram-negative bacteria and *Mycobacterium tuberculosis*, respectively (Fig. 1b) (Imai et al., 2019; Imai et al., 2022). As a promising drug lead, darobactin inhibits the outer membrane insertase BamA by mimicking a β-strand (Kaur et al., 2021). The cyclic depsipeptide evybactin targets DNA gyrase when transferred into *M. tuberculosis* by a membrane transporter, BacA, and does not harm human gut microbiomes because gut commensal bacteria lack a BacA-type protein (Imai et al., 2022).

3. Emerging technologies to unearth the hidden microbial biosynthetic potential from underexplored sources

In the last ten years, a wide variety of emerging technologies have been developed for BGC prediction, prioritization and engineering, which significantly facilitate the discovery of bioactive NPs with novel chemical entities. As sequencing data accumulate, BGC prioritization will be critical to discover new class of NPs or diversity among a family of known NPs. The simplest strategy for BGC prioritization is to access NPs' structures from DNA sequences but only a few classes of NPs, such as non-ribosomal peptides (NRPs) and ribosomally synthesized and post-translationally modified peptides (RiPPs), can be directly predicted. Interested readers are referred to some excellent reviews or articles on BGC prioritization (Bauman et al., 2021; Culp et al., 2022; Panter et al., 2021; Yan et al., 2018). Here, I focus on the discussion of recent technological developments in the last ten years that are enabling discovery of novel microbial NPs from underexplored sources, and highlight selected applications (Tables 1 and 2).

3.1. Computational tools and databases for NP discovery

To facilitate the discovery of microbial NPs, different computational tools and databases have been summaried in recent review papers (Chevrette et al., 2021; Hemmerling and Piel, 2022). Here a number of widely used tools and databases in the field of microbial NPs have been

Table 1
Computational tools and databases for microbial NP discovery.

Function	Name	Link	Description	Reference
BGC prediction	antiSMASH 6.0	https://antismash.secondarymetabolites.org/	A webserver for detecting and characterizing of microbial BGCs	Blin et al., 2021
	PRISM 3	http://magarveylab.ca/prism/	A computational tool for predicting bacterial NPs' structures	Skinnider et al., 2017
	gutSMASH	https://gutsmash.bioinformatics.nl/	A webserver for identifying primary pathways in gut microbiota	Pascal Andreu et al., 2021
	ARTS 2.0	http://arts.ziemertlab.com/	An antibiotic resistant target seeker for targeted genome mining	Mungan et al., 2020
	CONKAT-Seq	NA	A workflow for co-occurrence network analysis of targeted BGCs	Libis et al., 2019
BGC novelty	BiG-SCAPE/CORASON	https://bigscape-corason.secondarymetabolites.org/	A tool for analyzing sequence similarity of BGCs and their families	Navarro-Muñoz et al., 2020
	BiG-SLiCE	https://github.com/medema-group/bigslice/	A tool for visualizing the diversity of 1.2 million BGCs	Kautsar et al., 2021a, 2021b
BGC database	MIBIG 2.0	https://mibig.secondarymetabolites.org/	A repository for known microbial NPs' biosynthetic pathways	Kautsar et al., 2020
	antiSMASH Database v3	https://antismash-db.secondarymetabolites.org/	A database for predicted BGCs from high-quality genomes	Blin et al., 2021
	IMG-ABC v.5.0	https://img.jgi.doe.gov/abc-public/	An integrated microbial genomes atlas of predicted BGCs	Palaniappan et al., 2020
	Big-FAM	https://bigfam.bioinformatics.nl/	A database for predicted microbial BGC families	Kautsar et al., 2021a, 2021b
	ARTS-DB	https://arts-db.ziemertlab.com/	A database for pre-computed antibiotic resistant targets	Mungan et al., 2022
	ClusterMine360	http://www.clustermine360.ca/	A database of microbial PKS/NRPS biosynthesis	Conway and Boddy, 2013
	MycoCosm	http://jgi.doe.gov/fungi	An interactive web-based portal for fungal genomics resources	Grigoriev et al., 2014
NP database	Natural Products Atlas 2.0	https://www.npatlas.org/	A database for microbially-derived known natural products	van Santen et al., 2022
	SuperNatural 3.0	http://bioinf-applied.charite.de/supernatural_3	A database of natural compounds and their derivatives	Gallo et al., 2023
	NPASS	http://bidd2.nus.edu.sg/NPASS/	A database of natural product activity and species source	Zeng et al., 2018
	StreptomeDB	http://132.230.56.4/streptomedb2/	A resource of natural products produced by streptomycetes	Klementz et al., 2016
	GNPS	http://gnps.ucsd.edu/	A resourece of tandem mass (MS/MS) spectrometry data	Wang et al., 2016

Table 2
Enabling technologies to discover novel microbial NPs from underexplored sources. AMP: antimicrobial peptide. NRP: non-ribosomal peptide.

Strategy	Method	Method description	NP type	Reference
Activation of silent BGCs				
BGC activation in native producers	HiTES	High-throughput elicitor screening to activate cryptic BGCs in fungi	Any type	Lee and Seyedsayamdost, 2022
	HiTES-imaging MS	A genetic-free method by combining high-throughput chemical genetic screening with imaging MS in rare actinomycetes	Any type	Xu et al., 2019
Direct cloning of BGCs	CAT-FISHNG	An in vitro platform for CRISPR/Cas12a-mediated fast direct BGC cloning	Any type	Liang et al., 2022
	CAPTURE	Cas12a-assisted precise targeted BGC cloning using in vivo Cre-lox recombination	Any type	Enghiad et al., 2021
	NabLC	A natural competence based large DNA fragment cloning in the anaerobe *Streptococcus mutans*	Any type	Hao et al., 2019
Bottom-up BGC refactoring	HEx	A scalable approach to heterologous expression of fungal cryptic BGCs	Any type	Harvey et al., 2018
	auto-HTP	An automated and high-throughput biofoundry workflow for discovery of fungal terpenoids	Terpenoid	Yuan et al., 2022; Tao et al., 2022
	Cross-kingdom expression platform	Redesign BGCs into synthetic genetic elements that can be expressed in prokaryote and eukaryote hosts	Any type	Patel et al., 2022
AI-derived AMP discovery	Deep learning approach	A unified pipeline for candidate AMP identification from human microbiome	Peptide	Ma et al., 2022
	Classifier model design approach	In silico identification of AMPs using computational analysis from rumen microbiome	Peptide	Oyama et al., 2022
syn-BNP for antibiotic discovery	De novo syn-BNP	Bioinformatic prediction of random NRP BGCs followed by chemical synthesis	NRP	Chu et al., 2016, 2020
	Targeted syn-BNP	Bioinformatic prediction of a type of NRP BGCs-of-interest followed by chemical synthesis	NRP	Li et al., 2022; Wang et al., 2022a, 2022b

discussed about their advantages and limitations of usage. A variety of powerful computational tools, such as antiSMASH 7.0, gutSMASH and ARTS 2.0, have been developed for prediction of BGCs from gut microbiota or other microbial sources (Blin et al., 2023; Mungan et al., 2020; Pascal Andreu et al., 2021). Particularly, antiSMASH 7.0 is currently the most widely used computational platform to detect and characterize up to 81 BGC types in microbes. Meanwhile, the accuracy of chemical structure prediction in antiSMASH 7.0 has further been improved, especially for non-ribosomal peptides. However, there is still a big obstacle to optimize the structural prediction accuracy of other NP types, including polyketides (PKs) and terpenes. Meanwhile, a panel of unsupervised algorithms such as BiG-SCAPE/CORASON and more recently BiG-SLiCE, have been developed to analyze sequence similarity of BGCs and gene cluster families (Kautsar et al., 2021a, 2021b; Navarro-Muñoz et al., 2020). Nowadays, both algorithms have been widely used to identify novel features within and between taxonomic lineages by clustering BGCs into gene cluster families (GCFs). Noteworthy, high-quality genomic and chemical databases are still required for these unsupervised platforms. Furthermore, a bioinformatic workflow termed CONKAT-seq was developed for co-occurrence network analysis of targeted DNA sequences, which allows scalable discovery of novel natural compounds from soil microbiomes (Libis et al., 2019). The CONKAT-seq approach can be used to rapidly detect and characterize novel NRP or PK gene clusters, but requires to be further optimized for other BGC types in the future.

