J Ind Microbiol Biotechnol (2016) 43:129–141
DOI 10.1007/s10295-015-1706-6

NATURAL PRODUCTS

Culture-independent discovery of natural products from soil metagenomes

Micah Katz[1] · Bradley M. Hover[1] · Sean F. Brady[1]

Received: 26 September 2015 / Accepted: 29 October 2015 / Published online: 19 November 2015

Abstract Bacterial natural products have proven to be invaluable starting points in the development of many currently used therapeutic agents. Unfortunately, traditional culture-based methods for natural product discovery have been deemphasized by pharmaceutical companies due in large part to high rediscovery rates. Culture-independent, or "metagenomic," methods, which rely on the heterologous expression of DNA extracted directly from environmental samples (eDNA), have the potential to provide access to metabolites encoded by a large fraction of the earth's microbial biosynthetic diversity. As soil is both ubiquitous and rich in bacterial diversity, it is an appealing starting point for culture-independent natural product discovery efforts. This review provides an overview of the history of soil metagenome-driven natural product discovery studies and elaborates on the recent development of new tools for sequence-based, high-throughput profiling of environmental samples used in discovering novel natural product biosynthetic gene clusters. We conclude with several examples of these new tools being employed to facilitate the recovery of novel secondary metabolite encoding gene clusters from soil metagenomes and the subsequent heterologous expression of these clusters to produce bioactive small molecules.

Keywords Metagenomics · Drug discovery · Natural products · Culture-independent

Introduction

In the complex web of interactions that comprises life on earth, small molecule natural products play crucial and ubiquitous roles in governing relations within and between species [13, 65, 72]. While humans have used plant extracts for therapeutic purposes for millennia [47, 72], it was Alexander Fleming's isolation of the antibiotic, penicillin, from the *Penicillium rubens* fungus in 1929 that marked the beginning of a new age of natural product-driven medicine [34]. Since Fleming's discovery, natural products derived from environmental microbes have been a major source of lead structures in the development of clinically useful therapeutics—representing more than 60 % of the FDA-approved anti-infective and anti-tumorigenic agents currently on the market [72]. Easily cultured environmental bacteria have been the most productive source of natural products. However, despite decades of historical productivity, pharmaceutical companies largely deemphasized natural product discovery at the end of the past century due to high rediscovery rates from their continued examination of cultured bacteria.

The large-scale sequencing of bacterial genomic DNA over the past 2–3 decades has led to the observation that previous, culture-based natural product discovery programs have likely failed to access the vast majority of bacterial biosynthetic diversity present in the environment. These studies suggest that while a single gram of soil can contain

Special Issue: Natural Product Discovery and Development in the Genomic Era. Dedicated to Professor Satoshi Ōmura for his numerous contributions to the field of natural products.

M. Katz and B. M. Hover contributed equally to the writing of this review.

✉ Sean F. Brady
sbrady@rockefeller.edu

[1] Laboratory of Genetically Encoded Small Molecules, Howard Hughes Medical Institute, The Rockefeller University, 1230 York Avenue, New York, NY 10065, USA

as many as 10,000 unique bacterial species [81, 91], 99 % of these have not been cultivated in the laboratory, and in the small fraction of cultured species [38, 78, 88, 89], large numbers of natural product biosynthetic gene clusters are not expressed in pure culture [2, 5]. A number of different approaches are now being explored to address these problems and gain access to a larger fraction of nature's true biosynthetic diversity. These efforts include such approaches as the re-exploration of ancestral culture collections using next generation sequencing methods, attempts to culture a larger fraction of environmental bacteria in the laboratory, and the development of culture-independent methods for accessing natural products from environmental samples [68]. This review is intended to provide a brief history of culture-independent approaches for the discovery of novel natural products from soil microbiomes and an overview of recent efforts to develop more systematic approaches for accessing biologically active natural products from soil environments.

A culture-independent approach for natural product discovery

The uncultured bacterial majority undoubtedly produces small molecules with therapeutic potential; however, there are currently no general strategies for culturing them [78]. Alternatively, DNA extracted directly from environmental samples (eDNA) can be cloned and accessibly stored in a host organism as a living genetic library. Subsequently, the secondary metabolite-encoding gene clusters contained in this library can be heterologously expressed to access the natural products encoded within the environmental bacterial genomes. This general approach has been termed "metagenomics" and its application to the study of bacterial secondary metabolism is particularly appealing in light of the fact that all of the genes required for natural product biosynthesis are typically tightly clustered on bacterial chromosomes [42]. Approaches used to harvest small molecules from metagenomic libraries (Fig. 1a) can be divided into two general categories: (1) functional screening, which relies on the random, unbiased screening of individual eDNA clones in phenotypic assays to identify bioactive metabolite producing clones; and (2) homology screening, which relies on DNA sequence similarity to identify clones containing a specific gene of interest. These clones are then heterologously expressed to generate molecules of interest.

Direct functional metagenomic screening efforts and their limitations

Early efforts to mine metagenomes for the purpose of natural product discovery began with the construction of eDNA cosmid libraries in *E. coli*, and, in at least one instance, *Streptomyces*, followed by visual or chromatographic screening of these libraries for easily observable phenotypes commonly associated with natural product expression from a clone (e.g., color, antibiosis, HPLC peak) [10–12, 22, 31, 64, 82, 83, 94]. While these simple screens yielded some interesting metabolites (Fig. 2, **1–6**), they were less productive than originally hoped. In retrospect, three easily identifiable issues have likely limited the success of functional metagenomic studies: (1) limitations in the heterologous expression capabilities of the model host organisms currently in use, (2) the insert size of the cosmid based libraries that have dominated the field to date, and (3) the rarity of secondary metabolism biosynthetic machinery in prokaryotic genomes [5, 24, 37, 45, 68, 97]. Although none of these bottlenecks to functional metagenomics have been cleared outright, a number of advances have been made in recent years.

