Chinese Journal of Natural Medicines

Recent advances in the culture-independent discovery of natural products using metagenomic approaches

Yiping SHEN, Nan LIU, Zongqiang WANG

Citation: Yiping SHEN, Nan LIU, Zongqiang WANG, Recent advances in the culture-independent discovery of natural products using metagenomic approaches, Chinese Journal of Natural Medicines, 2024, 22(2), 1-12. doi: 10.1016/S1875-5364(23)60498-4.

View online: https://doi.org/10.1016/S1875-5364(23)60498-4

Available online at www.sciencedirect.com
ScienceDirect
Chinese Journal of Natural Medicines 2024, **22**(2): 100-111
doi: 10.1016/S1875-5364(24)60585-6
Chinese Journal of Natural Medicines

•Review•

Recent advances in the culture-independent discovery of natural products using metagenomic approaches

SHEN Yiping, LIU Nan, WANG Zongqiang*

Laboratory of Microbial Drug Discovery, China Pharmaceutical University, Nanjing 211198, China

Available online 20 Feb., 2024

[ABSTRACT] Natural products derived from bacterial sources have long been pivotal in the discovery of drug leads. However, the cultivation of only about 1% of bacteria in laboratory settings has left a significant portion of biosynthetic diversity hidden within the genomes of uncultured bacteria. Advances in sequencing technologies now enable the exploration of genetic material from these metagenomes through culture-independent methods. This approach involves extracting genetic sequences from environmental DNA and applying a hybrid methodology that combines functional screening, sequence tag-based homology screening, and bioinformatic-assisted chemical synthesis. Through this process, numerous valuable natural products have been identified and synthesized from previously uncharted metagenomic territories. This paper provides an overview of the recent advancements in the utilization of culture-independent techniques for the discovery of novel biosynthetic gene clusters and bioactive small molecules within metagenomic libraries.

[KEY WORDS] Metagenomics; Culture-independent approach; Natural product; Functional screening; Sequence tag-based homology screening; Syn-BNP

[CLC Number] R284 **[Document code]** A **[Article ID]** 2095-6975(2024)02-0100-12

Introduction

The environment serves as a rich reservoir of microbial entities, with their secondary metabolites, widely recognized as natural products, garnering considerable research interest due to their varied and substantial medicinal properties. Over the past few decades, the extraction of microbial-derived drugs primarily relied on culture-dependent techniques, focusing on bacteria amenable to laboratory cultivation [1]. This process, encompassing cultivation, fermentation, and activity screening, has been instrumental in identifying significant drug leads. Despite its past efficacy, the primary limitation currently lies in the minimal proportion of bacteria that are cultured and the frequent rediscovery of known compounds, which impedes the progress in novel drug discovery [2]. Advances in sequencing technologies have unveiled that over 99% of environmental bacteria elude cultivation under standard laboratory conditions [3], suggesting that a vast array of biosynthetic diversity within the genomes of these uncultured bacteria remains largely unexplored.

Numerous strategies have been employed to access this expansive resource, including the exploration of suitable media for cultivating elusive bacteria [4], the invention of tools to culture these organisms *in situ* [5], and the adoption of culture-independent methodologies. The advent of metagenomics, bolstered by sequencing advancements, has facilitated access to the genetic information of uncultured bacteria. Environmental DNA (eDNA) is extracted from environmental samples and processed through metagenomic libraries to yield physical DNA sequences [6]. In the absence of prior knowledge about specific natural products, functional screening is applied to these metagenomic libraries to identify bioactive compounds based on the properties of the molecules screened [7]. A sequence tag-based homology screening approach is employed to target biosynthetic gene clusters (BGCs) within the vast metagenomic pool, focusing on conserved regions within reference genes. This technique aids in the retrieval of complete BGCs, which are subsequently introduced into suitable model hosts for heterologous expression [8]. Additionally, deep sequencing of eDNA from metagenomic libraries, coupled with bioinformatic algorithms, allows for the prediction and subsequent chemical synthesis of natural product analogs from targeted BGCs [9]. This culture-independent methodology not only circumvents the limitations of bacterial cultivability but also broadens our understanding of microbial genetic diversity and enhances the potential for novel natural product discovery.

The fundamental difference between culture-dependent and independent approaches is the source of targeted molecules. The former relies on cultivating and screening native

[Received on] 24-Aug.-2023
[Research funding] This work was supported by the Start-Up Grant from China Pharmaceutical University (No. 3154070046).
[*Corresponding author] E-mail: zqwang02@cpu.edu.cn
These authors have no conflict of interest to declare.

hosts to biosynthesize these molecules, whereas the latter involves extracting genetic sequences from eDNA, expressing them in model hosts, or employing chemical synthesis based on bioinformatics predictions. This review aims to highlight recent progress in culture-independent metagenomics, particularly in the discovery of bioactive natural products and novel BGCs from environmental samples, providing valuable insights and references for future advancements in the field of metagenomics.

Metagenomic Library Construction

Bacteria are a prolific source of pharmaceutically relevant natural products [10-12]. However, the inability to culture the majority of natural bacterial species in laboratory settings presents a significant challenge for obtaining these products through traditional isolation methods, such as fermentation [13-16]. Compared with other genetic materials like proteins and RNA, which are produced by uncultured bacteria, environmental DNA (eDNA) is more accessible in laboratory conditions. The extraction of eDNA offers a feasible approach for exploring the taxonomy and functional diversity of bioactive small molecules encoded within the DNA of uncultured bacteria [17]. The culture-independent method, which involves heterologous expression of eDNA in a culturable model host, stands out as an innovative strategy to study the diversity of metagenome-derived natural products [18]. The eDNA samples can be extracted using direct or indirect extraction methods [19]. The direct method involves lysing whole bacterial cells in environmental samples and precipitating the eDNA using an organic solvent like isopropanol [20]. In contrast, the indirect method first isolates bacteria from environmental samples through density gradient centrifugation. These bacteria are then embedded in agarose gel for lysis, a technique designed to minimize mechanical damage [21]. Although the indirect method can yield larger eDNA fragments, it generally produces a lower quantity relative to the direct method and involves a more complex and time-consuming library construction process.

Depending on the specific downstream applications, the acquisition of metagenomic DNA can be categorized into two primary methods: DNA entity sequence-dependent and DNA entity sequence-independent methods. In the sequence-dependent method, metagenomic DNA is physicochemically captured using a shuttle vector, after which it is transformed into an appropriate host. This process facilitates either long-term storage or further analyses, such as functional and sequence-based screening for bioactive natural products (Fig. 1). Currently, cosmid, fosmid, and bacterial artificial chromosome (BAC) vectors are the three main types of vectors employed in constructing metagenomic libraries, each differing in their DNA insert size capacity and copy number. Cosmid and fosmid vectors incorporate phage mobilization elements, allowing them to shuttle between a phage and an *E. coli* host. A notable difference between these two vectors is the incorporation of an F-origin element in fosmids, which restricts the number of fosmids to one per cell [22]. This unique feature is crucial as it prevents the recombination of highly repetitive sequences within the host cell, thereby avoiding undesirable deletions or cell death [23]. Moreover, the lower copy number of these vectors reduces the production of potentially toxic metabolites. Due to their high efficiency in phage transfection, cosmid and fosmid vectors are adept at capturing a wide range of BGCs with diverse biosynthetic capabilities. The Brady group demonstrated this through a detailed protocol for constructing metagenomic libraries using a commercial cosmid vector (pWEB), successfully generating millions of cosmid clones from less than 1 kg of soil [6]. Despite these advantages [24], a significant limitation of these vectors is their insert size, typically 35−45 kb, which may not cover the full length of larger BGCs that can extend beyond 100 kilobases [25]. To address this issue, several assembly strategies have been developed for reconstructing complete gene clusters from eDNA-derived cosmid clones, including restriction-site ligation, λ-based recombination, and transformation-associ-

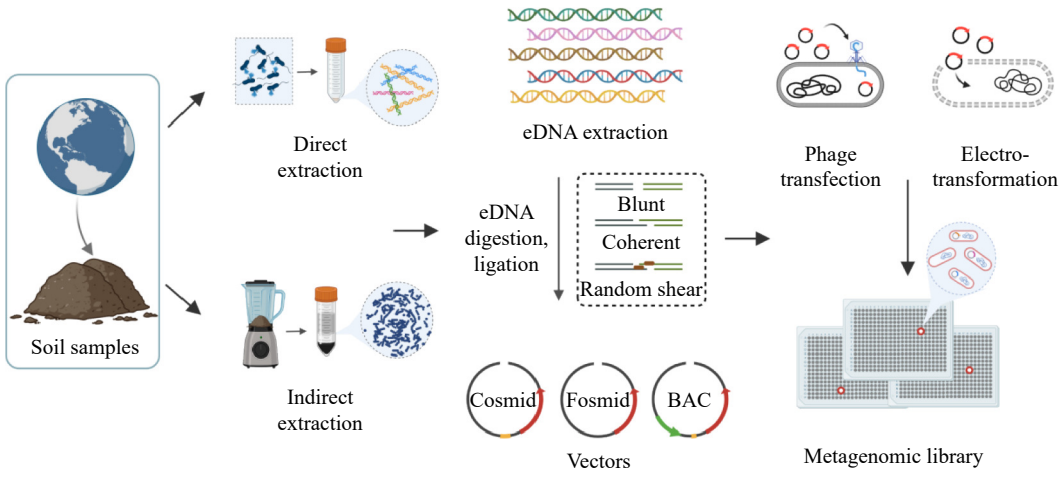

Fig. 1 The workflow for metagenomic library construction. The eDNA samples collected from diverse environmental sources are fragmented, ligated, and transformed into a model host for downstream analysis.