Besides these computational toos discussed above, repositories for BGCs encoded by known NPs (e.g., MIBiG 2.0) or unknown chemical entities (e.g., antiSMASH database v3, IMG-ABC v.5.0, BiG-FAM and ARTS-DB), have been established to strengthen in silico analysis of novel BGCs (Blin et al., 2021; Kautsar et al., 2020; Kautsar et al., 2021a, 2021b; Mungan et al., 2022; Palaniappan et al., 2020;). For example, the BiG-FAM database groups homologous BGCs into Gene Cluster Families (GCFs) and includes 29,955 GCFs covering the global diversity of 1,225,071 BGCs, which provides new insights into rapidly accessing the novelty of BGCs by mapping their architectural and taxonomic diversity (Kautsar et al., 2021a, 2021b). ARTS-DB, a database for antibiotic resistance targets from >70,000 genomes and metagenomes, opens up new possibilities for rapid target-directed genome mining, including BGC proximity and duplication of housekeeping genes (Mungan et al., 2022). In addition, the Natural Products Atlas 2.0, a chemical database of known microbial NP structures, has been released, which contains the structural and content changes of 32,552 compounds as well as taxonomic descriptions for all bacterial or fungal taxa (van Santen et al., 2022). Recently, a larger database of NPs and derivatives, SuperNatural 3.0, was released, which contains 449,058 compounds with their physicochemical and structural information (Gallo et al., 2023). These powerful computational tools and databases, which have been listed in Table 1, open up new opportunities to explore hidden microbial biosynthetic potential from underexplored sources.

3.2. Activating silent BGCs from undomesticated microorganisms

Generally, underexplored natural compound producers, such as most of Gram-negative proteobacteria, anaerobic bacteria and filamentous fungi, are genetically intractable. A large variety of novel enabling technologies is required to be developed to access their biosynthetic potential, including species-specific genome editing tools for in situ BGC activation and high-efficient BGC engineering approaches for heterologous expression.

3.2.1. BGC activation in native producers

In situ BGC activation mainly relies on high-efficient genome editing for promoter engineering. By combining environmental transformation sequencing and CRISPR-Cas transposase-mediated DNA insertion systems, Rubin et al., recently developed a panel of species- and site-specific genome editing tools in complex bacterial communities, such as soil and infant gut microbiota, which open up new possibilities to access novel BGCs in non-model bacteria (Rubin et al., 2022). A variety of CRISPR-Cas9-based genome editing and transcriptional regulation technologies have been also developed in diverse filamentous fungi, which facilitates the exploitation of new bioactive fungal metabolites (Jiang et al., 2021). Besides rational genetic manipulation of BGCs, several novel chemical genetics approaches such as high-throughput elicitor screening (HiTES)-imaging MS platform have been used to rapidly unearth biosynthetic potential in underexplored microorganisms in a genetics-free manner (Lee and Seyedsayamdost, 2022; Xu et al., 2019). For example, Lee et al., extended their previously developed HiTES platform from unconventional actinomycetes to rare filamentous fungi, and successfully uncovered 13 structurally diverse metabolites, including sclerocyclane with a novel scaffold (Lee and Seyedsayamdost,

2022). Considering that it is a direct route for NP discovery by in situ genetic engineering but >90% BGCs are silent or cryptic under standard laboratory conditions (Rutledge and Challis, 2015; Walsh and Fischbach, 2010), more general synthetic biology approaches are still needed, especially for genetically intractable strains.

3.2.2. *Heterologous expression of native or refactored BGCs*

In the last five years, a variety of versatile surrogate heterologous hosts has been developed for novel drug lead discovery from distinct bacterial or fungal taxa (Liu et al., 2022a, 2022b; Meng et al., 2022; Roux and Chooi, 2022). For example, genome-reduced *Schlegelella brevitalea* classis exhibited alleviated cell autolysis and was better than two commonly used Gram-negative chassis *Escherichia coli* and *Pseudomonas putida* for production of cryptic metabolites from proteobacteria (Liu et al., 2021). Meanwhile, a variety of Cas12a-assisted approaches has been developed for targeted and precise BGC cloning, including CATFISHING and CAPTURE (Enghiad et al., 2021; Liang et al., 2022). For example, Enghiad et al., used CAPTURE to clone 47 BGCs from 10 to 113 kb with ~100% efficiency, significantly facilitating large-scale discovery of novel chemical entities (Enghiad et al., 2021). Intriguingly, Hao et al., developed a natural competence based large DNA fragment cloning (NabLC) approach in the facultative anaerobe *Streptococcus mutans* UA159 and successfully identified an anti-infiltration compound from anaerobic bacteria in human oral microbiomes (Hao et al., 2019). Compared with in situ BGC activation, heterologous expression not only provides an alternative route to access nature's abundance of small molecules without the need of genetic manipulation of native hosts, but also shows unique advantages for NP discovery from obligate symbionts and environmental DNA.

Although direct cloning and heterologous expression of novel BGCs provides an important route to access microbial biosynthetic capacity from underexplored sources, a majority of BGCs are lowly or not expressed even in engineered heterologous hosts. With rapid advances in synthetic biology, BGC refactoring has become a general strategy to access silent pathways under standard laboratory conditions. In 2018, Harvey et al., developed a heterologous expression (HEx) discovery platform for the high-efficient expression of large-scale fungal BGCs in engineered *Saccharomyces cerevisiae* (Harvey et al., 2018). Using the integrated platform, 41 cryptic BGCs from diverse fungal species were heterologously expressed. In 2022, Yuan et al., developed an automated and high-throughput (auto-HTP) biofoundry workflow for efficient genome mining of bioactive fungal terpenoids (Fig. 2a) (Yuan et al., 2022). Using this workflow, >180 distinct terpenoids, including a type of novel non-squalene triterpenes, were identified by refactoring 38 BGCs into 208 engineered strains (Tao et al., 2022; Yuan et al., 2022). The approach can be easily extended to other NP types in diverse microbial classis and then accelerate the discovery of new pharmaceutical agents. Notably, Patel et al., recently developed a series of synthetic genetic elements for cross-kingdom expression of biosynthetic pathways in Gram-negative and -positive bacteria as well as eukaryotes, decoupling biosynthetic potential from host-range constraints to unlock cryptic BGCs. Via computer-aided design of targeted BGCs, they identified a new class of human microbiome-derived nucleotide metabolites (Fig. 2b) (Patel et al., 2022). As the cost of gene synthesis continues to decrease and DNA assembly capacity becomes more powerful, it is expected that scientists will shift toward the bottom-up BGC refactoring strategy to rapidly activate BGCs of interest.

3.3. *Artificial intelligence-derived AMP discovery from human or animal microbiomes*

According to prediction, drug-resistant microbial infection will kill >10 million people worldwide by 2050 and new classes of antimicrobial therapies are urgently needed (Antimicrobial Resistance Collaborators, 2022; Fisher et al., 2022). Antimicrobial peptides (AMPs), such as small proteins typically 8–50 amino acids in length, have low susceptibility to resistance development, which provide an important alternative to traditional antibiotics (Privalsky et al., 2021). With rapid advances in artificial intelligence (AI), including natural language processing and machine learning algorithms, an AI era that molecules can be recognized by computer instead of traditional trial-and-error strategy is coming for drug discovery (Saldívar-González et al., 2021). Recently, some groups are using machine learning to discover or design new antibiotics with predictive or generative models-based approaches. For example, a deep neural network was trained to predict structurally different antibiotics from known small molecules, and helicin was found to exhibit potent efficacy and low toxicity in mice (Stokes et al., 2020). Das et al., combined the deep generative classifier with molecular dynamics to design