New model hosts for heterologous expression. Model strains used to date were largely identified decades ago for their ability to overproduce a single specific metabolite as opposed to the ability to induce and support the production of diverse natural product biosynthesis at levels suitable for initial natural product discovery. For a metagenomic-based approach to drug discovery to be successful, new model hosts must be identified that can transcriptionally activate and produce molecules encoded by a more diverse set of clusters than is currently accessible. More recently, several studies involving both eDNA and pure culture DNA have employed a broad range of genetically tractable *Actinobacterial* and *Proteobacterial* hosts for the heterologous expression of biosynthetic clusters [1, 6, 8, 14, 16, 23, 40, 46, 48, 49, 53, 56, 74, 79]. These studies were made possible by the use of specially constructed broad-host-range shuttle vectors, such as pJWC1, pTARa or pMBD14, which allow effective eDNA clone shuttling between such taxonomically diverse organisms. While efforts to improve broad-host-range cloning methods continue [97], direct host engineering also provides a promising direction for the heterologous expression of natural products from metagenomes. For heterologous expression experiments, many problems are thought to arise from issues associated with the transcription and translation of exogenous DNA by the host. For example, *E. coli* contains only half of the RNA polymerase sigma factors contained in *Streptomyces* and has been shown to express a significantly lower number of metagenomic genes in comparison [36, 97]. To this end, the overexpression of one of these missing sigma factors, σ^{54}, in *E. coli* facilitated the heterologous expression of an otherwise silent polyketide biosynthetic pathway [87]. These studies suggest