ated recombination (TAR) [26, 27]. The restriction-site ligation strategy relies on the unique restriction sites present in overlapping clones, while λ-based recombination is beneficial for assembling large BGCs [28]. Using the high recombination efficacy of *Saccharomyces cerevisiae*, the yeast-based TAR assembly technology enables the *in vivo* assembly of eDNA-derived overlapping cosmid clones with a pTARa-based vector to form complete BGCs [29]. Among these methods, TAR assembly is particularly common for constructing large pathways involving more than two overlapping clones [26]. However, finding compatible overlapping clones in metagenomic libraries and the complexity of high-throughput assembly of complete BGCs remain significant challenges. To overcome the inherent shortcomings of the cosmid library, BAC vectors, capable of inserting up to 300 kb of foreign DNA, have been utilized in metagenomic library construction [30]. Based on a modified F plasmid, BAC vectors have found widespread use in constructing libraries for single organism genomes, such as those of humans [31], plants [32], and bacteria [33]. A random shear strategy is employed in metagenomic DNA preparation to avoid biases associated with traditional restriction digestion-ligation [34]. Libraries generated using BAC vectors typically have an average insert size of about 100 kb, sufficient to contain a complete natural product BGC within a single clone. However, one drawback of BAC cloning is that the resulting metagenomic libraries are smaller by two to three orders of magnitude compared to those produced using cosmid-based cloning strategies [30].

In the DNA entity sequence-independent method, metagenomic DNA is processed without the need for incorporation into a vector-based library. Instead, this method involves randomly fragmenting the DNA into short sequences, which are then directly sequenced using next-generation sequencing (NGS) technologies (i.e., Illumina [35, 36], PacBio [37], Ion Torrent [38, 39], MGISEQ [39], and Nanopore [40] sequencing). The sequencing reads are either mapped to reference genomes or assembled *de novo* to form extensive contigs. Subsequently, advanced binning algorithms are employed to organize these contigs into coherent genome bins [41-44]. This sequence-independent method differs from the sequence-dependent method in several aspects. Its primary challenges lie in the length and depth of sequencing reads, as well as the overall quality of the assembly. The introduction of PacBio and Nanopore technologies has been pivotal in this context, markedly increasing the length of sequencing reads from a few hundred bases to tens of kilobases [37, 40]. Moreover, the rapid development of binning algorithms is progressively enabling the complete *de novo* assembly of metagenomes from environmental samples [41, 42].

Functional Screening-based Culture-independent Discovery of Metagenomic Natural Products

One of the main purposes of exploring metagenomes is to discover the diversity of natural products encoded by the genetic sequences of eDNA. The success of this endeavor hinges on various factors that influence the activation of natural product gene clusters, such as the completeness of the BGCs recovered from the metagenomic libraries, appropriate selection of heterologous expression hosts, and suitable detection methods [45]. Several culture-independent methods have been developed to explore these untapped resources, with functional screening emerging as the most intuitive method based on phenotypic variation. Functional screening begins with the construction of metagenomic libraries using shuttle vectors such as cosmid/fosmid or BAC. These vectors are employed to capture eDNA sequences, which are then transferred into suitable heterologous expression hosts, such as *E. coli* and *Streptomyces* (Fig. 2). Once the libraries are established, the next phase involves screening for clones that demonstrate a wide range of bioactive natural productions [46]. Based on the characterizations of isolated molecules, functional screening can be divided into colorization-based screening (production of colorful pigments) and inhibition zone-based screening (production of antimicrobial metabolites). In the colorization-based screening method, the production of visible pigments by eDNA-containing colonies serves as a marker for the potential synthesis of previously unknown natural products. A notable example of this method's effectiveness is the discovery of clone CSL51. This clone, containing a four-gene BGC, was found to produce the broad-spectrum antibiotic violacein and its derivative, deoxyviolacein, from soil-derived eDNA libraries [47]. Another instance involved screening purple clones from forest soil-derived libraries, leading to the identification of indirubin and two related operons that produce pigments with antimicrobial activity [48]. This approach has also facilitated the discovery of other bioactive compounds, such as turbomycin B and metatricycloene [7, 49] (Fig. 2).

Inspired by the discovery of penicillin [1], inhibition zone-based functional screening has become a widely used method in the field of metagenomic natural product discovery. This technique involves distributing heterologous expression hosts, each harboring different metagenomic clones, across a soft agar plate. This plate also includes bacteria, such as *Bacillus subtilis*, which serve as indicators for activity testing. The emergence of inhibition zones around certain clones signals the production of antimicrobial natural products by these clones. The isolation and purification of such colonies can lead to the discovery of novel antibiotics. With this approach, CSLG18, an eDNA clone that produced isonitrile functionalized indole antibiotic, was found in a cosmid library using a top agar overlay containing *Bacillus subtilis* [50]. Similarly, Brady *et al*. employed this method to identify clones from eDNA libraries that produced *N*-acyltyrosine [51] and palmitoylputrescine [52].

The advent of advanced sequencing technologies has dramatically reduced the cost of sequencing eDNA, paving the way for more extensive metagenomic studies. However, despite the vast scale of DNA sequences in metagenomic libraries, often extending to terabytes, only a small fraction of

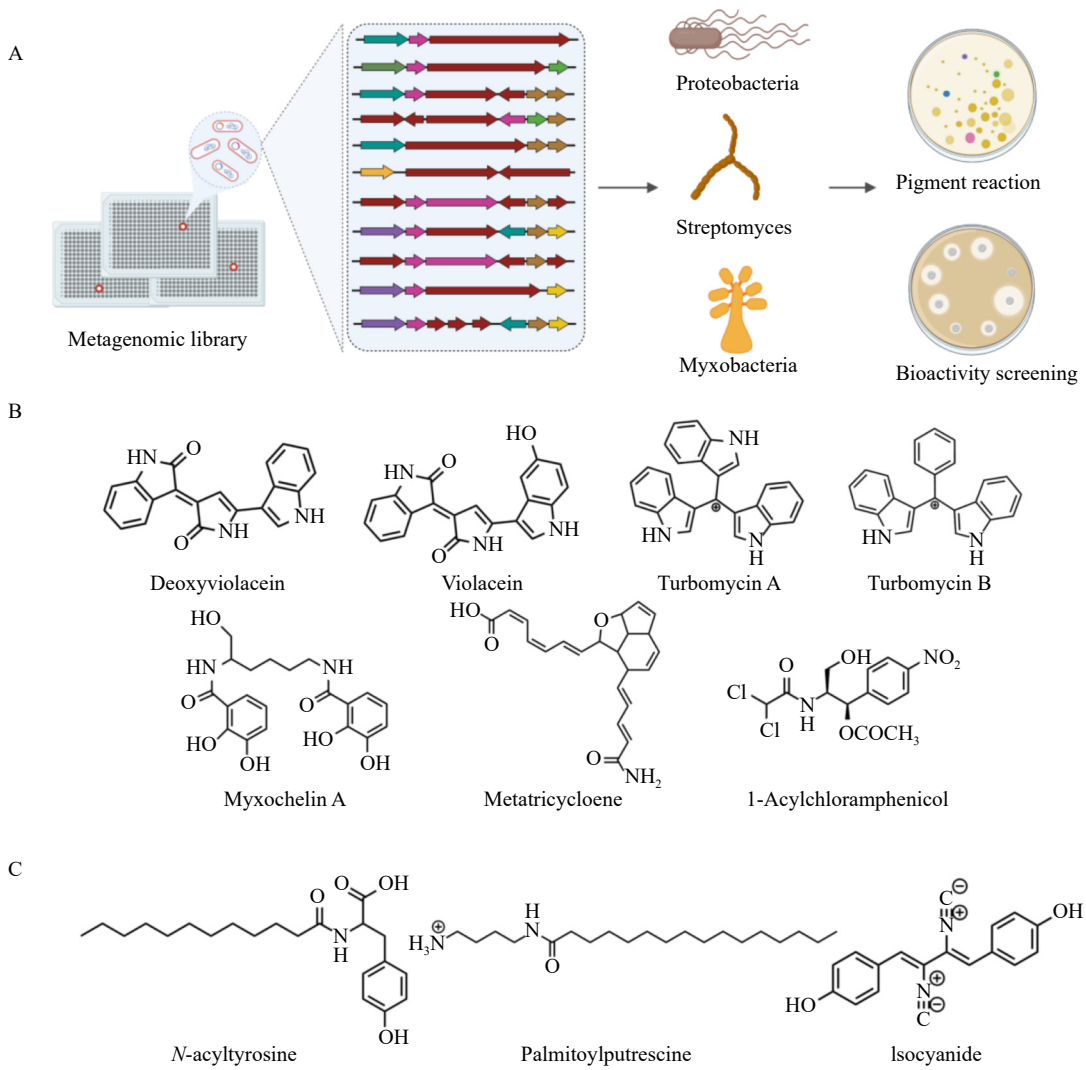

Fig. 2 Functional screening-based culture-independent discovery of metagenomic natural products. (A) General pipeline of functional screening. (B and C) Representative natural product structures obtained by functional screening methods.