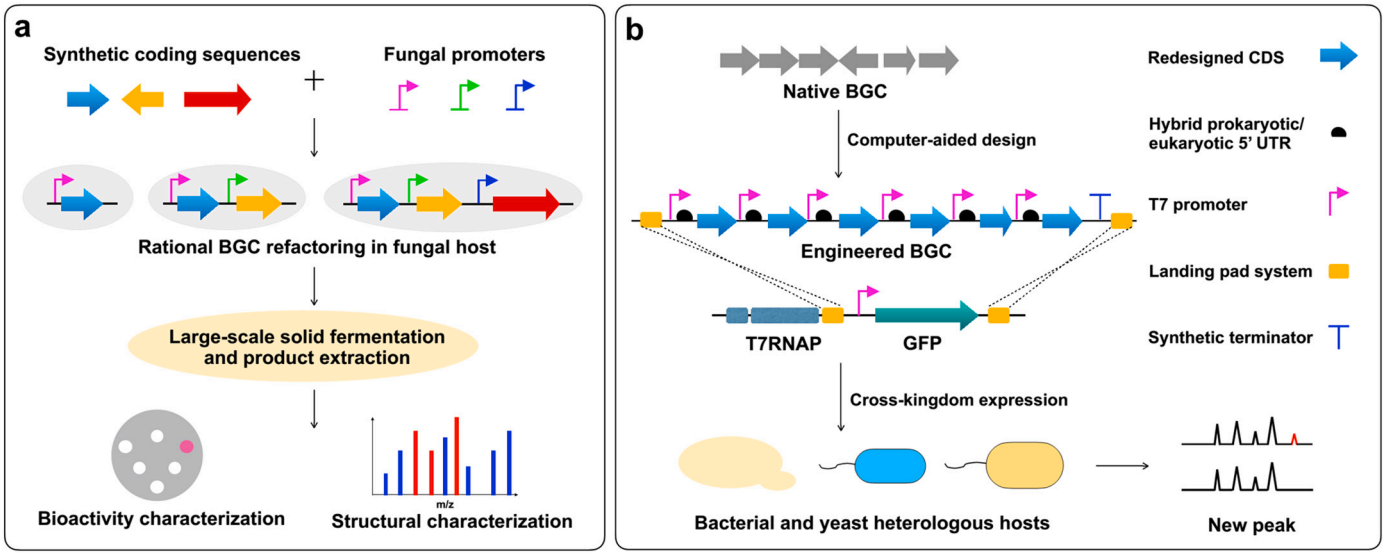

Fig. 2. De novo Assembly and Expression of BGCs. (a) An automated and high-throughput (auto-HTP) biofoundry workflow for discovery of fungal NPs. The workflow includes four modules: 1) bioinformatic prediction and automated assembly of fungal BGCs; 2) rational BGC refactoring in fungal chassis; 3) high-throughput fermentation and product extraction; 4) bioactivity and structural characterization. (b) Cross-kingdom expression of synthetic genetic elements for NP discovery from human microbiomes. These include three modules in the platform: 1) computer-aided design to generate synthetic genetic elements (SGE); 2) assembly and mobilization of SGE to heterologous hosts; 3) cross-kingdom expression of engineered biosynthetic pathways.

and synthesize 20 candidate AMPs within 48 days (Das et al., 2021). Notably, Torres et al., used the similar algorithm to identify 2603 encrypted peptide antibiotics from the human proteome (Torres et al., 2022). These studies open up new routes for the discovery of candidate antibiotics from host or environmental microbiomes (Fig. 3).

In 2022, based on natural language processing and deep learning, Ma et al., developed a clever AI approach to access complex human metagenomic information (Fig. 3) (Ma et al., 2022). They first combined different deep-learning-based natural language processing models to construct an optimized pipeline for AMP discovery. Then, 4409 qualified bacterial genomes from human microbiomes were screened using the advanced AI pipeline and >20 million predicted AMPs were found from non-redundant sORF sequences. By cross-checking with the available metaproteomic data, 2349 candidate AMPs were eventually identified. Combined with relative abundance, gene expression and association with selected bacterial taxa, 241 peptides were finally selected for chemical synthesis. A total 216 peptides were successfully synthesized and 181 of these showed antimicrobial activity with a high hit rate. More importantly, they found that most of these synthesized AMPs show low identities of <40% to previously conventional peptides, which provide novel chemical entities for drug discovery. Among all 181 AMPs, 11 were the most potent effective against drug-resistant pathogens and three showed high anti-infective efficacy and low toxicity in a mouse lung infection model (Fig. 3) (Ma et al., 2022).

Similarly, Linda et al., used the classifier model design to discover two AMPs from a rumen microbiome metagenomic dataset, which were also active against methicillin-resistant *Staphylococcus aureus* (MRSA) in a *Galleria mellonella* infection model (Oyama et al., 2022). Collectively, the AI driven chemical synthesis of linear peptides or small proteins presents a powerful methodology for rapidly accessing novel antibiotics from complex host or environmental microbiomes and has a promising potential to replenish our depleted arsenal. One possible limitation of AI-driven AMP discovery is their propensity toward biases. More robust training sets and experimentally validated models will further enhance success rate and decrease sequence biases. Additionally, AI-derived discovery pipeline requires to be further optimized and extended to access other structural classes of microbial NPs.

3.4. *De novo and targeted syn-BNP for antibiotic discovery*

Nonribosomal peptides (NRPs) show high chemical and mechanistic diversity, which have been widely used as lead compounds for drug development, such as daptomycin and colistin (Liu et al., 2019). Thanks to the deep understanding of biosynthesis for known NRPs, chemical structures of a large variety of NRPs can be predicted with high confidence (Süssmuth and Mainz, 2017). This provides an alternative route to access BGC-encoded products via bioinformatics prediction followed by chemical synthesis, that is a synthetic-Bioinformatics Natural Product (syn-BNP) approach. Here, I discuss some representative examples that demonstrate the utility of this new methodology with an emphasis on the targeted syn-BNP approach to unlock biosynthetic diversity of a panel of known antibiotics or new antibiotic family.

3.4.1. *De novo syn-BNP for random discovery of antibiotics*

In 2016, the Brady group first accessed NRP BGCs from human microbiomes using the syn-BNP approach and humimycins from 25 synthesized linear syn-BNPs were found to be active against MRSA by inhibiting lipid II flippase (Chu et al., 2016). Then, this approach was extended to explore cyclic peptides for novel antibiotic discovery (Chu et al., 2020). 157 designed peptides with different cyclization modes were successfully synthesized and nine inhibited bacterial growth by diverse modes of action (Chu et al., 2020). In 2021, using penicillin binding protein-like cyclase as target, Hostetler et al., also developed a library of predicted head-to-tail cyclic peptides encoded by >500 NRP BGCs from underexplored actinobacteria and identified 14 antibacterial compounds (Hostetler et al., 2021). The traditional NP discovery paradigm that involves in fermentation of producers is laborious and time-consuming to identify and isolate metabolites, and even the best activation strategies can only unlock a small fraction of silent BGCs. On the contrast, the *syn*-BNP approach completely skips the paradigm and opens up new opportunities to unlock >90% silent BGCs under standard laboratory conditions.

3.4.2. *Targeted syn-BNP for the discovery of menaquinone-binding antibiotics*

Although the de novo syn-BNP approach can access novel peptides, some obvious obstacles, including the low success rate and moderate activity of syn-BNPs, are required to be overcome. Recently, a targeted syn-BNP approach was developed in the Brady group, which can be used to discover more potent compounds, including congeners of known AMPs and new antibiotic families (Li et al., 2022; Wang et al., 2021; Wang et al., 2022a, 2022b, 2022c). They developed a predicted NRP database that contains >36,000 linear peptides from >10,000 bacterial genomes (Li et al., 2022). Using the predicted database combined with sequence tag-based metagenomic mining, Li et al., developed a novel NP discovery approach, that is motif search of predicted structures, and identified six novel menaquinone-binding antibiotics (MBAs) from untapped Gram-negative bacteria or soil metagenomes (Fig. 4a). Furthermore, MBAs were firstly found to show potent activity against *Mycobacterium tuberculosis* by targeting menaquinone in bacterial membrane (Li et al., 2022). Identifying NPs with a desired MOA by searching for a substructure among a database of predicted structures