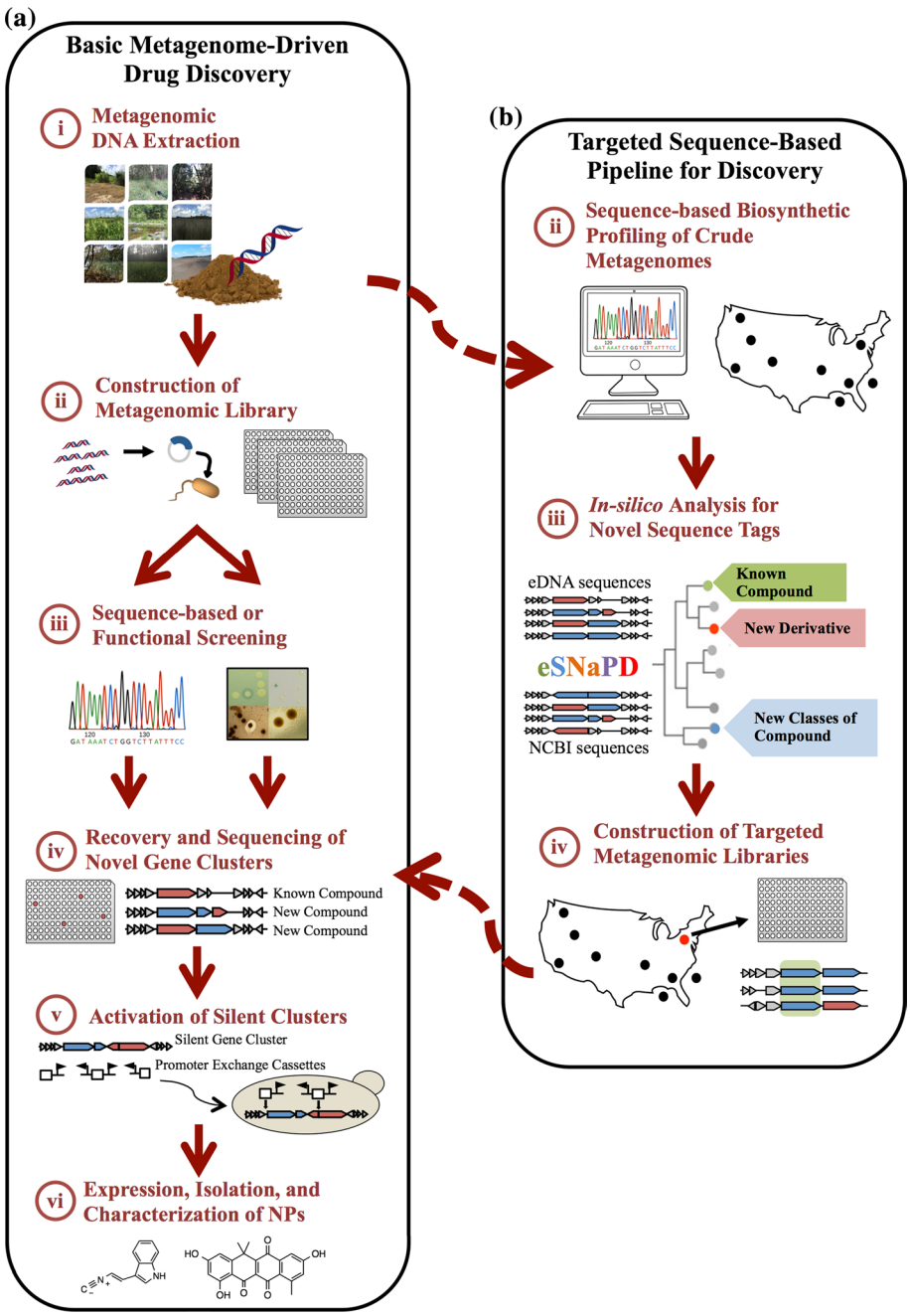

Fig. 1 Basic metagenome-driven drug discovery and the development of a sequence-based pipeline. **a** Metagenomic samples are collected from ecologically and geographically diverse environments, such as soil microbiomes. DNA from the microorganisms that inhabit these environments is then extracted for further analysis. Once this environmental DNA (eDNA) is cloned and ligated into a shuttle vector, it is then transformed into a host cell, creating a living library of organisms housing the eDNA. These metagenomic libraries can then be functionally screened by searching for phenotypes produced by eDNA-encoded natural products or the eDNA can be directly examined using sequence-based screening methods such as PCR. Clones of interest are recovered from metagenomic libraries and the biosynthetic pathways they contain are sequenced and annotated. The natural product encoding pathways of interest are assembled in full and activated through a variety of methods. The activated gene clusters are heterologously expressed, and the resulting natural products are isolated and characterized. **b** Alternatively, crude eDNA from various environmental samples can be screened by PCR to construct a profile of the biosynthetic pathways they contain. The sequences resulting from this screen are phylogenetically organized so as to make predictions of the chemical structures encoded by the greater biosynthetic pathways to which they belong. Samples predicted to contain novel or diverse biosynthetic pathways are prioritized for library construction and subsequent clone recovery, assembly, and heterologous expression

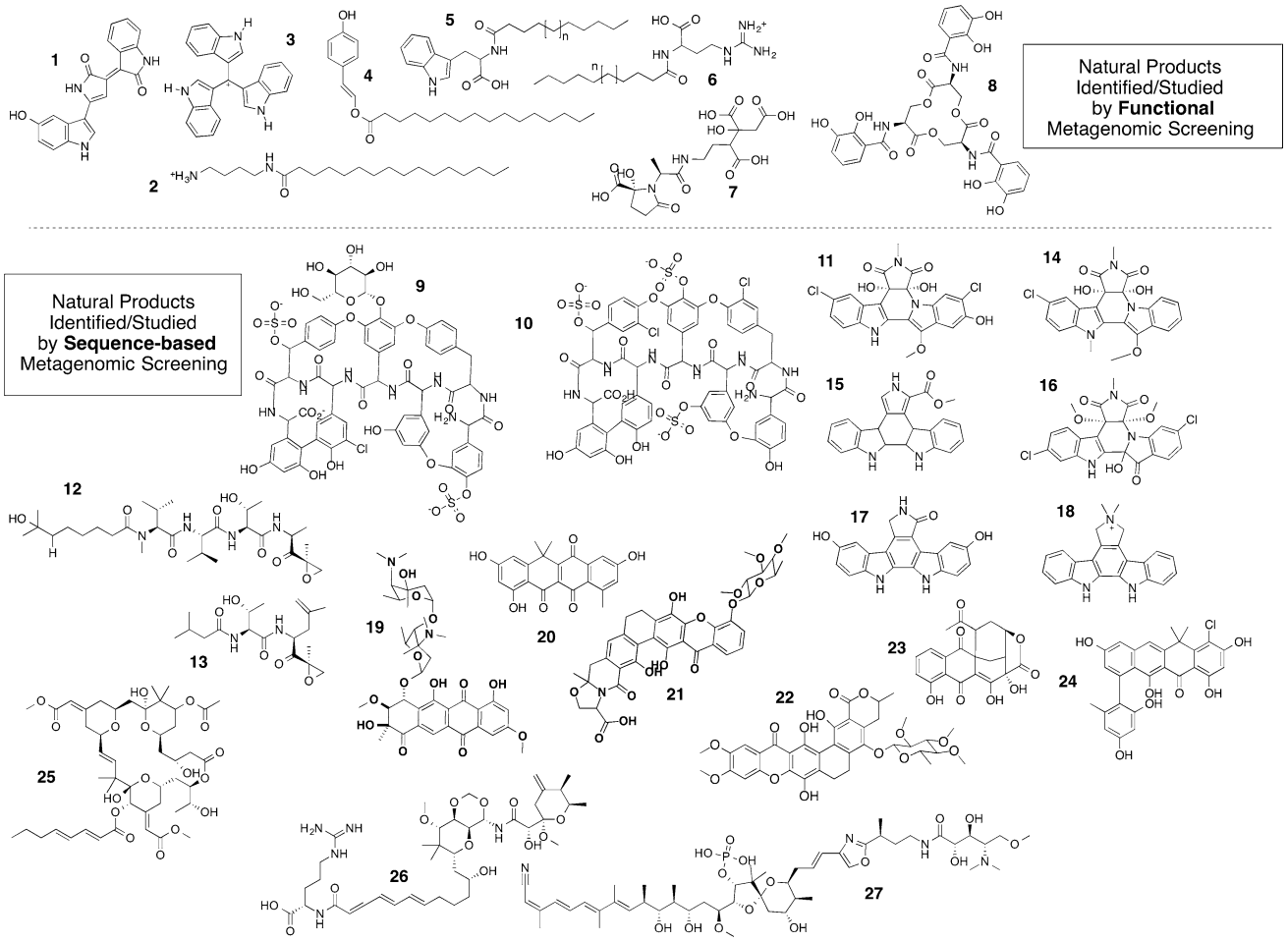

Fig. 2 Examples of microbial natural products discovered and studied through functional and sequence-based metagenomics. (*1*) violacein, (*2*) turbomycin A, (*3*) palmitoylputrescine, (*4*) a long-chain fatty acid enol ester compound, (*5*) long-chain *N*-acyl tryptophan, (*6*) long-chain *N*-acyl arginine; (*7*) vibrioferrin; (*8*) enterobactin; (*9*) and (*10*) sulfo-glycopeptide compounds from Owen et al. [75], (*11*) lazarimide A, (*12*) clarepoxin D, (*13*) landepoxin A, (*14*) BE54017, (*15*) erdasporine A, (*16*) borregomycin A, (*17*) hydroxysporine, (*18*) reductasporine; (*19*) arimetamycin A, (*20*) tetarimycin A, (*21*) arixanthomycin A, (*22*) calixanthomycin A; (*23*) UT-X26, (*24*) AZ154, (*25*) broystatin, (*26*) onnamide A, (*27*) calyculin A

that a combination of both better host selection and host engineering will likely be required to yield truly useful heterologous expression hosts.