the clones encode gene sequences with biosynthetic functions [53]. Key among these are non-ribosomal peptide synthetase (NRPS) and polyketide synthase (PKS), two major classes of BGCs that produce pharmacologically significant microbial secondary metabolites [54, 55]. Therefore, it becomes essential to implement BGC enrichment strategies to increase the probability of identifying eDNA clones containing NRPS or PKS BGCs. One crucial element in natural product biosynthesis is phosphopantetheine transferase (PPTase), a family of enzymes that post-translationally modifies the peptidyl-carrier-protein (PCP) and acyl-carrier-protein (ACP) domains in NRPS and PKS gene clusters [56]. Leveraging this, a PPTase-deficient *Streptomyces albus* strain, combined with an indigodine expression system, was developed to screen eDNA clones for PPTase genes, using color production as an indicator. This innovative approach has led to the discovery of a variety of NRPS, PKS, and NRPS/PKS hybrid BGCs, including the natural product myxochelin A [57]—a dihydrobenzoate-containing siderophore first identified in myxobacteria [58]. For functional screening, metagenomic libraries constructed using cosmid or fosmid vectors are prevalently employed, with comprehensive protocols for eDNA extraction, restriction digestion, and phage infection well-documented [6]. A significant challenge with these libraries is their limited capacity for inserting foreign DNA. Due to the limitation of phage packaging, the cosmid [59] and fosmid [23] libraries can incorporate DNA fragments of less than 40 kb, a size that is often inadequate to cover entire BGCs and, therefore, cannot express corresponding metabolites. Natural product BGCs typically span 60−110 kb, like those of daptomycin (63 kb) [60], erythromycin (61 kb) [61], and rifamycin (109 kb) [62-64]. To overcome this limitation, BAC vector-based metagenomic libraries have been developed to discover bioactive natural products. The average insert size of 120 kb in BAC libraries increases the probability of harboring complete natural product gene clusters. Using soil-derived metagenomic BAC libraries, Rondon *et al*. [65] identified a BAC clone with antibacterial activity against *Bacillus subtilis* BR151 and *Staphyl-*

ococcus aureus. Moreover, to avoid the biases introduced by the restriction digestion of metagenomic DNA, a random shear technology is applied in BAC library construction, ensuring a more uniform fragmentation. Using this technology, Mark *et al*. [34] found chloramphenicol derivatives with activity against methicillin-resistant *Staphylococcus aureus* (MRSA), along with the broad-spectrum antibiotic turbomycin B, complementing the earlier discovery of turbomycin A from a fungus [49].

The key step in functional screening is the selection of a heterologous expression host that accurately transcribes, translates, and biosynthesizes proteins from eDNA sequences. The ideal host for metagenomic libraries should not only demonstrate high diversity in natural product synthesis but also possess robust transformation capabilities for constructing large-insertion metagenomic libraries effectively. Currently, *E. coli* is the most commonly used host for functional metagenomics due to its rapid growth and genetic tractability. However, it is also clear that most BGCs captured from diverse environmental samples cannot be heterologously expressed in a single host, especially those eDNA isolated from different families of bacteria. A variety of alternative hosts have been developed to accommodate a wide range of eDNA-containing BGCs. For example, *Pseudomonas aeruginosa*, a Gram-negative soil bacterium, has shown promise in producing inhibitory proteins from metagenome-derived clones, which led to the discovery of two novel lactonase family proteins with the capability to inhibit biofilm formation [24]. To increase the transcription efficiency of foreign genes, the bifunctional cosmid vector pFX583, which enables T7 RNA polymerase-directed transcription, has been used for the construction and screening of metagenomic libraries, promoting heterologous expression in both *E. coli* and *S. lividans* [66, 67]. The use of diverse proteobacteria as hosts, including *Agrobacterium tumefaciens*, *Burkholderia graminis*, and *Caulobacter vibrioides*, has expanded the biosynthetic diversity of eDNA clones obtained from functional screening [68]. In addition, eukaryotes such as *Saccharomyces cerevisiae* have been used to optimize the expression of exogenous DNA fragments [69].

Apart from proteobacteria and eukaryotes, actinomyces have been recognized as prolific hosts of therapeutically relevant metabolites [70]. Previous attempts at functional screening in streptomyces have been made using *Streptomyces lividans* as the host [71, 72]. To enhance the discovery of novel secondary metabolites in *Streptomyces*, Iqbal *et al*. conducted a host screening of 38 species to identify species with the best innate ability to heterologously express BGCs [7]. Among the screened species, *S. albus* was identified as the most efficient host, facilitating the production of several novel natural products, including metatricycloene [7], nocardamine [73], myxochelin A [57], malacidins [29], and cadasides [8].

Sequence Tag-based Culture-independent Discovery of Natural Products from Metagenomes

Functional screening, in combination with metagenomic libraries, has produced numerous bioactive natural products and functional enzymes. This approach is especially valuable for identifying "unknown-unknown" molecules, which are characterized by novel chemical scaffolds and unique biochemical properties and are not reliant on known gene sequences or molecule structures. However, the current technologies for metagenomic library construction using cosmid/fosmid and BAC vectors encounter limitations in meeting the insert size requirements necessary to express complete BGCs encoding natural products [6]. Although the insert size is large enough for screening bioactive enzymes and some short pathways, it hardly covers the full-length natural product BGCs within a single vector. This shortfall impedes the production of related metabolites in functional screening and contributes to the dormancy of many natural product BGCs.

An alternative approach named sequence tag-based natural product discovery from metagenomes has been developed to uncover the metabolite diversity in nature (Fig. 3). This method uses polymerase chain reaction (PCR) to screen eDNA libraries with degenerate primers that target the conserved regions of BGCs, enabling the targeted discovery of specific types of natural products [74]. Unlike functional screening, sequence tag-based screening does not rely on the functionality of inserted foreign DNA sequences. Instead, it focuses on identifying short sequences or sequence tags that target the desired gene clusters. The gene sequence encoding special module in BGCs can be used as a searching tag. These degenerate primers amplify the conserved regions of the target sequences and screen against the entire metagenomic DNA library to find evolutionally related amplicons. The resulting amplicon sequences are then analyzed for similarities and evolutionary relationships with known reference sequences. Next, the complete BGCs related to the amplicon sequence are recovered from corresponding metagenomic libraries. The recovered full BGCs are then transformed into an appropriate host for expression. It is worth noting that the sequence tag itself is not expressive, so it needs to be combined with a heterologous expression system or a synthetic-bioinformatics natural product (syn-BNP) approach. This combination is essential to decode the genetic information and facilitate the biological or chemical synthesis of the final molecules.