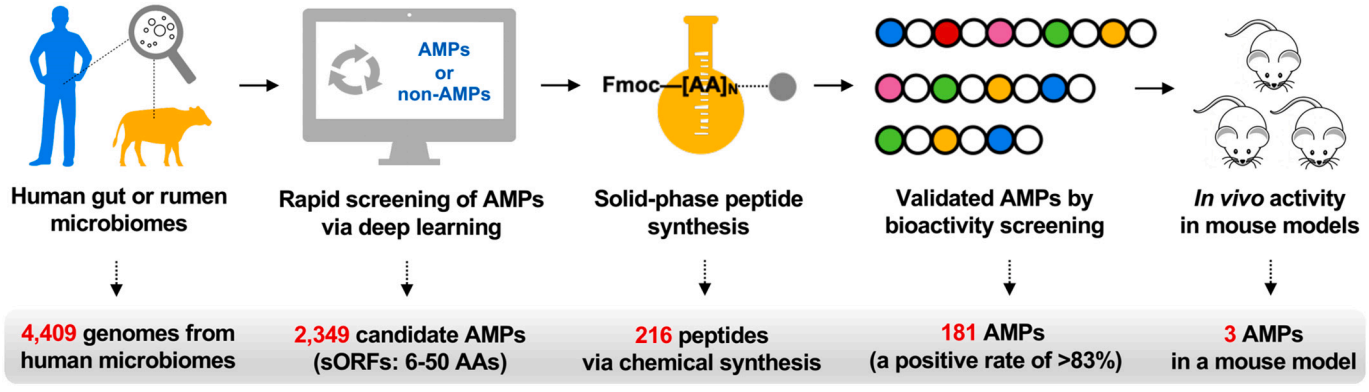

Fig. 3. AI-derived AMP Discovery. To develop a powerful AMP prediction pipeline, a panel of neural network models are required to be builded and optimized by collecting sequences of known AMPs to build training and test sets. Then, promising candidates are synthesized and tested against MDR bacteria in an animal model. A representative work from Ma et al. was shown for searching for novel AMPs from human gut microbiomes using deep learning and natural language processing. Abbreviations: AMP, anti-microbial peptides; sORFs, small open reading frames.

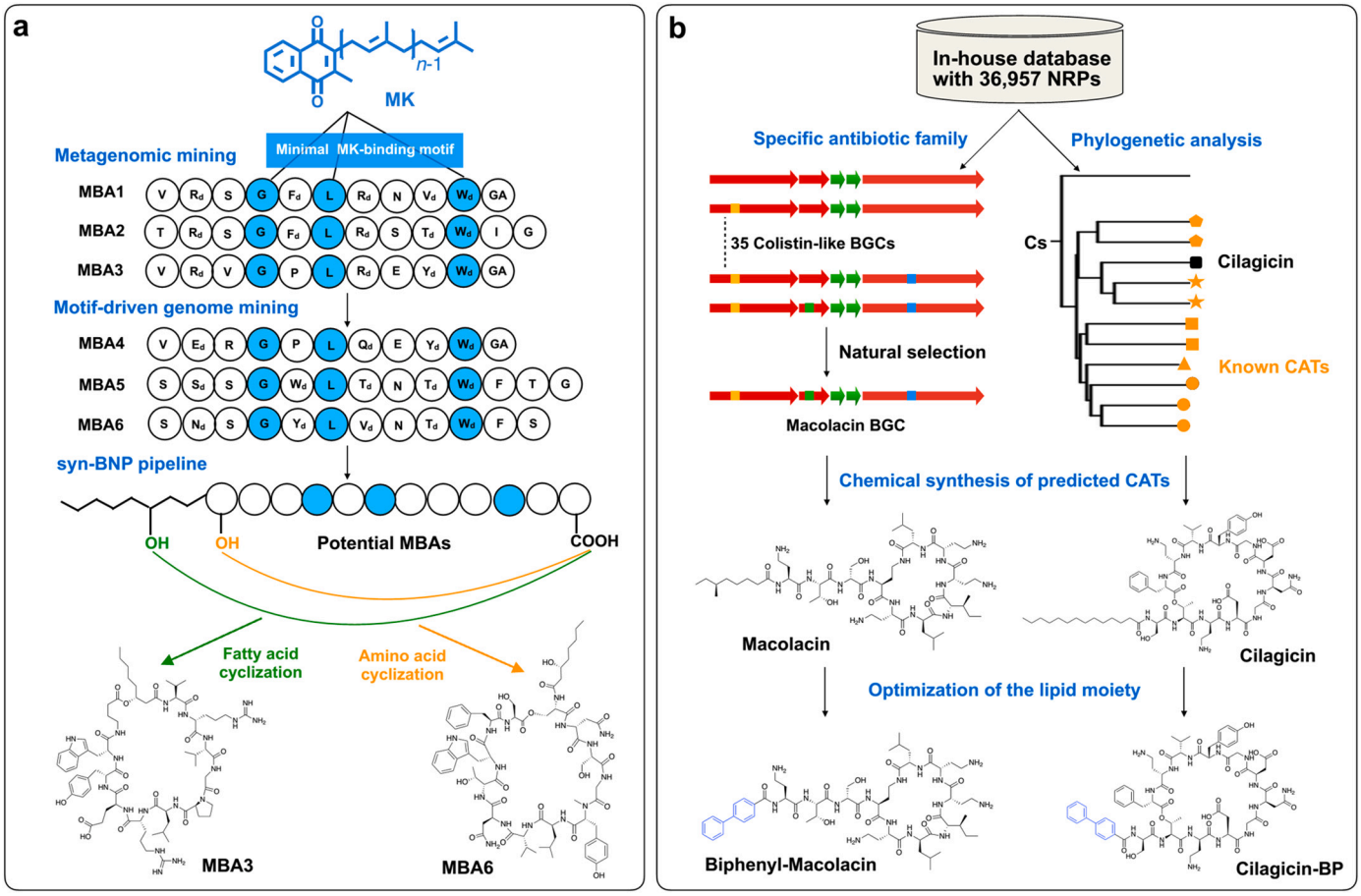

Fig. 4. Targeted *syn*-BNP for Antibiotic Discovery. (a) (Meta)genomic mining of menaquinone-binding antibiotics (MBAs). Two orthogonal pipelines, including sequence tag-based metagenomic mining and motif search-based genome mining, were used for MBA discovery. Linear MBAs were cyclized through the β-hydroxyl of the fatty acid or through a nucleophilic amino acid side chain. (b) Targeted genome mining of cationic lipopeptides (CATs). Based on the concept of natural selection, macolacin was identified by searching for colistin congeners to overcome its resistance. Phylogenetic analysis of Cs domain sequences of known or potential CATs was used to guide the discovery of cilagicin. To further optimize in vivo efficacy, biphenyl lipid moiety was introduced into macolacin or cilagicin. Abbreviations: syn-BNP, synthetic-Bioinformatics Natural Product; NRPs: non-ribosomal peptides; Cs: condensation start domain.

provides a generalizable and high-throughput approach for discovery of novel bioactive NPs.

3.4.3. Targeted syn-BNP for the discovery of cationic lipopeptide antibiotics

Cationic lipopeptides (CATs) are generally capable of penetrating the outer membranes of Gram-negative bacteria and interacting with different anionic intracellular targets, which show potent activity against a large variety of MDR Gram-negative pathogens (Liu et al., 2019). Intriguingly, a few CATs (e.g., Lysocin E and WAP-8492A2) have been proven to possess narrow-spectrum activity against Gram-positive pathogens (Yue et al., 2022). The diverse antimicrobial activities make CATs an appealing class of antibacterial agents for in-depth (meta) genomic mining. By systematically analyzing 5585 complete bacterial genomes, Li et al., firstly identified two Gram-negative-active CATs, brevicidine and laterocidine (Li et al., 2018). Instead of the time-consuming fermentation, the Brady group recently used the targeted *syn*-BNP approach to rapidly discover two novel CATs with unique features (Fig. 4b) (Wang et al., 2022a, 2022b). In the first study, to overcome colistin resistance, they set out to identify BGCs that encode the colistin congeners from the predicted NRP database (Wang et al., 2022a). Totally, 35 colistin-like BGCs were identified and the predicted product encoded by one of these BGCs, macolacin, was synthesized and found to be active against a variety of colistin-resistant clinical isolates. (Fig. 4b) (Wang et al., 2022a). In the second study, via phylogenetic analysis of condensation start domain, the same group identified a distinct CAT BGC that encodes cilagicin. Cilagicin caused cell death by directly binding to both undecaprenyl phosphate (C55—P) and undecaprenyl pyrophosphate (C55-PP). Notably, they did not detect resistance to cilagicin due to its dual MOA (Fig. 4b) (Wang et al., 2022b). The targeted syn-BNP approach provides an inspirational interdisciplinary and high-throughput pipeline for future antibiotic discovery to overcome antimicrobial resistance as well as high-throughput discovery of other drugs.