Improving library construction methods. Advancements in eDNA cloning methods should dramatically enhance metagenomic screening efforts by facilitating the capture of a larger fraction of complete gene clusters on individual fragments of eDNA. The most efficient eDNA cloning methods are currently limited to capturing ~40 kb of eDNA, thus not allowing for the cloning of the majority of gene clusters, which can exceed 150 kb, on a single clone. Ideally, a metagenomic library would contain large-insert (>100 kb) clones that redundantly cover the total genetic material of an environmental sample. While to date such large-insert metagenomic libraries suitable for molecule discover do not exist, multiple different protocols for metagenomic library construction have been developed recently, each resulting in different average insert-size, total library size, and efficiency of creation, making each differently suited for the various methods of metagenomic screening [8, 42, 62, 99, 103]. Furthermore, a number of methods have been described to improve different aspects of library construction. New technologies such as synchronous coefficient of drag alteration (SCODA) [30, 76], indirect DNA extraction through microbial cell separation [62], and formamide treatment [62] result in the extraction of high molecular weight DNA and eliminate environmental inhibitors, helping to facilitate the transfer of eDNA to metagenomic libraries [30, 76].

Selective library enrichment by functional complementation. As only a small fraction of clones in a metagenomic library will contain biosynthetic clusters, the utility of libraries would be greatly improved by the develop-

ment of methods to selectively enrich for clones with functionally active gene clusters. It has been estimated that secondary metabolite biosynthesis comprises less than 2 % of many bacterial genomes [37]. As a result, screening a complex metagenome for natural product biosynthesis poses a difficult challenge given the small fraction of clones that are likely to contain biosynthetic genes of interest. A potential approach to address this rarity issue is selectively enriching metagenomic libraries for secondary metabolite biosynthetic machinery. This has been illustrated by the functional screening of eDNA clones for phosphopantetheine transferase (PPTase) genes using a PPTase deficient *E. coli* strain [19]. PPTases are responsible for activating nonribosomal peptide synthetase (NRPS) and polyketide synthase (PKS) modules by post-translational attachment of phosphopantetheine (PPT) and are required for the function of NRPS and PKS gene clusters [60], including those which produce the secondary metabolite iron-chelators, siderophores, required for bacterial growth under iron limiting conditions [66] (Fig. 2, **7–8**). Like some NRPS and PKS clusters, some siderophore-producing gene clusters contain their own PPTase-encoding genes, meaning that PPTase-deficient *E. coli* will selectively grow in low-iron conditions only if they harbor eDNA clones that can functionally complement the PPTase activity. While not a universal solution to the rarity of biosynthetic clusters, these experiments illustrate one method to capitalize on the unique functions of natural products to enrich for their biosynthetic machinery.

Sequence-based metagenomic discovery efforts in natural product discovery

Sequence-based metagenomics differs from functional metagenomics in that it does not require heterologous expression to identify metagenomic clones of interest. Instead, this approach circumvents the challenges associated with heterologous expression by identifying gene clusters of interest through sequence analysis. The target pathways can be subsequently explored in heterologous expression studies to generate molecules. Although shotgun-sequencing approaches have been useful for guiding the identification of biosynthetic gene clusters in individual genomes [5] and small, endosymbiont metagenomes [28, 51], the immense size and complexity of soil metagenomes has limited their application for such samples. It has proved extraordinarily challenging to sequence deep enough into soil microbiomes, which can contain up to 10^{4-5} unique species, to generate sequencing data that is broadly useful for natural products gene cluster discovery [78, 81, 88, 91]. To parse this surfeit of genomic sequence in a cost effective and bioinformatically manageable manner, a more targeted approach is required. As a result, most sequence-based metagenomic discovery efforts using soil microbiomes have used degenerate PCR primers to amplify conserved biosynthetic genes of interest followed by amplicon sequencing as a means of identifying clones that contain genes and, correspondingly, gene clusters of interest. In a direct comparison of biosynthetic domain detection from metagenomic samples, the PCR-based sequence targeting approach was shown to be 10–100 times more sensitive than shotgun sequencing in identifying unique sequences [100].

Initial sequence-based approaches targeting biosynthetic genes

The first study to explore a soil microbiome for natural product biosynthetic diversity using a sequence-based metagenomic approach was carried out by Seow et al. in 1997, even before the term "metagenome" was coined [42]. In this study, degenerate primers were designed to target conserved regions of the ketoacyl synthase-β (KS-β) and chain-length-factor (CLF) genes found in all type II PKS biosynthetic gene clusters. Type II PKS gene fragments amplified by PCR from crude eDNA were cloned to yield hybrid type II PKS gene cassettes. The resulting constructs were used in complement with type II PKS gene clusters from cultured bacteria to generate new octaketide and decaketide structures. This work was arguably the first study to show that gene clusters from the soil "multigenome" could be used to potentially generate novel secondary metabolites [84]. In subsequent studies, primers targeting these and other conserved biosynthetic enzymes have been used to interrogate metagenomic libraries instead of crude eDNA. This allowed for the recovery of complete gene clusters in place of isolated biosynthetic genes. In one such study, degenerate primers targeting the isonitrile biosynthetic enzyme family, *isnA*, and oxidative coupling enzyme, *OxyC*, from glycopeptide biosynthesis, were used to probe eDNA libraries [9]. This led to the identification of biosynthetic gene clusters predicted to encode for either novel isonitrile- or glycopeptide-containing natural products, respectively. A small collection of isonitrile-functionalized metabolites was generated through heterologous expression of the *isnA*-containing gene clusters in *E. coli* [9]. In the case of the glycopeptide gene clusters, the sulfotransferases encoded in one group of eDNA gene clusters from a metagenomic sample were used in vitro to modify a teicoplanin aglycone scaffold to generate a novel family of polysulfate glycopeptides [4]. This in vitro approach allowed for the generation of novel glycopeptide congeners using biosynthetic diversity cloned from the soil metagenome without the heterologous expression of the complete