Microbial natural products, such as non-ribosomal peptides (NRPS) and polyketides (PKS), are primary targets for metagenomic research due to their significant medicinal potential. These secondary metabolites are synthesized through pathways involving specific biosynthetic domains, notably the adenylation (A) and ketosynthase (KS) domains. Researchers utilize degenerate primers targeting the conserved regions of these domains to screen metagenomic libraries for novel NRPS and PKS biosynthetic gene clusters (BGCs) related to reference sequences. Malacidins [29] and cadasides [8], two new calcium-dependent antibiotics, were identified from soil metagenomic libraries using an A probe targeting the Asp4 motif of NRPS gene clusters. Both compounds showed

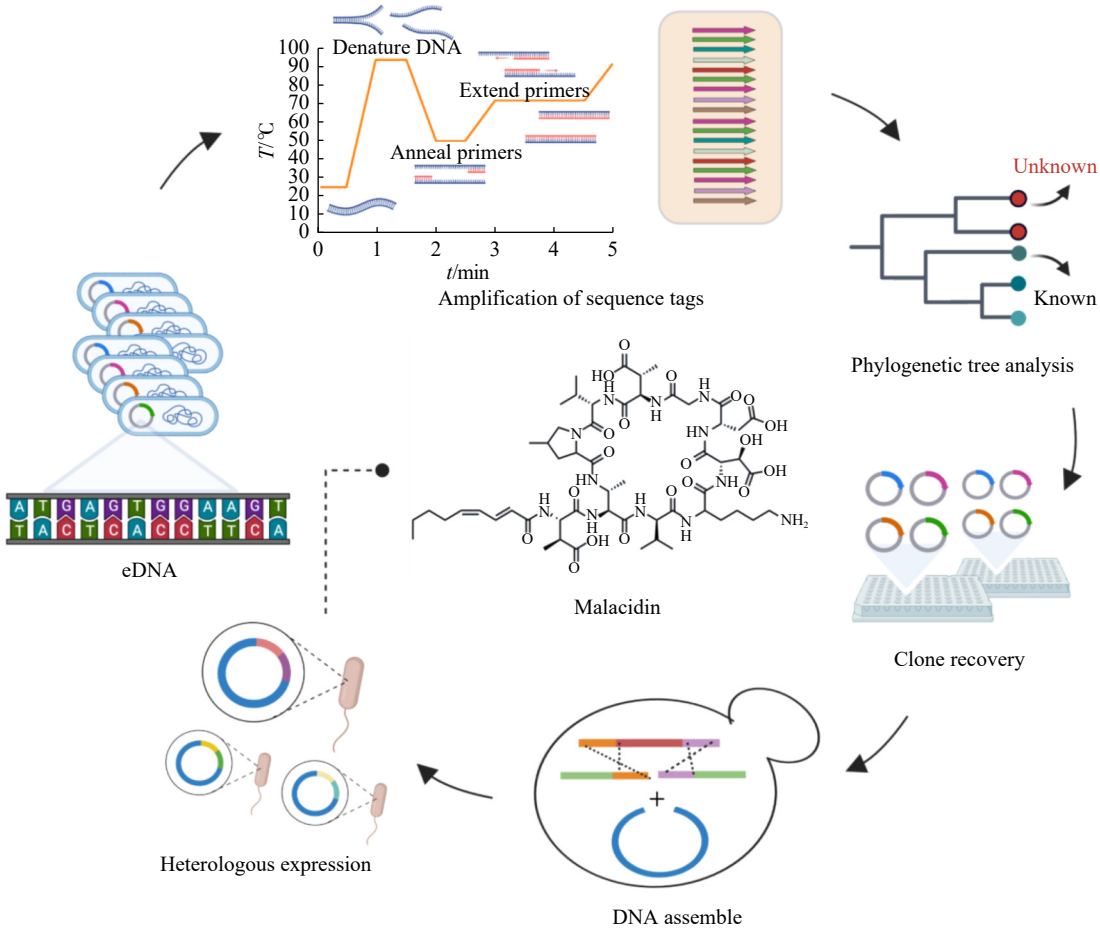

Fig. 3 Sequence tag-based culture-independent discovery of metagenomic natural products. The eDNA samples are immobilized in metagenomic libraries.

potent antimicrobial activity *via* a calcium-dependent mechanism, yet they exhibit variations in their molecular structures and specific activities. Moreover, molecules containing para-aromatic benzoic acid (PABA) are of significant interest due to their backbone comprising multiple non-proteogenic amino acids, their capacity for topoisomerase inhibition, and their potent bioactivity. Utilizing A probes to target PABA-encoding amplicons in soil-derived metagenomic libraries has led to the recovery of several novel PABA-related BGCs. Subsequently, this approach facilitated the synthesis of various naturally inspired antibiotics and antiproliferative compounds, including PABA34, PABA48, and lapcin [9, 75]. Additionally, a sequence tag targeting KSα or KSβ was designed to extract BGCs of polyketide compounds from soil eDNA and led to the discovery of arixanthomycins A−C [76], calixanthomycin A [77], and arenimycins C and D [77] as well as clinically relevant arimetamycin A [78] and antibiotic fasamycins A and B [79].

To investigate the diversity of bacterial secondary metabolites in metagenomic libraries sourced from various soil environments, researchers have adopted probes targeting both the AD domain of NRPS and the KS domain of PKS. This method was particularly effective in analyzing a collection of marine sediment-derived actinobacteria, and it enabled a bioinformatic assessment of the biosynthetic potential of secondary metabolites, even in the absence of fully assembled pathways or complete genome sequences, helping identify strains that possess the highest potential to produce similar or new secondary metabolites. Intriguingly, this study uncovered that metagenomes from ecologically similar environments, even those with comparable phylum-level 16S rRNA gene distributions, can possess almost completely distinct sets of biosynthetic gene sequences [80]. In summary, utilizing sequence tags facilitates an in-depth analysis of the diversity of secondary metabolites without the need to express the molecules involved. In addition to targeting the biosynthetic domains of NRPS/PKS, focusing on tailoring enzymes implicated in post-modification processes can lead to the discovery of natural products with distinct bioactivities and structures. Isonitrile A was the first characterized isonitrile enzyme that catalyzed free primary amine in tryptophan to an isonitrile group. Researchers designed degenerate primers targeting the isonitrile synthase (isnA) gene to scan both fully sequenced bacterial genomes and eDNA, resulting in the identification of 12 eDNA clones containing operons homologous to isnA. The heterologous expression of these operons in *E. coli* and

Pseudomonas aeruginosa led to the discovery of several natural products encoded by isnA-contained operons [81]. To expand the genetic source of self-sufficient cytochrome P450 monooxygenases (CYPs), researchers identified the first CYP gene, *SYK181*, with significant hydroxylase activity from a soil-derived metagenomic library based on CYP sequence homology screening [82]. Furthermore, the investigation into the diversity of microviridin gene clusters, directed by homologous sequences of the ATP grasp-type ligase MdnB, enabled the discovery of novel tricyclic depsipeptides [83].

Tryptophan dimers (TDs) are a group of structurally and pharmacologically diverse natural products that have garnered significant interest [84]. Their unique substructures make them compelling targets for sequence-guided screening in the quest for bioactive natural products. Employing soil-derived metagenomic libraries and homology-based screening of targeted TD BGCs, researchers have successfully identified a variety of natural products containing TD structures, each exhibiting a range of biological activities [85, 86]. A key strategy in isolating unique TD BGCs within expansive soil metagenomic libraries involves using the oxy-tryptophan dimerization gene as a sequence tag. This approach has facilitated the discovery of several bioactive compounds, such as BE-54017, a substance with antitumor properties [87], borregomycin A, an antiproliferative agent, and a series of compounds including borregomycins B−D, which have shown potential as anticancer and antibiotic agents [88]. In addition, researchers have explored the chromopyrrolic acid (CPA) synthase gene as a navigational tool for uncovering novel TDs from environmental samples. By probing eDNA libraries for CPA synthase genes and constructing phylogenetic trees, they have been able to pinpoint and express specific gene clusters heterologously in *E. coli*, which has led to the characterization of erdasporines, a group of cytotoxins with a novel carboxy-indolocarbazole TD substructure and hydroxysporine/reductasporine with antifungal activity [85].

As a significant challenge in metagenomic natural product discovery, the size limitations of cosmid/fosmid and even BAC vectors are noteworthy. These vectors often fall short in covering the entire length of BGCs, particularly when used to recover clones containing targeted probe genes. After a target synthetase or synthase gene has been identified and amplified from eDNA libraries using homologous sequences, the subsequent step involves the reassembly of overlapping clones to reconstruct a complete BGC. Several recombination techniques, both *in vivo* and *in vitro*, have been developed to address this challenge, such as TAR and Red/ET, which have been reported thoroughly in previous researches [26, 28, 89]. However, despite the efficiency of sequence tag-based approaches in identifying targeted natural product gene clusters, these methods are not without limitations. Significant challenges include the inherent bias associated with degenerate PCR and the "silence" of recovered BGCs [90, 91].