4. Concluding remarks

The study of microbial NPs can not only help to better understand their functions in native environment, but also discover more drug leads with novel chemical entities or distinct modes of action. With biological and technological advances in the last two decades, the cost of DNA sequencing has been dramatically decreased and the capacity of sequencing complex samples has been significantly improved. Thanks to this unprecedented access to (meta)genomes of unconventional microbial taxa and global microbiomes, we are witnessing a renaissance in natural product discovery for novel drug development. Major bottlenecks in accessing the hidden biosynthetic potential are being overcome by the enabling technologies discussed in this review (Table 2) and future increasing developments.

Several important questions remain and require to be addressed in the near future. For example, how to develop more effective strategies of

BGC prioritization for the discovery of new chemical entities? Big data-driven computational tools (e.g., BiG-SCAPE/CORASON) to analyze BGC novelty at scale and targeted (meta)genomic mining of a type of NPs (e.g., motif search of predicted NP database) could be efficient solutions for BGC prioritization. Will the cost of gene synthesis be decreased significantly in the near future? If so, it will be available to refactor large-size BGCs (>100 kb), such as non-ribosomal peptide and polyketide biosynthetic pathways. Additionally, how to improve prediction accuracy and decrease sequence biases of artificial intelligence (AI) driven discovery of antimicrobial peptides? Specially, is it possible to extend the AI technique or the syn-BNP approach to access other classes of NPs besides bioactive peptides from untapped microbial sources? Except for peptides, building blocks recognized by acyltransferase (AT) domains in type I polyketides (PKs) and hybrid polyketide-non-ribosomal peptide (PK-NRP) compounds could be predicted based on ATSignature or Minowa software with low confidence. However, it is expected that structures of PKs or hybrid PK-NPR molecules could be accurately predicted along with rapid development of AI and computational tools. Nevertheless, considering the efficiency and breadth of new (meta)genomic mining approaches and given the increased extent of underexplored sources, I believe that NP-based drug discovery and development will continue to make major contributions to human health.

Declaration of Competing Interest

The author declares no competing financial interest.

Acknowledgment

Research on natural product discovery, bioengineering and synthetic biology in the Li group is currently supported by Shanghai Pujiang Program (22PJ1406000) and Shanghai Jiao Tong University Start-up Funds.

References

Almeida, A., Nayfach, S., Boland, M., Strozzi, F., Beracochea, M., Shi, Z.J., Pollard, K.S., Sakharova, E., Parks, D.H., Hugenholtz, P., Segata, N., Kyrpides, N.C., Finn, R.D., 2021. A unified catalog of 204,938 reference genomes from the human gut microbiome. Nat. Biotechnol. 39 (1), 105–114.

Antimicrobial Resistance Collaborators, 2022. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. Lancet 399 (10325), 629–655.

Atanasov, A.G., Zotchev, S.B., Dirsch, V.M., International Natural Product Sciences Taskforce, Supuran, C.T., 2021. Natural products in drug discovery: advances and opportunities. Nat. Rev. Drug Discov. 20 (3), 200–216.

Bauman, K.D., Butler, K.S., Moore, B.S., Chekan, J.R., 2021. Genome mining methods to discover bioactive natural products. Nat. Prod. Rep. 38 (11), 2100–2129.

Bickhart, D.M., Kolmogorov, M., Tseng, E., Portik, D.M., Korobeynikov, A., Tolstoganov, I., Uritskiy, G., Liachko, I., Sullivan, S.T., Shin, S.B., Zorea, A., Andreu, V.P., Panke-Buisse, K., Medema, M.H., Mizrahi, I., Pevzner, P.A., Smith, T.P. L., 2022. Generating lineage-resolved, complete metagenome-assembled genomes from complex microbial communities. Nat. Biotechnol. 40 (5), 711–719.

Blin, K., Shaw, S., Kautsar, S.A., Medema, M.H., Weber, T., 2021. The antiSMASH database version 3: increased taxonomic coverage and new query features for modular enzymes. Nucleic Acids Res. 49 (D1), D639–D643.

Blin, K., Shaw, S., Augustijn, H.E., Reitz, Z.L., Biermann, F., Alanjary, M., Fetter, A., Terlouw, B.R., Metcalf, W.W., Helfrich, E.J.N., van Wezel, G.P., Medema, M.H., Weber, T., 2023. antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. Nucleic Acids Res. https://doi.org/10.1093/nar/gkad344.

Carroll, A.R., Copp, B.R., Davis, R.A., Keyzers, R.A., Prinsep, M.R., 2022. Marine natural products. Nat. Prod. Rep. 39 (6), 1122–1171.

Chevrette, M.G., Gavrilidou, A., Mantri, S., Selem-Mojica, N., Ziemert, N., Barona-Gómez, F., 2021. The confluence of big data and evolutionary genome mining for the discovery of natural products. Nat. Prod. Rep. 38 (11), 2024–2040.

Chu, J., Vila-Farres, X., Inoyama, D., Ternei, M., Cohen, L.J., Gordon, E.A., Reddy, B.V., Charlop-Powers, Z., Zebroski, H.A., Gallardo-Macias, R., Jaskowski, M., Satish, S., Park, S., Perlin, D.S., Freundlich, J.S., Brady, S.F., 2016. Discovery of MRSA active antibiotics using primary sequence from the human microbiome. Nat. Chem. Biol. 12 (12), 1004–1006.

Chu, J., Koirala, B., Forelli, N., Vila-Farres, X., Ternei, M.A., Ali, T., Colosimo, D.A., Brady, S.F., 2020. Synthetic-bioinformatic natural product antibiotics with diverse modes of action. J. Am. Chem. Soc. 142 (33), 14158–14168.

Conway, K.R., Boddy, C.N., 2013. ClusterMine360: a database of microbial PKS/NRPS biosynthesis. Nucleic Acids Res. 41, D402–D407.

Crits-Christoph, A., Diamond, S., Butterfield, C.N., Thomas, B.C., Banfield, J.F., 2018. Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis. Nature 558 (7710), 440–444.

Culp, E.J., Sychantha, D., Hobson, C., Pawlowski, A.C., Prehna, G., Wright, G.D., 2022. ClpP inhibitors are produced by a widespread family of bacterial gene clusters. Nat. Microbiol. 7 (3), 451–462.

Das, P., Sercu, T., Wadhawan, K., Padhi, I., Gehrmann, S., Cipcigan, F., Chenthamarakshan, V., Strobelt, H., Dos Santos, C., Chen, P.Y., Yang, Y.Y., Tan, J.P. K., Hedrick, J., Crain, J., Mojsilovic, A., 2021. Accelerated antimicrobial discovery via deep generative models and molecular dynamics simulations. Nat. Biomed. Eng. 5 (6), 613–623.

Enghiad, B., Huang, C., Guo, F., Jiang, G., Wang, B., Tabatabaei, S.K., Martin, T.A., Zhao, H., 2021. Cas12a-assisted precise targeted cloning using *in vivo* Cre-lox recombination. Nat. Commun. 12 (1), 1171.

Fisher, M.C., Alastruey-Izquierdo, A., Berman, J., Bicanic, T., Bignell, E.M., Bowyer, P., Bromley, M., Brüggemann, R., Garber, G., Cornely, O.A., Gurr, S.J., Harrison, T.S., Kuijper, E., Rhodes, J., Sheppard, D.C., Warris, A., White, P.L., Xu, J., Zwaan, B., Verweij, P.E., 2022. Tackling the emerging threat of antifungal resistance to human health. Nat. Rev. Microbiol. 20 (9), 557–571.

Gallo, K., Kemmler, E., Goede, A., Becker, F., Dunkel, M., Preissner, R., Banerjee, P., 2023. SuperNatural 3.0-a database of natural products and natural product-based derivatives. Nucleic Acids Res. 51 (D1), D654–D659.

Gavriilidou, A., Kautsar, S.A., Zaburannyi, N., Krug, D., Müller, R., Medema, M.H., Ziemert, N., 2022. Compendium of specialized metabolite biosynthetic diversity encoded in bacterial genomes. Nat. Microbiol. 7 (5), 726–735.

Grigoriev, I.V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., Riley, R., Salamov, A., Zhao, X., Korzeniewski, F., Smirnova, T., Nordberg, H., Dubchak, I., Shabalov, I., 2014. MycoCosm portal: gearing up for 1000 fungal genomes. Nucleic Acids Res. 42, D699–D704.