eDNA-derived glycopeptide clusters. The identification of individual biocatalysts, such as sulfotransferases, esterases [102], lipases [96], and β-galactosidases [95] from metagenomic samples through either sequence- or functional-based screening provides an independent strategy for the development of new natural products without requiring the identification, assembly, or activation of complete biosynthetic pathways [46, 92].

While early sequence-based metagenomic studies illustrated the potential of this approach to unlock the biosynthetic diversity that has remained hidden in soil microbiomes, it still falls short of being a truly systematic method for screening inherently complex metagenomes [3, 9]. Described below are a number of recent advances relevant to the development of a more efficient, sequence-based metagenomic pipeline (Fig. 1b). When utilized together, these methods have the potential to create a higher throughput approach to identifying novel natural products from previously unexplored biosynthetic clusters in diverse microbiomes.

Developing tools for barcoding PKS and NRPS gene clusters in microbiomes

Early applications of sequence-based metagenomic methods used primers tailored to genes conserved among relatively small biosynthetic families of natural products. A number of recent studies have focused on expanding these efforts to look for the presence of a broader range of conserved natural product biosynthetic domains, with a major focus on domains found in PKS and NRPS biosynthetic gene clusters [20, 21, 79, 80]. Degenerate primers targeting these common biosynthesis domains are intended to generate complex mixtures of PCR amplicons consisting of domains from hundreds, if not thousands, of gene clusters present in individual environmental samples or metagenomic libraries. Individual next-generation sequencing reads derived from such PCR amplicons have been termed Natural Product Sequence Tags (NPSTs). The scale of the data generated in these studies has necessitated the development of new bioinformatic platforms that can interrogate and organize NPST datasets. While a number of bioinformatics tools have been developed to mine complete sequenced genomes for secondary metabolite biosynthetic clusters (e.g., ClustScan [86] and antiSMASH [7, 67]) these programs generally rely on much larger DNA fragment inputs than are available from metagenomic sequencing efforts.

In response to the need for the tools to analyze very large NPST datasets, programs such as eSNaPD (Environmental Surveyor of Natural Product Diversity) [80] and NaPDoS (Natural Product Domain Search) [104] were developed. These software packages compare NPSTs to reference databases of sequences obtained from characterized biosynthetic gene clusters to predict gene cluster content and potential chemical output of an environmental sample or library. This process is similar to reconstructing the phylogeny of entire organisms within an environment using 16S rRNA sequences [75]. The phylogenetic organization of biosynthetic gene clusters offers two routes for environmental natural product mining. Congeners of characterized natural products can be mined by pursuing gene clusters associated with NPSTs that group closely with the biosynthetic domains encoding for known natural products. Conversely, potentially novel natural products can be pursued by focusing on gene clusters associated with NPSTs that do not group closely with any previously characterized sequences in these families [18, 52, 74]. Such close and distant relative searches are computationally simple; however, once the appropriate sequence similarity cutoffs are defined empirically, their output has proven to be a robust predictor of pathway gene content and chemical output [75]. In addition to its NPST analysis function, eSNaPD organizes data to allow for the archival storage and comparative analysis of NPST datasets from diverse environments. These methods for interrogating biosynthetic diversity in environmental samples and metagenomic libraries are low cost and need little computational power, as they only require the sequencing and subsequent computational comparison of small PCR amplicons instead of complete genomes.

Surveying environments for novel biosynthetic clusters

With the ability to cost effectively analyze the biosynthetic diversity present in a metagenome using a NPST approach, it has been possible to conduct larger scale studies of secondary metabolism in the environment. This involves comparative analysis of multiple environmental samples with the dual goals of both studying the differential distribution of biosynthetic gene clusters in the global microbiome and the identification of specific environments with gene clusters of interest that might serve as starting points for metagenomic library construction efforts. Work in marine environments has constructed geographic secondary metabolite surveys by screening representative collections of marine organism genomes from different locations using ketosynthase (KS) and adenylation domain-derived (AD) degenerate primers to detect the presence of PKS and NRPS biosynthetic machinery in these organisms [29, 39]. These studies indicate that biosynthetic gene cluster families are partitioned across geographically distinct marine environments [29]. The geographic mapping of secondary metabolism has also been applied to diverse soils from around the world using a metagenomic approach to provide more comprehensive profiles of natural product biosynthetic gene clusters in diverse environments [20, 21,

79, 80]. These survey results were similar to those seen in marine culture collections, indicating that sequence similarity occurs in only the most similar of soil types and that geographically and/or ecologically dissimilar environments contain little overlap in their natural product biosynthetic gene clusters [21, 79]. Furthermore, by correlating environment type with secondary metabolite production profiles, specific soil types could be deemed as either rich or poor in NRPS and PKS biosynthetic diversity [20]. Expanding the profiling of biosynthetic diversity present in environmental samples should allow for more strategic decisions to be made as to which metagenomic samples are probed more in depth for gene clusters that either make desired congeners of known molecules or potentially novel families of secondary metabolites.