Natural Product Structure Prediction and Total Chemical Synthesis

The advancements in next-generation sequencing (NGS) technologies have significantly revolutionized bacterial genomic sequencing [10, 92, 93]. Although numerous BGCs have been identified through these technologies, the secondary metabolites they produce often remain elusive. Traditionally, the biosynthesis of natural products has been the primary method for unraveling the genetic codes of these BGCs and producing the corresponding metabolites. However, this approach faces significant challenges: most bacteria cannot be cultured under laboratory conditions, and many gene clusters remain silent and unexpressed during fermentation processes, hindering the discovery of potentially valuable natural products. To bypass these obstacles, researchers have combined bioinformatics prediction with total chemical synthesis methods. This innovative approach, known as synthetic-bioinformatics natural product (syn-BNP) discovery, enables the acquisition of natural product small molecules without the need for culturing bacteria or heterologously expressing BGCs (Fig. 4) [94, 95]. The process involves obtaining the primary DNA sequences of relevant BGCs from bacterial genomes or assembled metagenomic sequences and then analyzing the encoded chemical structures using bioinformatics algorithms. The syn-BNP method is particularly adept at predicting BGCs related to NRPS. NRPS-derived secondary metabolites are synthesized through repetitive modules, allowing for precise bioinformatic predictions. Various algorithms have been developed to predict the substrate specificity of the A domains in NRPS [96-100]. The critical aspect of these predictions hinges on the ten essential binding-pocket residues (positions 235, 236, 239, 278, 299, 301, 322, 330, 331, and 517) in the active center of the A domain [96]. However, it has been observed that expanding the number of binding residues or using the entire A domain sequence does not enhance prediction efficiency. This approach has been found to decrease the precision of predictions [100]. To improve the accuracy of predictions for A domains that lack exact matches in the training data, an evolutionary-based phylogenetic algorithm known as SANDPUMA has been incorporated into these prediction models [100].

The discovery of humimycin, achieved through syn-BNP technology, represents a significant advancement in drug discovery. Humimycin is a synthesized lipopeptide antibiotic notable for enhancing the efficacy of *β*-lactam antibiotics against MRSA, specifically targeting lipid II flippase, a vital component in bacterial cell wall synthesis. Brady and his team used bioinformatic tools to analyze the human microbiome pool. They successfully identified 25 complete NRPS gene clusters, each encoding peptides composed of more than five amino acids. The team then deduced the amino acid structures and sequences encoded by each A domain *in silico*, employing established bioinformatic algorithms, such as

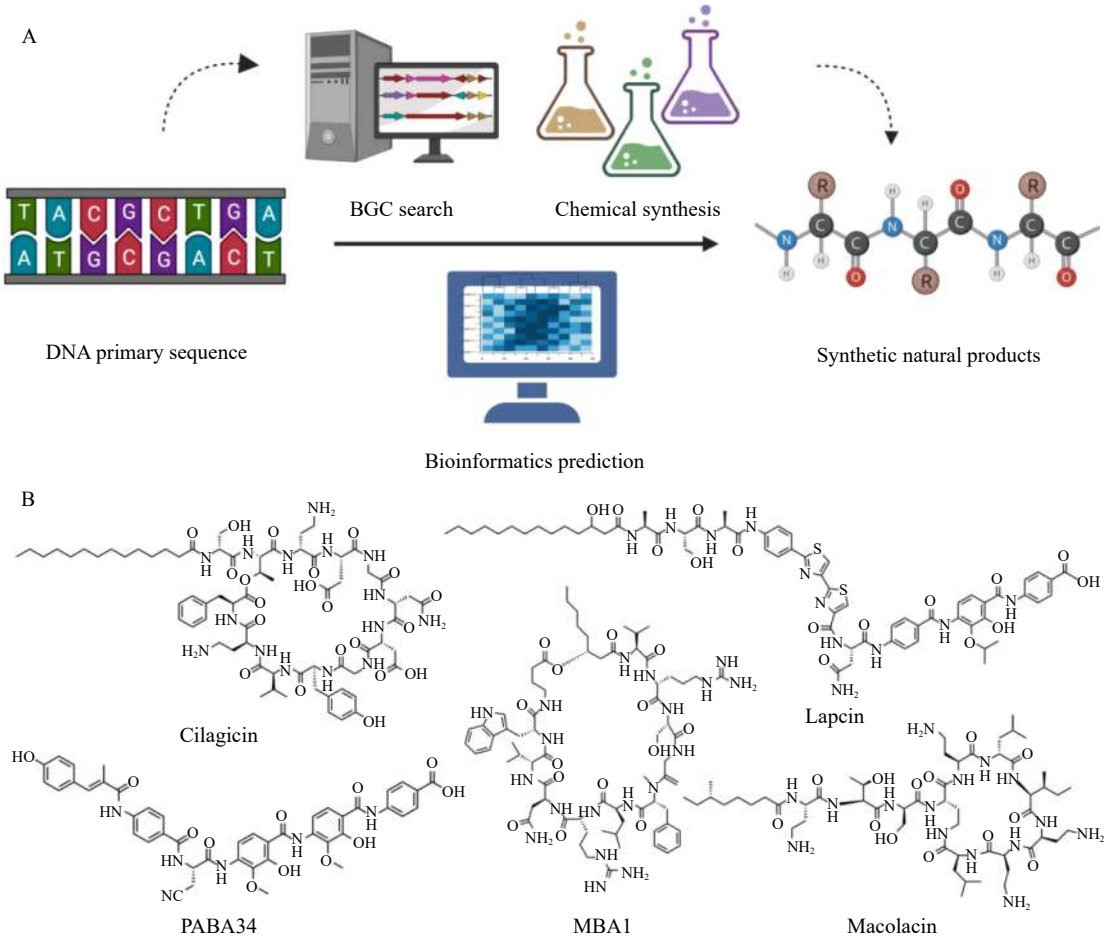

Fig. 4 Discovery of inspired natural products using natural product structure prediction and total chemical synthesis. (A) Pipeline of synthetic-bioinformatics natural product (syn-BNP) approach. (B) Representative structures of identified natural products using the syn-BNP approach.

Stachelhaus, Minowa, and NRPSPredictor2, now integrated into the antiSMASH software. The predicted structures were then synthesized using solid-phase peptide synthesis (SPPS) technology [96, 97, 101, 102]. The resulting molecules were tested against a panel of common human commensal and pathogenic bacteria, leading to the discovery of antibiotic compounds [94]. Expanding upon this method, Vila-Farres *et al*. then screened syn-BNPs against diverse microbial pathogens and discovered two novel antibiotics, syn-BNP1 and syn-BNP2, which showed antibacterial and antifungal activity, respectively [103]. Although the specificity of the A domain in NRPS can be predicted using bioinformatic algorithms, the cyclization site of each NRP cannot be predicted at the genetic level. To figure out the correct cyclization pattern, the researchers synthesized various cyclic peptides using three common cyclization methods: head-to-tail (cyclizing the terminal carboxylic acid to the free amine of the first amino acid), head-to-lipid chain (cyclizing the terminal carboxylic acid to a hydroxyl group on a lipid chain), and head-to-middle (cyclizing the terminal carboxylic acid to an amine/hydroxyl group of middle amino acids in the peptide chain). The most bioactive variant was then determined through screening for biological activity.

In addition to obtaining NRPS BGCs through a cultured bacterial genome database, biosynthetic instructions are used to guide the synthesis of natural products involved in metagenome-derived BGCs. One successful example of this approach is the targeting of PABA-specific A domains to extract PABA-containing BGCs from complex eDNA libraries. This strategy, coupled with bioinformatics prediction and total chemical synthesis, led to the discovery of several bioactive PABA-containing molecules. Among these, PABA34, PABA48 [75], and lapcin [9] stand out, with lapcin showing potent cytotoxicity against a range of cancer cells and a unique mechanism of action through dual topoisomerase I/II enzyme binding. The ability to decode genetic information from sequenced metagenomes has allowed researchers to uncover the metabolite diversity hidden under previous unknown gene cluster pools. In a related study, Wang *et al*. [104] utilized syn-BNP methodology to analyze the functionality of previously unknown colistin-related BGCs and found a natural way to overcome antibiotic resistance, a growing concern in clinical settings. By analyzing the evolutionary relationships of colistin-related BGCs in sequenced bacterial genomes, they

identified a novel colistin BGC predicted to encode a distinctive colistin structure. Employing syn-BNP, they synthesized macolacin, a novel nature-inspired antibiotic against extensive drug-resistant *Acinetobacter baumannii* with colistin resistance. Building on this evolutionary idea, Wang *et al.* used the conserved condensation starter (Cs) domain as a common sequence tag to analyze the evolutionary relationships among known lipopeptide antibiotics. This exploration led to the identification and synthesis of cilagicin, a naturally inspired lipopeptide antibiotic demonstrating potent activity against Gram-positive pathogens. Its mechanism, targeting crucial undecaprenyl phosphates in cell wall biosynthesis, combined with an absence of detectable resistance in laboratory tests, positions cilagicin as a promising candidate in combating antibiotic-resistant pathogens [105]. Additionally, menaquinone (MK) has been identified as an attractive target for addressing infections caused by multidrug-resistant bacteria [106]. Using the conserved minimal MK binding motif GXLXXXW detected in structurally diverse MK-binding antibiotics (MBAs), researchers have identified potential BGCs predicted to encode new MBAs with *in vivo* activities against multidrug-resistant pathogens [107].