Hao, T., Xie, Z., Wang, M., Liu, L., Zhang, Y., Wang, W., Zhang, Z., Zhao, X., Li, P., Guo, Z., Gao, S., Lou, C., Zhang, G., Merritt, J., Horsman, G.P., Chen, Y., 2019. An anaerobic bacterium host system for heterologous expression of natural product biosynthetic gene clusters. Nat. Commun. 10 (1), 3665.

Harvey, C.J.B., Tang, M., Schlecht, U., Horecka, J., Fischer, C.R., Lin, H.C., Li, J., Naughton, B., Cherry, J., Miranda, M., Li, Y.F., Chu, A.M., Hennessy, J.R., Vandova, G.A., Inglis, D., Aiyar, R.S., Steinmetz, L.M., Davis, R.W., Medema, M.H., Sattely, E., Khosla, C., St Onge, R.P., Tang, Y., Hillenmeyer, M.E., 2018. HEx: a heterologous expression platform for the discovery of fungal natural products. Sci. Adv. 4 (4), eaar5459.

Hemmerling, F., Piel, J., 2022. Strategies to access biosynthetic novelty in bacterial genomes for drug discovery. Nat. Rev. Drug Discov. 21 (5), 359–378.

Hostetler, M.A., Smith, C., Nelson, S., Budimir, Z., Modi, R., Woolsey, I., Frerk, A., Baker, B., Gantt, J., Parkinson, E.I., 2021. Synthetic natural product inspired cyclic peptides. ACS Chem. Biol. 16 (11), 2604–2611.

Hug, J.J., Krug, D., Müller, R., 2020. Bacteria as genetically programmable producers of bioactive natural products. Nat. Rev. Chem. 4, 172–193.

Imai, Y., Meyer, K.J., Iinishi, A., Favre-Godal, Q., Green, R., Manuse, S., Caboni, M., Mori, M., Niles, S., Ghiglieri, M., Honrao, C., Ma, X., Guo, J.J., Makriyannis, A., Linares-Otoya, L., Böhringer, N., Wuisan, Z.G., Kaur, H., Wu, R., Mateus, A., Typas, A., Savitski, M.M., Espinoza, J.L., O'Rourke, A., Nelson, K.E., Hiller, S., Noinaj, N., Schäberle, T.F., D'Onofrio, A., Lewis, K., 2019. A new antibiotic selectively kills Gram-negative pathogens. Nature 576 (7787), 459–464.

Imai, Y., Hauk, G., Quigley, J., Liang, L., Son, S., Ghiglieri, M., Gates, M.F., Morrissette, M., Shahsavari, N., Niles, S., Baldisseri, D., Honrao, C., Ma, X., Guo, J.J., Berger, J.M., Lewis, K., 2022. Evybactin is a DNA gyrase inhibitor that selectively kills *Mycobacterium tuberculosis*. Nat. Chem. Biol. 18 (11), 1236–1244.

Jiang, C., Lv, G., Tu, Y., Cheng, X., Duan, Y., Zeng, B., He, B., 2021. Applications of CRISPR/Cas9 in the synthesis of secondary metabolites in filamentous fungi. Front. Microbiol. 12, 638096.

Kalkreuter, E., Pan, G., Cepeda, A.J., Shen, B., 2020. Targeting bacterial genomes for natural product discovery. Trends Pharmacol. Sci. 41 (1), 13–26.

Kaur, H., Jakob, R.P., Marzinek, J.K., Green, R., Imai, Y., Bolla, J.R., Agustoni, E., Robinson, C.V., Bond, P.J., Lewis, K., Maier, T., Hiller, S., 2021. The antibiotic darobactin mimics a β-strand to inhibit outer membrane insertase. Nature 593 (7857), 125–129.

Kautsar, S.A., Blin, K., Shaw, S., Navarro-Muñoz, J.C., Terlouw, B.R., van der Hooft, J.J. J., van Santen, J.A., Tracanna, V., Suarez Duran, H.G., Pascal Andreu, V., Selem-Mojica, N., Alanjary, M., Robinson, S.L., Lund, G., Epstein, S.C., Sisto, A.C., Charkoudian, L.K., Collemare, J., Linington, R.G., Weber, T., Medema, M.H., 2020. MIBiG 2.0: a repository for biosynthetic gene clusters of known function. Nucleic Acids Res. 48 (D1), D454–D458.

Kautsar, S.A., Blin, K., Shaw, S., Weber, T., Medema, M.H., 2021a. BiG-FAM: the biosynthetic gene cluster families database. Nucleic Acids Res. 49 (D1), D490–D497.

Kautsar, S.A., van der Hooft, J.J.J., de Ridder, D., Medema, M.H., 2021b. BiG-SLiCE: a highly scalable tool maps the diversity of 1.2 million biosynthetic gene clusters. Gigascience 10 (1), giaa154.

Keller, N.P., 2019. Fungal secondary metabolism: regulation, function and drug discovery. Nat. Rev. Microbiol. 17 (3), 167–180.

Klementz, D., Döring, K., Lucas, X., Telukunta, K.K., Erxleben, A., Deubel, D., Erber, A., Santillana, I., Thomas, O.S., Bechthold, A., Günther, S., 2016. StreptomeDB 2.0–an extended resource of natural products produced by streptomycetes. Nucleic Acids Res. 44 (D1), D509–D514.

Lee, S.R., Seyedsayamdost, M.R., 2022. Induction of diverse cryptic fungal metabolites by steroids and channel blockers. Angew. Chem. Int. Ed. Eng. 61 (29), e202204519.