Utilizing NPST data to recover natural product gene clusters from metagenomic libraries

The NPST approach has also proved to be a productive method for guiding the recovery of specific gene clusters captured in large soil DNA libraries. To do this, large metagenomic libraries are arrayed as manageably sized subpools and screened with collections of primers containing unique barcodes that are used to track NPSTs corresponding to gene clusters of interest to their location of origin within the arrayed library. Natural product gene clusters of interest can then be recovered from the specific subpool of the library, sequenced, and analyzed in heterologous expression studies. This concept was illustrated using a mulit-million membered metagenomic library constructed from eDNA isolated from Arizona desert soil [75]. The screen utilized degenerate NRPS- and KS-targeting primers to identified tags belonging to gene clusters predicted to encode congeners of a variety of antibiotics and anticancer agents. One gene cluster, predicted to encode for a glycopeptide-like antibiotic, was recovered in full and heterologously expressed in *S. toyocaensis:ΔStaL*, a known glycopeptide producer, to produce three novel glycopeptide congeners (Fig. 2, **9**, **10**). This work demonstrated the utility of NPST screening methods in the identification and characterization of diverse natural product biosynthetic clusters captured in soil metagenomic libraries.

Reassembly, activation, and expression of recovered biosynthetic clusters

The ultimate goal of sequence-based metagenomics is the production and characterization of novel natural products through heterologous expression in a model cultured bacterial host. A number of barriers however remain to the routine application of this process. Methods designed to address common barriers standing between gene cluster recovery from metagenomic cosmid libraries and natural product heterologous expression are discussed here.

Reassembly of clusters

Because of the large size of many biosynthetic gene clusters, often clusters of interest cannot be captured on a single cosmid clone and therefore must be recovered on a series of overlapping cosmid clones. To facilitate the study of large gene clusters in a model cultured heterologous expression host, overlapping soil DNA cosmid clones comprising a complete biosynthetic pathway are reassembled into a bacterial artificial chromosome (BAC) using transformation-associated recombination (TAR). In a typical TAR reaction, linearized eDNA cosmids containing components of a complete biosynthetic gene cluster are co-transformed into yeast with a linearized *E. coli:*yeast:*Streptomyces* shuttle capture vector (e.g., pTARa) containing pathway-specific homology arms [32, 49, 101]. Once in yeast, the overlapping cosmids assemble into a continuous fragment of DNA and the shuttle vector captures this target DNA sequence through homologous recombination to yield a BAC containing the reassembled biosynthetic cluster [56]. Reassembled gene clusters can then be transferred from yeast into genetically tractable model bacterial strains for heterologous expression [1, 6, 8, 14, 23, 40, 46, 48, 49, 56].

Activation of clusters

Although some eDNA derived gene clusters will immediately lead to the production of the metabolite(s) they encode upon introduction into a heterologous host, most remain silent. The activation of silent biosynthetic gene clusters remains a significant challenge for almost all microbial natural product discovery programs and the development of methods to overcome this challenge is a very active area of research. Several approaches have been successful in the activation of otherwise silent gene clusters in native host organisms, such as the simulation of environmental conditions [33], co-culturing [25], and the use of histone deacetylase (HDAC) inhibitors [70]. To the best of our knowledge, similar methods have not yet been widely examined as potential mechanisms for inducing eDNA derived gene cluster expression in heterologous hosts.

The exploitation of genetic regulatory elements has been used to activate silent gene clusters in both native and heterologous host organisms. Strong promoters, such as the ermE* erythromycin resistance encoding promoter, can be placed in front of positive regulatory elements [50, 61, 73], or individual biosynthetic genes themselves [73] to constitutively activate biosynthetic pathways, resulting in the production of secondary metabolites. This method has been used to produce 6-*epi*-alteramides, candicidins, antimycins

[73] and the 51-membered macrolide, stambomycin [61] in native hosts and the eDNA-encoded tetarimycin A [50] in a heterologous host. A number of studies have shown complete refactoring of biosynthetic pathways as a means of accessing metabolites from silent gene clusters, resulting in the heterologous expression of spectinabilin [85] and several new polycyclic tetramate macrolactams [63]. Recently, yeast homologous recombination was employed to expedite multiplex promoter exchange in a silent eDNA-derived gene cluster, resulting in the production of the indolotryptoline antiproliferative agents, lazarimides A and B (Fig. 2, **11**) [69]. In another example, every gene in the eDNA-derived *erd* gene cluster was cloned under the control of an inducible promoter to yield the novel carboxy-indolocarbazole containing tryptophan dimer erdasporine [17]. To date, complete gene cluster refactoring approaches have yet to be extensively applied to soil metagenomic discovery efforts.

Sequence-based metagenomic natural product mining in action

A growing number of natural products have been identified from soil metagenomes using sequence-based discovery efforts. These studies can largely be divided into two groups. One effort is focused on NPST's that are closely related to known sequences in an effort to identify additional congeners of a known natural product or natural product class. The second effort is focused on NPST's that are phylogenetically distinct from any known sequence in the target gene family in an effort to identify new structural classes of natural products. Examples of natural products arising from our own research using each approach are highlighted here.