Advances in bioinformatics, particularly in BGC analysis, combined with growing expertise in chemical synthesis, have enabled scientists to more accurately deduce and synthesize the structures of natural products encoded by BGCs identified within metagenomic sequencing pools. While metabolites synthesized *via* the syn-BNP method may not be exact replicas of naturally biosynthesized compounds, they often closely approximate the structures and biological activities of these natural products. The identification of bioactive metabolites using the syn-BNP approach has opened new avenues for the discovery of novel, pharmaceutically relevant drug candidates without relying on natural product biosynthesis. However, despite its promise and achievements, the syn-BNP approach faces challenges that need to be addressed to fully realize its potential. Enhancing the accuracy of prediction algorithms is one of the primary areas for improvement. Additionally, the scope of chemical synthesis also presents limitations.

Conclusion

Secondary metabolites produced by uncultured microorganisms represent a new, vast, and diverse reservoir of therapeutic microbial drugs. The development of NGS technology, metagenomics, and syn-BNP approach has revolutionized the discovery of microbial drugs derived from uncultured environmental microbes. By applying these culture-independent approaches, the biosynthetic information hidden in those untapped microorganisms can be captured by either metagenomic libraries or eDNA sequencing. The acquired BGCs are then decoded using heterologous expression and total chemical synthesis. However, the field of metagenomics-based natural product discovery still faces several challenges. A primary issue is the limited capacity of current metagenomic libraries to cover the full length of complex BGCs. Additionally, successfully identifying BGCs does not always translate to the production of the corresponding natural products, especially considering the presence of silent or cryptic gene clusters. Therefore, the development of metagenomic libraries capable of larger insertions, innovative methods for activating BGCs, enhanced algorithms for structure prediction, and broader breakthroughs in molecular biology techniques are all expected to address the current bottlenecks in metagenomics technology. Looking forward, metagenomics, enriched by multidisciplinary collaboration, is poised to become a mainstay in microbial natural product research.

References

[1] Kardos N, Demain AL. Penicillin: the medicine with the greatest impact on therapeutic outcomes [J]. *Appl Microbiol Biotechnol*, 2011, **92**(4): 677-687.
[2] Xu F, Wu Y, Zhang C, *et al*. A genetics-free method for high-throughput discovery of cryptic microbial metabolites [J]. *Nat Chem Biol*, 2019, **15**(2): 161-168.
[3] Lewis K. Platforms for antibiotic discovery [J]. *Nat Rev Drug Discov*, 2013, **12**(5): 371-387.
[4] Gutleben J, Chaib DM, van-Elsas JD, *et al*. The multi-omics promise in context: from sequence to microbial isolate [J]. *Crit Rev Microbiol*, 2018, **44**(2): 212-229.
[5] Ling LL, Schneider T, Peoples AJ, *et al*. A new antibiotic kills pathogens without detectable resistance [J]. *Nature*, 2015, **517**(7535): 455-459.
[6] Brady SF. Construction of soil environmental DNA cosmid libraries and screening for clones that produce biologically active small molecules [J]. *Nat Protoc*, 2007, **2**(5): 1297-1305.
[7] Iqbal HA, Low-Beinart L, Obiajulu JU, *et al*. Natural product discovery through improved functional metagenomics in streptomyces [J]. *J Am Chem Soc*, 2016, **138**(30): 9341-9344.
[8] Wu C, Shang Z, Lemetre C, *et al*. Cadasides, calcium-dependent acidic lipopeptides from the soil metagenome that are active against multidrug-resistant bacteria [J]. *J Am Chem Soc*, 2019, **141**(9): 3910-3919.
[9] Wang Z, Forelli N, Hernandez Y, *et al*. Lapcin, a potent dual topoisomerase I/II inhibitor discovered by soil metagenome guided total chemical synthesis [J]. *Nat Commun*, 2022, **13**(1): 842.
[10] Kalkreuter E, Pan G, Cepeda AJ, *et al*. Targeting bacterial genomes for natural product discovery [J]. *Trends Pharmacol Sci*, 2020, **41**(1): 13-26.
[11] Newman DJ, Cragg GM. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019 [J]. *J Nat Prod*, 2020, **83**(3): 770-803.
[12] Zhang H, Tang X. Combining microbial and chemical syntheses for the production of complex natural products [J]. *Chin J Nat Med*, 2022, **20**(10): 729-736.
[13] Hugenholtz P, Goebel BM, Pace NR. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity [J]. *J Bacteriol*, 1998, **180**(18): 4765-4774.
[14] Rappe MS, Giovannoni SJ. The uncultured microbial majority [J]. *Annu Rev Microbiol*, 2003, **57**: 369-394.
[15] Torsvik V, Goksoyr J, Daae FL. High diversity in DNA of soil bacteria [J]. *Appl Environ Microbiol*, 1990, **56**(3): 782-787.
[16] Torsvik V, Daae FL, Sandaa RA, *et al*. Novel techniques for analysing microbial diversity in natural and perturbed environments [J]. *J Biotechnol*, 1998, **64**(1): 53-62.
[17] Kallifidas D, Kang HS, Brady SF. Tetarimycin A, an MRSA-active antibiotic identified through induced expression of en-