Lewis, K., 2021. The science of antibiotic discovery. Cell 181 (1), 29–45.
Li, Y.X., Zhong, Z., Zhang, W.P., Qian, P.Y., 2018. Discovery of cationic nonribosomal peptides as Gram-negative antibiotics through global genome mining. Nat. Commun. 9 (1), 3273.
Li, L., Koirala, B., Hernandez, Y., MacIntyre, L.W., Ternei, M.A., Russo, R., Brady, S.F., 2022. Identification of structurally diverse menaquinone-binding antibiotics with *in vivo* activity against multidrug-resistant pathogens. Nat. Microbiol. 7 (1), 120–131.
Liang, M., Liu, L., Xu, F., Zeng, X., Wang, R., Yang, J., Wang, W., Karthik, L., Liu, J., Yang, Z., Zhu, G., Wang, S., Bai, L., Tong, Y., Liu, X., Wu, M., Zhang, L.X., Tan, G.Y., 2022. Activating cryptic biosynthetic gene cluster through a CRISPR-Cas12a-mediated direct cloning approach. Nucleic Acids Res. 50 (6), 3581–3592.
Libis, V., Antonovsky, N., Zhang, M., Shang, Z., Montiel, D., Maniko, J., Ternei, M.A., Calle, P.Y., Lemetre, C., Owen, J.G., Brady, S.F., 2019. Uncovering the biosynthetic potential of rare metagenomic DNA using co-occurrence network analysis of targeted sequences. Nat. Commun. 10 (1), 3848.
Liu, Y., Ding, S., Shen, J., Zhu, K., 2019. Nonribosomal antibacterial peptides that target multidrug-resistant bacteria. Nat. Prod. Rep. 36 (4), 573–592.
Liu, J., Zhou, H., Yang, Z., Wang, X., Chen, H., Zhong, L., Zheng, W., Niu, W., Wang, S., Ren, X., Zhong, G., Wang, Y., Ding, X., Müller, R., Zhang, Y., Bian, X., 2021. Rational construction of genome-reduced Burkholderiales chassis facilitates efficient heterologous production of natural products from proteobacteria. Nat. Commun. 12 (1), 4347.
Liu, Y., Ji, M., Yu, T., Zaugg, J., Anesio, A.M., Zhang, Z., Hu, S., Hugenholtz, P., Liu, K., Liu, P., Chen, Y., Luo, Y., Yao, T., 2022a. A genome and gene catalog of glacier microbiomes. Nat. Biotechnol. 40 (9), 1341–1348.
Liu, J., Wang, X., Dai, G., Zhang, Y., Bian, X., 2022b. Microbial chassis engineering drives heterologous production of complex secondary metabolites. Biotechnol. Adv. 59, 107966.
Ma, Y., Guo, Z., Xia, B., Zhang, Y., Liu, X., Yu, Y., Tang, N., Tong, X., Wang, M., Ye, X., Feng, J., Chen, Y., Wang, J., 2022. Identification of antimicrobial peptides from the human gut microbiome using deep learning. Nat. Biotechnol. 40 (6), 921–931.
Medema, M.H., de Rond, T., Moore, B.S., 2021. Mining genomes to illuminate the specialized chemistry of life. Nat. Rev. Genet. 22 (9), 553–571.
Meng, X., Fang, Y., Ding, M., Zhang, Y., Jia, K., Li, Z., Collemare, J., Liu, W., 2022. Developing fungal heterologous expression platforms to explore and improve the production of natural products from fungal biodiversity. Biotechnol. Adv. 54, 107866.
Miethke, M., Pieroni, M., Weber, T., Brönstrup, M., Hammann, P., Halby, L., Arimondo, P.B., Glaser, P., Aigle, B., Bode, H.B., Moreira, R., Li, Y., Luzhetskyy, A., Medema, M.H., Pernodet, J.L., Stadler, M., Tormo, J.R., Genilloud, O., Truman, A. W., Weissman, K.J., Takano, E., Sabatini, S., Stegmann, E., Brötz-Oesterhelt, H., Wohlleben, W., Seemann, M., Empting, M., Hirsch, A.K.H., Loretz, B., Lehr, C.M., Titz, A., Herrmann, J., Jaeger, T., Alt, S., Hesterkamp, T., Winterhalter, M., Schiefer, A., Pfarr, K., Hoerauf, A., Graz, H., Graz, M., Lindvall, M., Ramurthy, S., Karlén, A., van Dongen, M., Petkovic, H., Keller, A., Peyrane, F., Donadio, S., Fraisse, L., Piddock, L.J.V., Gilbert, I.H., Moser, H.E., Müller, R., 2021. Towards the sustainable discovery and development of new antibiotics. Nat. Rev. Chem. 5 (10), 726–749.
Mungan, M.D., Alanjary, M., Blin, K., Weber, T., Medema, M.H., Ziemert, N., 2020. ARTS 2.0: feature updates and expansion of the antibiotic resistant target seeker for comparative genome mining. Nucleic Acids Res. 48 (W1), W546–W552.
Mungan, M.D., Blin, K., Ziemert, N., 2022. ARTS-DB: a database for antibiotic resistant targets. Nucleic Acids Res. 50 (D1), D736–D740.
Navarro-Muñoz, J.C., Selem-Mojica, N., Mullowney, M.W., Kautsar, S.A., Tryon, J.H., Parkinson, E.I., De Los Santos, E.L.C., Yeong, M., Cruz-Morales, P., Abubucker, S., Roeters, A., Lokhorst, W., Fernandez-Guerra, A., Cappelini, L.T.D., Goering, A.W., Thomson, R.J., Metcalf, W.W., Kelleher, N.L., Barona-Gomez, F., Medema, M.H., 2020. A computational framework to explore large-scale biosynthetic diversity. Nat. Chem. Biol. 16 (1), 60–68.
Nayfach, S., Roux, S., Seshadri, R., Udwary, D., Varghese, N., Schulz, F., Wu, D., Paez-Espino, D., Chen, I.M., Huntemann, M., Palaniappan, K., Ladau, J., Mukherjee, S., Reddy, T.B.K., Nielsen, T., Kirton, E., Faria, J.P., Edirisinghe, J.N., Henry, C.S., Jungbluth, S.P., Chivian, D., Dehal, P., Wood-Charlson, E.M., Arkin, A.P., Tringe, S. G., Visel, A., IMG/M Data Consortium, Woyke, T., Mouncey, N.J., Ivanova, N.N., Kyrpides, N.C., Eloe-Fadrosh, E.A., 2021. A genomic catalog of Earth's microbiomes. Nat. Biotechnol. 39 (4), 499–509.
Newman, D.J., Cragg, G.M., 2020. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. J. Nat. Prod. 83 (3), 770–803.
Oyama, L.B., Olleik, H., Teixeira, A.C.N., Guidini, M.M., Pickup, J.A., Hui, B.Y.P., Vidal, N., Cookson, A.R., Vallin, H., Wilkinson, T., Bazzolli, D.M.S., Richards, J., Wootton, M., Mikut, R., Hilpert, K., Maresca, M., Perrier, J., Hess, M., Mantovani, H. C., Fernandez-Fuentes, N., Creevey, C.J., Huws, S.A., 2022. *In silico* identification of two peptides with antibacterial activity against multidrug-resistant *Staphylococcus aureus*. NPJ Biofilms Microb. 8 (1), 58.
Palaniappan, K., Chen, I.A., Chu, K., Ratner, A., Seshadri, R., Kyrpides, N.C., Ivanova, N. N., Mouncey, N.J., 2020. IMG-ABC v.5.0: an update to the IMG/Atlas of biosynthetic gene clusters knowledgebase. Nucleic Acids Res. 48 (D1), D422–D430.
Panter, F., Bader, C.D., Müller, R., 2021. Synergizing the potential of bacterial genomics and metabolomics to find novel antibiotics. Chem. Sci. 12 (17), 5994–6010.
Paoli, L., Ruscheweyh, H.J., Forneris, C.C., Hubrich, F., Kautsar, S., Bhushan, A., Lotti, A., Clayssen, Q., Salazar, G., Milanese, A., Carlström, C.I., Papadopoulou, C., Gehrig, D., Karasikov, M., Mustafa, H., Larralde, M., Carroll, L.M., Sánchez, P., Zayed, A.A., Cronin, D.R., Acinas, S.G., Bork, P., Bowler, C., Delmont, T.O., Gasol, J. M., Gossert, A.D., Kahles, A., Sullivan, M.B., Wincker, P., Zeller, G., Robinson, S.L., Piel, J., Sunagawa, S., 2022. Biosynthetic potential of the global ocean microbiome. Nature 607 (7917), 111–118.
Pascal Andreu, V., Roel-Touris, J., Dodd, D., Fischbach, M.A., Medema, M.H., 2021. The gutSMASH web server: automated identification of primary metabolic gene clusters from the gut microbiota. Nucleic Acids Res. 49 (W1), W263–W270.
Patel, J.R., Oh, J., Wang, S., Crawford, J.M., Isaacs, F.J., 2022. Cross-kingdom expression of synthetic genetic elements promotes discovery of metabolites in the human microbiome. Cell 185 (9), 1487–1505.e14.
Privalsky, T.M., Soohoo, A.M., Wang, J., Walsh, C.T., Wright, G.D., Gordon, E.M., Gray, N.S., Khosla, C., 2021. Prospects for antibacterial discovery and development. J. Am. Chem. Soc. 143 (50), 21127–21142.
Robey, M.T., Caesar, L.K., Drott, M.T., Keller, N.P., Kelleher, N.L., 2021. An interpreted atlas of biosynthetic gene clusters from 1,000 fungal genomes. Proc. Natl. Acad. Sci. U. S. A. 118 (19), e2020230118.
Roux, I., Chooi, Y.H., 2022. Cre/lox-mediated chromosomal integration of biosynthetic gene clusters for heterologous expression in *Aspergillus nidulans*. ACS Synth. Biol. 11 (3), 1186–1195.
Rubin, B.E., Diamond, S., Cress, B.F., Crits-Christoph, A., Lou, Y.C., Borges, A.L., Shivram, H., He, C., Xu, M., Zhou, Z., Smith, S.J., Rovinsky, R., Smock, D.C.J., Tang, K., Owens, T.K., Krishnappa, N., Sachdeva, R., Barrangou, R., Deutschbauer, A. M., Banfield, J.F., Doudna, J.A., 2022. Species- and site-specific genome editing in complex bacterial communities. Nat. Microbiol. 7 (1), 34–47.
Rutledge, P.J., Challis, G.L., 2015. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. Nat. Rev. Microbiol. 13, 509–523.
Saldívar-González, F.I., Aldas-Bulos, V.D., Medina-Franco, J.L., Plisson, F., 2021. Natural product drug discovery in the artificial intelligence era. Chem. Sci. 13 (6), 1526–1546.
Scherlach, K., Hertweck, C., 2021. Mining and unearthing hidden biosynthetic potential. Nat. Commun. 12 (1), 3864.
Scott, T.A., Piel, J., 2019. The hidden enzymology of bacterial natural product biosynthesis. Nat. Rev. Chem. 3 (7), 404–425.
Skinnider, M.A., Merwin, N.J., Johnston, C.W., Magarvey, N.A., 2017. PRISM 3: expanded prediction of natural product chemical structures from microbial genomes. Nucleic Acids Res. 45 (W1), W49–W54.
Stokes, J.M., Yang, K., Swanson, K., Jin, W., Cubillos-Ruiz, A., Donghia, N.M., MacNair, C.R., French, S., Carfrae, L.A., Bloom-Ackermann, Z., Tran, V.M., Chiappino-Pepe, A., Badran, A.H., Andrews, I.W., Chory, E.J., Church, G.M., Brown, E.D., Jaakkola, T.S., Barzilay, R., Collins, J.J., 2020. A deep learning approach to antibiotic discovery. Cell 180 (4), 688–702.e13.
Süssmuth, R.D., Mainz, A., 2017. Nonribosomal peptide synthesis-principles and prospects. Angew. Chem. Int. Ed. Eng. 56 (14), 3770–3821.
Tao, H., Lauterbach, L., Bian, G., Chen, R., Hou, A., Mori, T., Cheng, S., Hu, B., Lu, L., Mu, X., Li, M., Adachi, N., Kawasaki, M., Moriya, T., Senda, T., Wang, X., Deng, Z., Abe, I., Dickschat, J.S., Liu, T., 2022. Discovery of non-squalene triterpenes. Nature 606 (7913), 414–419.
Torres, M.D.T., Melo, M.C.R., Crescenzi, O., Notomista, E., de la Fuente-Nunez, C., 2022. Mining for encrypted peptide antibiotics in the human proteome. Nat. Biomed. Eng. 6 (1), 67–75.
van Santen, J.A., Poynton, E.F., Iskakova, D., McMann, E., Alsup, T.A., Clark, T.N., Fergusson, C.H., Fewer, D.P., Hughes, A.H., McCadden, C.A., Parra, J., Soldatou, S., Rudolf, J.D., Janssen, E.M., Duncan, K.R., Linington, R.G., 2022. The natural products atlas 2.0: a database of microbially-derived natural products. Nucleic Acids Res. 50 (D1), D1317–D1323.
Walsh, C.T., Fischbach, M.A., 2010. Natural products version 2.0: connecting genes to molecules. J. Am. Chem. Soc. 132, 2469–2493.
Wang, M., Carver, J.J., Phelan, V.V., Sanchez, L.M., Garg, N., Peng, Y., Nguyen, D.D., Watrous, J., Kapono, C.A., Luzzatto-Knaan, T., Porto, C., Bouslimani, A., Melnik, A. V., Meehan, M.J., Liu, W.T., Crüsemann, M., Boudreau, P.D., Esquenazi, E., Sandoval-Calderón, M., Kersten, R.D., Pace, L.A., Quinn, R.A., Duncan, K.R., Hsu, C. C., Floros, D.J., Gavilan, R.G., Kleigrewe, K., Northen, T., Dutton, R.J., Parrot, D., Carlson, E.E., Aigle, B., Michelsen, C.F., Jelsbak, L., Sohlenkamp, C., Pevzner, P., Edlund, A., McLean, J., Piel, J., Murphy, B.T., Gerwick, L., Liaw, C.C., Yang, Y.L., Humpf, H.U., Maansson, M., Keyzers, R.A., Sims, A.C., Johnson, A.R., Sidebottom, A. M., Sedio, B.E., Klitgaard, A., Larson, C.B.P.C.A.B., Torres-Mendoza, D., Gonzalez, D. J., Silva, D.B., Marques, L.M., Demarque, D.P., Pociute, E., O'Neill, E.C., Briand, E., Helfrich, E.J.N., Granatosky, E.A., Glukhov, E., Ryffel, F., Houson, H., Mohimani, H., Kharbush, J.J., Zeng, Y., Vorholt, J.A., Kurita, K.L., Charusanti, P., McPhail, K.L., Nielsen, K.F., Vuong, L., Elfeki, M., Traxler, M.F., Engene, N., Koyama, N., Vining, O. B., Baric, R., Silva, R.R., Mascuch, S.J., Tomasi, S., Jenkins, S., Macherla, V., Hoffman, T., Agarwal, V., Williams, P.G., Dai, J., Neupane, R., Gurr, J., Rodríguez, A. M.C., Lamsa, A., Zhang, C., Dorrestein, K., Duggan, B.M., Almaliti, J., Allard, P.M., Phapale, P., Nothias, L.F., Alexandrov, T., Litaudon, M., Wolfender, J.L., Kyle, J.E., Metz, T.O., Peryea, T., Nguyen, D.T., VanLeer, D., Shinn, P., Jadhav, A., Müller, R., Waters, K.M., Shi, W., Liu, X., Zhang, L., Knight, R., Jensen, P.R., Palsson, B.O., Pogliano, K., Linington, R.G., Gutiérrez, M., Lopes, N.P., Gerwick, W.H., Moore, B.S., Dorrestein, P.C., Bandeira, N., 2016. Sharing and community curation of mass spectrometry data with global natural products social molecular networking. Nat. Biotechnol. 34 (8), 828–837.
Wang, Z., Kasper, A., Mehmood, R., Ternei, M., Li, S., Freundlich, J.S., Brady, S.F., 2021. Metagenome-guided analogue synthesis yields improved gram-negative-active albicidin- and cystobactamid-type antibiotics. Angew. Chem. Int. Ed. Eng. 60 (41), 22172–22177.
Wang, Z., Forelli, N., Hernandez, Y., Ternei, M., Brady, S.F., 2022a. Lapcin, a potent dual topoisomerase I/II inhibitor discovered by soil metagenome guided total chemical synthesis. Nat. Commun. 13 (1), 842.
Wang, Z., Koirala, B., Hernandez, Y., Zimmerman, M., Brady, S.F., 2022b. Bioinformatic prospecting and synthesis of a bifunctional lipopeptide antibiotic that evades resistance. Science 376 (6596), 991–996.