From the assessment of biosynthetic diversity to the characterization of novel natural products

Epoxyketone protease inhibitors (EPIs) are mixed PKS:NRPS-derived natural products that inhibit the human 20S proteasome, proving cytotoxic through the accumulation of poly-ubiquitinated proteins in the cell [58, 59]. While no natural EPIs have progressed through clinical trials, synthetic EPI analogues have been successfully used to treat multiple myelomas [58, 59]. Their proven clinical track record, yet limited characterized natural chemical diversity, made them appealing targets for sequence guided metagenomic discovery efforts. As they are partially derived from PKS biosynthetic machinery, gene clusters encoding EPIs can be identified by comparing KS-domain sequence tags to the KS domains from known epoxyketone biosynthetic gene clusters. To identify the distribution of EPI-encoding clusters in the environment, the sequences of the reference KS domains were compared with the results from a screen of 185 metagenomic samples from around the globe, using bar-coded, degenerate KS-derived primers [74]. 99 unique EPI-like sequences that grouped into six distinct clades were identified across these metagenomic samples. None of the 99 EPI relevant sequence tags were found in any whole genome sequences in publically available databases. Eleven potential EPI gene clusters were recovered from soil eDNA libraries. Characterization of these full gene clusters indicated that nine of them were predicted to encode EPIs. Two clusters were chosen for heterologous expression studies based on a desire to both expand the congener diversity of known EPIs and to identify novel EPI subclasses. These two clusters were reassembled (as necessary) using TAR and shuttled into *Streptomcyes albus* for heterologous expression studies. This led to the characterization of seven novel EPI natural products: clarepoxins A-E (Fig. 2, **12**) and landepoxins A and B (Fig. 2, **13**).

Targeted approaches to recovering tryptophan dimers from the environment

Natural products that arise from the dimerization of tryptophan (i.e., tryptophan dimers, TDs) exhibit potent anticancer, antibacterial, and antifungal activity, making them appealing targets for sequence based mining efforts [71]. Early metagenomic discovery efforts focused on the screening of archived eDNA libraries for TD gene clusters using degenerate primers based on an alignment of chromopyrrolic acid synthase (CPAS) genes, which encode the enzymes involved in the dimerization of oxytryptophan [15, 17]. These experiments resulted in the discovery of a number of TD gene clusters and the characterization of both known [the antitumor molecule BE-54017 [14] (Fig. 2, **14**) and novel TDs (e.g., erdasporine [17], borregomycins [15]) (Fig. 2, **15**–**16**). While these efforts identified novel TD structures, the TD core upon which each structure was based had been reported previously. In an effort to expand the scope of our search for new TD structures, crude eDNA extracts from geographically diverse United States soils were screened using bar-coded degenerate primers targeting CPAS genes. To identify metagenomes with the potential to encode for new TD core structures, these data were searched for unique sequence tags that did not fall into any existing CPAS clades [18]. A soil sample from the Sonoran Desert was identified as containing two CPAS tags that did not match any previously characterized TD biosynthetic gene clusters. One of these eDNA-derived genes fell into the same clade as a known family of TD CPAS's while the other did not resemble any existing CPAS family, suggesting the cluster it arose from might encode a metabolite representing a novel class of TD structures. The soil of interest

was expanded into a full metagenomic library and the complete gene clusters containing the identified sequences were recovered. The first cluster was heterologously expressed to produce hydroxysporine (Fig. 2, **17**): a pyrrolinone indolocarbazole core containing TD that had been reported as a synthetic compound but never seen in nature. The second cluster, which was associated with a phylogenetically novel CPAS sequence tag, encoded the new TD, reductasporine (Fig. 2, **18**), which contains a never before seen pyrrolinium indolocarbazole core thus supporting the premise that "outlier" sequence tags have a high potential of being associated with functionally novel gene clusters. The targeting of phylogenetically unique sequence tags as means of identifying gene clusters that encode structurally novel classes of natural products has also been explored using KS_{β} genes from type II PKS biosynthesis. This work has led to the identification of a number of natural products with either new or rarely seen ring systems (Fig. 2, **19**–**24**) [31, 53, 54, 57], further illustrating the utility of targeted screens in conjunction with a metagenomic natural product-mining pipeline to discover diverse new secondary metabolites.

The application of sequence-based metagenomic natural product discovery to other environments

The utility of a sequence-based metagenomic pipeline for natural product discovery is not limited to soil microbiomes. Elements of this method have been utilized to find and elucidate biosynthetic gene clusters from marine environments that encode for previously known therapeutically relevant molecules that have not been natively expressed or synthesized with sufficient yields for practical utility [90]. Metagenomic analyses of the symbiotic bacteria found associated with organisms like marine sponges and tunicates have led to the discovery and/or further characterization of diverse metabolites, such as the antitumor polyketides, bryostatins [26, 27, 43, 77, 90] (Fig. 2, **25**), onnamide (Fig. 2, **26**) [44, 98], and polytheonamides [35, 41, 98]. In the case of the cytotoxic natural product calyculin A (Fig. 1, **27**), it took a targeted sequence-based metagenomic approach to identify the symbiotic organism and associated biosynthetic gene cluster responsible for its production [93] nearly three decades after its initial discovery from extracts of the marine sponge *Discodermia calyx* [55]. Because the structure of calyculin A suggested that its assembly required both NRPS and PKS machinery, metagenomic samples were collected from *D. calyx* and screened using both KS- and AD-derived degenerate primers to generate sequence tags that were phylogenetically organized and predicted to encode for calyculin A [93].

Conclusion

Metagenomic screening approaches offer an alternative method for natural products discovery that has the potential to unlock previously unexplored biosynthetic diversity. While there remains much potential for its improvement, the examples outlined above show that this general approach has now developed to the point where it can be used to routinely recover previously unstudied gene clusters and functionally access novel bioactive secondary metabolites from diverse soil metagenomes.