vironmental DNA gene clusters [J]. *J Am Chem Soc*, 2012, **134**(48): 19552-19555.
[18] Handelsman J, Rondon MR, Brady SF, *et al*. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products [J]. *Chem Biol*, 1998, **5**(10): R245-249.
[19] Delmont TO, Robe P, Clark I, *et al*. Metagenomic comparison of direct and indirect soil DNA extraction approaches [J]. *J Microbiol Methods*, 2011, **86**(3): 397-400.
[20] Echeverria-Beirute F, Varela-Benavides I, Jiménez-Madrigal JP, *et al*. eDNA extraction protocol for metagenomic studies in tropical soils [J]. *Biotechniques*, 2021, **71**(6): 580-586.
[21] Liu J, Li J, Feng L, *et al*. An improved method for extracting bacteria from soil for high molecular weight DNA recovery and BAC library construction [J]. *J Microbiol*, 2010, **48**(6): 728-733.
[22] Aakvik T, Degnes KF, Dahlsrud R, *et al*. A plasmid RK2-based broad-host-range cloning vector useful for transfer of metagenomic libraries to a variety of bacterial species [J]. *FEMS Microbiol Lett*, 2009, **296**(2): 149-158.
[23] Kim UJ, Shizuya H, de-Jong PJ, *et al*. Stable propagation of cosmid sized human DNA inserts in an F factor based vector [J]. *Nucleic Acids Res*, 1992, **20**(5): 1083-1085.
[24] Schipper C, Hornung C, Bijtenhoorn P, *et al*. Metagenome-derived clones encoding two novel lactonase family proteins involved in biofilm inhibition in *Pseudomonas aeruginosa* [J]. *Appl Environ Microbiol*, 2009, **75**(1): 224-233.
[25] Feng Z, Kallifidas D, Brady SF. Functional analysis of environmental DNA-derived type II polyketide synthases reveals structurally diverse secondary metabolites [J]. *Proc Natl Acad Sci U S A*, 2011, **108**(31): 12629-12634.
[26] Kim JH, Feng Z, Bauer JD, *et al*. Cloning large natural product gene clusters from the environment: piecing environmental DNA gene clusters back together with TAR [J]. *Biopolymers*, 2010, **93**(9): 833-844.
[27] Cheng J, Pinnell L, Engel K, *et al*. Versatile broad-host-range cosmids for construction of high quality metagenomic libraries [J]. *J Microbiol Methods*, 2014, **99**: 27-34.
[28] Abbasi MN, Fu J, Bian X, *et al*. Recombineering for genetic engineering of natural product biosynthetic pathways [J]. *Trends Biotechnol*, 2020, **38**(7): 715-728.
[29] Hover BM, Kim SH, Katz M, *et al*. Culture-independent discovery of the malacidins as calcium-dependent antibiotics with activity against multidrug-resistant Gram-positive pathogens [J]. *Nat Microbiol*, 2018, **3**(4): 415-422.
[30] Liles MR, Williamson LL, Rodbumrer J, *et al*. Recovery, purification, and cloning of high-molecular-weight DNA from soil microorganisms [J]. *Appl Environ Microbiol*, 2008, **74**(10): 3302-3305.
[31] Kim UJ, Birren BW, Slepak T, *et al*. Construction and characterization of a human bacterial artificial chromosome library [J]. *Genomics*, 1996, **34**(2): 213-218.
[32] Lin J, Kudrna D, Wing RA. Construction, characterization, and preliminary BAC-end sequence analysis of a bacterial artificial chromosome library of the tea plant (*Camellia sinensis*) [J]. *J Biomed Biotechnol*, 2011, **2011**: 476723.
[33] Ouyang Y, Dai S, Xie L, *et al*. Isolation of high molecular weight DNA from marine sponge bacteria for BAC library construction [J]. *Mar Biotechnol (NY)*, 2010, **12**(3): 318-325.
[34] Nasrin S, Ganji S, Kakirde KS, *et al*. Chloramphenicol derivatives with antibacterial activity identified by functional metagenomics [J]. *J Nat Prod*, 2018, **81**(6): 1321-1332.
[35] Pendleton M, Sebra R, Pang AW, *et al*. Assembly and diploid architecture of an individual human genome *via* single-molecule technologies [J]. *Nat Methods*, 2015, **12**(8): 780-786.
[36] Zheng GX, Lau BT, Schnall-Levin M, *et al*. Haplotyping germline and cancer genomes with high-throughput linked-read sequencing [J]. *Nat Biotechnol*, 2016, **34**(3): 303-311.
[37] Fu S, Wang A, Au KF. A comparative evaluation of hybrid error correction methods for error-prone long reads [J]. *Genome Biol*, 2019, **20**(1): 26.
[38] Whiteley AS, Jenkins S, Waite I, *et al*. Microbial 16S rRNA Ion Tag and community metagenome sequencing using the Ion Torrent (PGM) Platform [J]. *J Microbiol Methods*, 2012, **91**(1): 80-88.
[39] Park C, Kim SB, Choi SH, *et al*. Comparison of 16S rRNA gene based microbial profiling using five next-generation sequencers and various primers [J]. *Front Microbiol*, 2021, **12**: 715500.
[40] Payne A, Holmes N, Rakyan V, *et al*. BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files [J]. *Bioinformatics*, 2019, **35**(13): 2193-2198.
[41] Wu YW, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets [J]. *Bioinformatics*, 2016, **32**(4): 605-607.
[42] Kang DD, Froula J, Egan R, *et al*. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities [J]. *Peer J*, 2015, **3**: e1165.
[43] Lu YY, Chen T, Fuhrman JA, *et al*. COCACOLA: binning metagenomic contigs using sequence COmposition, read CoverAge, CO-alignment and paired-end read LinkAge [J]. *Bioinformatics*, 2017, **33**(6): 791-798.
[44] Alneberg J, Bjarnason BS, de-Bruijn I, *et al*. Binning metagenomic contigs by coverage and composition [J]. *Nat Methods*, 2014, **11**(11): 1144-1146.
[45] Katz M, Hover BM, Brady SF. Culture-independent discovery of natural products from soil metagenomes [J]. *J Ind Microbiol Biotechnol*, 2016, **43**(2-3): 129-141.
[46] Handelsman J. Metagenomics: application of genomics to uncultured microorganisms [J]. *Microbiol Mol Biol Rev*, 2004, **68**(4): 669-685.
[47] Brady SF, Chao CJ, Handelsman J, *et al*. Cloning and heterologous expression of a natural product biosynthetic gene cluster from eDNA [J]. *Org Lett*, 2001, **3**(13): 1981-1984.
[48] Lim HK, Chung EJ, Kim JC, *et al*. Characterization of a forest soil metagenome clone that confers indirubin and indigo production on *Escherichia coli* [J]. *Appl Environ Microbiol*, 2005, **71**(12): 7768-7777.
[49] Gillespie DE, Brady SF, Bettermann AD, *et al*. Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA [J]. *Appl Environ Microbiol*, 2002, **68**(9): 4301-4306.
[50] Brady SF, Clardy J. Cloning and heterologous expression of isocyanide biosynthetic genes from environmental DNA [J]. *Angew Chem Int Ed Engl*, 2005, **44**(43): 7063-7065.
[51] Brady SF, Chao CJ, Clardy J. Long-chain *N*-acyltyrosine synthases from environmental DNA [J]. *Appl Environ Microbiol*, 2004, **70**(11): 6865-6870.
[52] Brady SF, Clardy J. Palmitoylputrescine, an antibiotic isolated from the heterologous expression of DNA extracted from bromeliad tank water [J]. *J Nat Prod*, 2004, **67**(8): 1283-1286.
[53] Garcia JAL, Fernández-Guerra A, Casamayor EO. A close relationship between primary nucleotides sequence structure and the composition of functional genes in the genome of prokaryotes [J]. *Mol Phylogenet Evol*, 2011, **61**(3): 650-658.
[54] Sieber SA, Marahiel MA. Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics [J]. *Chem Rev*, 2005, **105**(2): 715-738.
[55] Stevenson LJ, Owen JG, Ackerley DF. Metagenome driven

discovery of nonribosomal peptides [J]. *ACS Chem Biol*, 2019, **14**(10): 2115-2126.
[56] Beld J, Sonnenschein EC, Vickery CR, *et al*. The phosphopantetheinyl transferases: catalysis of a post-translational modification crucial for life [J]. *Nat Prod Rep*, 2014, **31**(1): 61-108.
[57] Bitok JK, Lemetre C, Ternei MA, *et al*. Identification of biosynthetic gene clusters from metagenomic libraries using PPTase complementation in a *Streptomyces* host [J]. *FEMS Microbiol Lett*, 2017, **364**(16): fnx155.
[58] Kunze B, Bedorf N, Kohl W, *et al*. Myxochelin A, a new iron-chelating compound from *Angiococcus disciformis* (myxobacterales). Production, isolation, physico-chemical and biological properties [J]. *J Antibiot (Tokyo)*, 1989, **42**(1): 14-17.
[59] Collins J, Hohn B. Cosmids: a type of plasmid gene-cloning vector that is packageable *in vitro* in bacteriophage lambda heads [J]. *Proc Natl Acad Sci U S A*, 1978, **75**(9): 4242-4246.
[60] Miao V, Coeffet-LeGal MF, Brian P, *et al*. Daptomycin biosynthesis in *Streptomyces roseosporus*: cloning and analysis of the gene cluster and revision of peptide stereochemistry [J]. *Microbiology (Reading)*, 2005, **151**(Pt 5): 1507-1523.
[61] Brikun IA, Reeves AR, Cernota WH, *et al*. The erythromycin biosynthetic gene cluster of *Aeromicrobium erythreum* [J]. *J Ind Microbiol Biotechnol*, 2004, **31**(7): 335-344.
[62] August PR, Tang L, Yoon YJ, *et al*. Biosynthesis of the ansamycin antibiotic rifamycin: deductions from the molecular analysis of the rif biosynthetic gene cluster of *Amycolatopsis mediterranei* S699 [J]. *Chem Biol*, 1998, **5**(2): 69-79.
[63] Kim CG, Yu TW, Fryhle CB, *et al*. 3-Amino-5-hydroxybenzoic acid synthase, the terminal enzyme in the formation of the precursor of mC7N units in rifamycin and related antibiotics [J]. *J Biol Chem*, 1998, **273**(11): 6030-6040.
[64] Yu TW, Muller R, Muller M, *et al*. Mutational analysis and reconstituted expression of the biosynthetic genes involved in the formation of 3-amino-5-hydroxybenzoic acid, the starter unit of rifamycin biosynthesis in *Amycolatopsis mediterranei* S699 [J]. *J Biol Chem*, 2001, **276**(16): 12546-12555.
[65] Rondon MR, August PR, Bettermann AD, *et al*. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms [J]. *Appl Environ Microbiol*, 2000, **66**(6): 2541-2547.
[66] Lussier FX, Chambenoit O, Cote A, *et al*. Construction and functional screening of a metagenomic library using a T7 RNA polymerase-based expression cosmid vector [J]. *J Ind Microbiol Biotechnol*, 2011, **38**(9): 1321-1328.
[67] Lussier FX, Denis F, Shareck F. Adaptation of the highly productive T7 expression system to *Streptomyces lividans* [J]. *Appl Environ Microbiol*, 2010, **76**(3): 967-970.
[68] Craig JW, Chang FY, Kim JH, *et al*. Expanding small-molecule functional metagenomics through parallel screening of broad-host-range cosmid environmental DNA libraries in diverse proteobacteria [J]. *Appl Environ Microbiol*, 2010, **76**(5): 1633-1641.
[69] Kellner H, Luis P, Portetelle D, *et al*. Screening of a soil metatranscriptomic library by functional complementation of *Saccharomyces cerevisiae* mutants [J]. *Microbiol Res*, 2011, **166**(5): 360-368.
[70] Procopio RE, Silva IR, Martins MK, *et al*. Antibiotics produced by Streptomyces [J]. *Braz J Infect Dis*, 2012, **16**(5): 466-471.
[71] Gomez-Escribano JP, Bibb MJ. Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters [J]. *Microb Biotechnol*, 2011, **4**(2): 207-215.
[72] Komatsu M, Uchiyama T, Omura S, *et al*. Genome-minimized *Streptomyces* host for the heterologous expression of secondary metabolism [J]. *Proc Natl Acad Sci U S A*, 2010, **107**(6): 2646-2651.
[73] Wang GY, Graziani E, Waters B, *et al*. Novel natural products from soil DNA libraries in a *Streptomycete* host [J]. *Org Lett*, 2000, **2**(16): 2401-2404.
[74] Brady SF, Simmons L, Kim JH, *et al*. Metagenomic approaches to natural products from free-living and symbiotic organisms [J]. *Nat Prod Rep*, 2009, **26**(11): 1488-1503.
[75] Wang Z, Kasper A, Mehmood R, *et al*. Metagenome-guided analogue synthesis yields improved Gram-negative-active albicidin- and cystobactamid-type antibiotics [J]. *Angew Chem Int Ed Engl*, 2021, **60**(41): 22172-22177.
[76] Kang HS, Brady SF. Arixanthomycins A−C: phylogeny-guided discovery of biologically active eDNA-derived pentangular polyphenols [J]. *ACS Chem Biol*, 2014, **9**(6): 1267-1272.
[77] Kang HS, Brady SF. Mining soil metagenomes to better understand the evolution of natural product structural diversity: pentangular polyphenols as a case study [J]. *J Am Chem Soc*, 2014, **136**(52): 18111-18119.
[78] Kang HS, Brady SF. Arimetamycin A: improving clinically relevant families of natural products through sequence-guided screening of soil metagenomes [J]. *Angew Chem Int Ed Engl*, 2013, **52**(42): 11063-11067.
[79] Feng Z, Chakraborty D, Dewell SB, *et al*. Environmental DNA-encoded antibiotics fasamycins A and B inhibit FabF in type II fatty acid biosynthesis [J]. *J Am Chem Soc*, 2012, **134**(6): 2981-2987.
[80] Gontang EA, Gaudencio SP, Fenical W, *et al*. Sequence-based analysis of secondary-metabolite biosynthesis in marine actinobacteria [J]. *Appl Environ Microbiol*, 2010, **76**(8): 2487-2499.
[81] Brady SF, Bauer JD, Clarke-Pearson MF, *et al*. Natural products from isnA-containing biosynthetic gene clusters recovered from the genomes of cultured and uncultured bacteria [J]. *J Am Chem Soc*, 2007, **129**(40): 12102-12103.
[82] Kim BS, Kim SY, Park J, *et al*. Sequence-based screening for self-sufficient P450 monooxygenase from a metagenome library [J]. *J Appl Microbiol*, 2007, **102**(5): 1392-1400.
[83] Ziemert N, Ishida K, Weiz A, *et al*. Exploiting the natural diversity of microviridin gene clusters for discovery of novel tricyclic depsipeptides [J]. *Appl Environ Microbiol*, 2010, **76**(11): 3568-3574.
[84] Sanchez C, Mendez C, Salas JA. Indolocarbazole natural products: occurrence, biosynthesis, and biological activity [J]. *Nat Prod Rep*, 2006, **23**(6): 1007-1045.
[85] Chang FY, Ternei MA, Calle PY, *et al*. Discovery and synthetic refactoring of tryptophan dimer gene clusters from the environment [J]. *J Am Chem Soc*, 2013, **135**(47): 17906-17912.
[86] Chang FY, Ternei MA, Calle PY, *et al*. Targeted metagenomics: finding rare tryptophan dimer natural products in the environment [J]. *J Am Chem Soc*, 2015, **137**(18): 6044-6052.
[87] Chang FY, Brady SF. Cloning and characterization of an environmental DNA-derived gene cluster that encodes the biosynthesis of the antitumor substance BE-54017 [J]. *J Am Chem Soc*, 2011, **133**(26): 9996-9999.
[88] Chang FY, Brady SF. Discovery of indolotryptoline antiproliferative agents by homology-guided metagenomic screening [J]. *Proc Natl Acad Sci U S A*, 2013, **110**(7): 2478-2483.
[89] Wenzel SC, Gross F, Zhang Y, *et al*. Heterologous expression of a myxobacterial natural products assembly line in pseudomonads *via* red/ET recombineering [J]. *Chem Biol*, 2005, **12**(3): 349-356.
[90] Eloe-Fadrosh EA, Ivanova NN, Woyke T, *et al*. Metagenom-