Wang, Z., Koirala, B., Hernandez, Y., Zimmerman, M., Park, S., Perlin, D.S., Brady, S.F., 2022c. A naturally inspired antibiotic to target multidrug-resistant pathogens. Nature 601 (7894), 606–611.

Xu, F., Wu, Y., Zhang, C., Davis, K.M., Moon, K., Bushin, L.B., Seyedsayamdost, M.R., 2019. A genetics-free method for high-throughput discovery of cryptic microbial metabolites. Nat. Chem. Biol. 15 (2), 161–168.

Yan, Y., Liu, Q., Zang, X., Yuan, S., Bat-Erdene, U., Nguyen, C., Gan, J., Zhou, J., Jacobsen, S.E., Tang, Y., 2018. Resistance-gene-directed discovery of a natural-product herbicide with a new mode of action. Nature 559 (7714), 415–418.

Yuan, Y.J., Cheng, S., Bian, G.K., Yan, P., Ma, Z.N., Dai, W., Chen, R., Fu, S., Huang, H. W., Chi, H.M., Cai, Y.S., Deng, Z.X., Liu, T.G., 2022. Efficient exploration of terpenoid biosynthetic gene clusters in filamentous fungi. Nat. Catal. 5, 277–287.

Yue, H., Miller, A.L., Khetrapal, V., Jayaseker, V., Wright, S., Du, L., 2022. Biosynthesis, regulation, and engineering of natural products from *Lysobacter*. Nat. Prod. Rep. 39 (4), 842–874.

Zeng, X., Zhang, P., He, W., Qin, C., Chen, S., Tao, L., Wang, Y., Tan, Y., Gao, D., Wang, B., Chen, Z., Chen, W., Jiang, Y.Y., Chen, Y.Z., 2018. NPASS: natural product activity and species source database for natural product research, discovery and tool development. Nucleic Acids Res. 46 (D1), D1217–D1222.

Zhang, F., Zhao, M., Braun, D.R., Ericksen, S.S., Piotrowski, J.S., Nelson, J., Peng, J., Ananiev, G.E., Chanana, S., Barns, K., Fossen, J., Sanchez, H., Chevrette, M.G., Guzei, I.A., Zhao, C., Guo, L., Tang, W., Currie, C.R., Rajski, S.R., Audhya, A., Andes, D.R., Bugni, T.S., 2022. A marine microbiome antifungal targets urgent-threat drug-resistant fungi. Science 370 (6519), 974–978.