ics uncovers gaps in amplicon-based detection of microbial diversity [J]. *Nat Microbiol*, 2016, **1**: 15032.
[91] Nett M, Ikeda H, Moore BS. Genomic basis for natural product biosynthetic diversity in the actinomycetes [J]. *Nat Prod Rep*, 2009, **26**(11): 1362-1384.
[92] Libis V, Antonovsky N, Zhang M, *et al*. Uncovering the biosynthetic potential of rare metagenomic DNA using co-occurrence network analysis of targeted sequences [J]. *Nat Commun*, 2019, **10**(1): 3848.
[93] Nayfach S, Roux S, Seshadri R, *et al*. A genomic catalog of Earth's microbiomes [J]. *Nat Biotechnol*, 2021, **39**(4): 499-509.
[94] Chu J, Vila-Farres X, Inoyama D, *et al*. Discovery of MRSA active antibiotics using primary sequence from the human microbiome [J]. *Nat Chem Biol*, 2016, **12**(12): 1004-1006.
[95] Chu J, Vila-Farres X, Brady SF. Bioactive synthetic-bioinformatic natural product cyclic peptides inspired by nonribosomal peptide synthetase gene clusters from the human microbiome [J]. *J Am Chem Soc*, 2019, **141**(40): 15737-15741.
[96] Stachelhaus T, Mootz HD, Marahiel MA. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases [J]. *Chem Biol*, 1999, **6**(8): 493-505.
[97] Rottig M, Medema MH, Blin K, *et al*. NRPSpredictor2-a web server for predicting NRPS adenylation domain specificity [J]. *Nucleic Acids Res*, 2011, **39**: W362-367.
[98] Baranasic D, Zucko J, Diminic J, *et al*. Predicting substrate specificity of adenylation domains of nonribosomal peptide synthetases and other protein properties by latent semantic indexing [J]. *J Ind Microbiol Biotechnol*, 2014, **41**(2): 461-467.
[99] Xu QS, Zou HC, Pan C. Lysohexaenetides A and B, linear lipopeptides from *Lysobacter* sp. DSM 3655 identified by heterologous expression in Streptomyces [J]. *Chin J Nat Med*, 2023, **21**(6): 454-458.
[100] Chevrette MG, Aicheler F, Kohlbacher O, *et al*. SANDPUMA: ensemble predictions of nonribosomal peptide chemistry reveal biosynthetic diversity across *Actinobacteria* [J]. *Bioinformatics*, 2017, **33**(20): 3202-3210.
[101] Minowa Y, Araki M, Kanehisa M. Comprehensive analysis of distinctive polyketide and nonribosomal peptide structural motifs encoded in microbial genomes [J]. *J Mol Biol*, 2007, **368**(5): 1500-1517.
[102] Weber T, Blin K, Duddela S, *et al*. antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters [J]. *Nucleic Acids Res*, 2015, **43**(W1): W237-243.
[103] Vila-Farres X, Chu J, Inoyama D, *et al*. Antimicrobials inspired by nonribosomal peptide synthetase gene clusters [J]. *J Am Chem Soc*, 2017, **139**(4): 1404-1407.
[104] Wang Z, Koirala B, Hernandez Y, *et al*. A naturally inspired antibiotic to target multidrug-resistant pathogens [J]. *Nature*, 2022, **601**(7894): 606-611.
[105] Wang Z, Koirala B, Hernandez Y, *et al*. Bioinformatic prospecting and synthesis of a bifunctional lipopeptide antibiotic that evades resistance [J]. *Science*, 2022, **376**(6596): 991-996.
[106] Johnston JM, Bulloch EM. Advances in menaquinone biosynthesis: sublocalisation and allosteric regulation [J]. *Curr Opin Struct Biol*, 2020, **65**: 33-41.
[107] Li L, Koirala B, Hernandez Y, *et al*. Identification of structurally diverse menaquinone-binding antibiotics with *in vivo* activity against multidrug-resistant pathogens [J]. *Nat Microbiol*, 2022, **7**(1): 120-131.

Cite this article as: SHEN Yiping, LIU Nan, WANG Zongqiang. Recent advances in the culture-independent discovery of natural products using metagenomic approaches [J]. *Chin J Nat Med*, 2024, **22**(2): 100-111.

Dr. WANG Zongqiang is professor, doctoral supervisor of China Pharmaceutical University. He obtained his Ph.D degree in China Pharmaceutical University and complete postdoctoral training in the Sean F Brady group. Dr WANG's research mainly focus on the discovery and development of antimicrobial drugs from uncultured bacterial using BAC libraries as well as biosynthesis and chemical synthetic approaches. He has published numerous research papers as first author and corresponding author in prestigious journals such as *Nature*, *Science*, *Nat Comm*, and *Angew Chem Int Ed*, and has filed for six international and domestic invention patents. Dr. WANG has taken charged of several national-level projects, including the National Natural Science Youth Fund and the National Key R&D Program (as the project leader), and has led the preclinical research of an antibiotic lead compound developed by his team. He has been appointed as a distinguished professor in Jiangsu Province and a young editorial board member of *Chin J Nat Med*.