References

1. Aakvik T, Degnes KF, Dahlsrud R, Schmidt F, Dam R, Yu L, Volker U, Ellingsen TE, Valla S (2009) A plasmid RK2-based broad-host-range cloning vector useful for transfer of metagenomic libraries to a variety of bacterial species. FEMS Microbiol Lett 296:149–158. doi:10.1111/j.1574-6968.2009.01639.x
2. Abrudan MI, Smakman F, Grimbergen AJ, Westhoff S, Miller EL, van Wezel GP, Rozen DE (2015) Socially mediated induction and suppression of antibiosis during bacterial coexistence. Proc Natl Acad Sci USA 112:11054–11059. doi:10.1073/pnas.1504076112
3. Banik JJ, Brady SF (2008) Cloning and characterization of new glycopeptide gene clusters found in an environmental DNA megalibrary. Proc Natl Acad Sci USA 105:17273–17277. doi:10.1073/pnas.0807564105
4. Banik JJ, Craig JW, Calle PY, Brady SF (2010) Tailoring enzyme-rich environmental DNA clones: a source of enzymes for generating libraries of unnatural natural products. J Am Chem Soc 132:15661–15670. doi:10.1021/ja105825a
5. Bentley SD, Chater KF, Cerdeno-Tarraga AM, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D, Bateman A, Brown S, Chandra G, Chen CW, Collins M, Cronin A, Fraser A, Goble A, Hidalgo J, Hornsby T, Howarth S, Huang CH, Kieser T, Larke L, Murphy L, Oliver K, O'Neil S, Rabbinowitsch E, Rajandream MA, Rutherford K, Rutter S, Seeger K, Saunders D, Sharp S, Squares R, Squares S, Taylor K, Warren T, Wietzorrek A, Woodward J, Barrell BG, Parkhill J, Hopwood DA (2002) Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). Nature 417:141–147. doi:10.1038/417141a
6. Biggins JB, Kang HS, Ternei MA, DeShazer D, Brady SF (2014) The chemical arsenal of *Burkholderia pseudomallei* is essential for pathogenicity. J Am Chem Soc 136:9484–9490. doi:10.1021/ja504617n
7. Blin K, Medema MH, Kazempour D, Fischbach MA, Breitling R, Takano E, Weber T (2013) antiSMASH 2.0–a versatile platform for genome mining of secondary metabolite producers. Nucleic Acids Res 41:W204–W212. doi:10.1093/nar/gkt449
8. Brady SF (2007) Construction of soil environmental DNA cosmid libraries and screening for clones that produce biologically active small molecules. Nat Protoc 2:1297–1305. doi:10.1038/nprot.2007.195
9. Brady SF, Bauer JD, Clarke-Pearson MF, Daniels R (2007) Natural products from isnA-containing biosynthetic gene clusters recovered from the genomes of cultured and uncultured bacteria. J Am Chem Soc 129:12102–12103. doi:10.1021/ja075492v

10. Brady SF, Chao CJ, Clardy J (2002) New natural product families from an environmental DNA (eDNA) gene cluster. J Am Chem Soc 124:9968–9969
11. Brady SF, Chao CJ, Clardy J (2004) Long-chain *N*-acyltyrosine synthases from environmental DNA. Appl Environ Microbiol 70:6865–6870. doi:10.1128/AEM.70.11.6865-6870.2004
12. Brady SF, Chao CJ, Handelsman J, Clardy J (2001) Cloning and heterologous expression of a natural product biosynthetic gene cluster from eDNA. Org Lett 3:1981–1984
13. Butler MS (2008) Natural products to drugs: natural product-derived compounds in clinical trials. Natural Prod Rep 25:475–516. doi:10.1039/b514294f
14. Chang FY, Brady SF (2011) Cloning and characterization of an environmental DNA-derived gene cluster that encodes the biosynthesis of the antitumor substance BE-54017. J Am Chem Soc 133:9996–9999. doi:10.1021/ja2022653
15. Chang FY, Brady SF (2013) Discovery of indolotryptoline antiproliferative agents by homology-guided metagenomic screening. Proc Natl Acad Sci USA 110:2478–2483. doi:10.1073/pnas.1218073110
16. Chang FY, Brady SF (2014) Characterization of an environmental DNA-derived gene cluster that encodes the bisindolylmaleimide methylarcyriarubin. Chembiochem Eur J Chem Biol 15:815–821. doi:10.1002/cbic.201300756
17. Chang FY, Ternei MA, Calle PY, Brady SF (2013) Discovery and synthetic refactoring of tryptophan dimer gene clusters from the environment. J Am Chem Soc 135:17906–17912. doi:10.1021/ja408683p
18. Chang FY, Ternei MA, Calle PY, Brady SF (2015) Targeted metagenomics: finding rare tryptophan dimer natural products in the environment. J Am Chem Soc 137:6044–6052. doi:10.1021/jacs.5b01968
19. Charlop-Powers Z, Banik JJ, Owen JG, Craig JW, Brady SF (2013) Selective enrichment of environmental DNA libraries for genes encoding nonribosomal peptides and polyketides by phosphopantetheine transferase-dependent complementation of siderophore biosynthesis. ACS Chem Biol 8:138–143. doi:10.1021/cb3004918
20. Charlop-Powers Z, Owen JG, Reddy BV, Ternei MA, Brady SF (2014) Chemical-biogeographic survey of secondary metabolism in soil. Proc Natl Acad Sci USA 111:3757–3762. doi:10.1073/pnas.1318021111
21. Charlop-Powers Z, Owen JG, Reddy BV, Ternei MA, Guimaraes DO, de Frias UA, Pupo MT, Seepe P, Feng Z, Brady SF (2014) Global biogeographic sampling of bacterial secondary metabolism. eLife 4:e05048. doi:10.7554/eLife.05048
22. Courtois S, Frostegard A, Goransson P, Depret G, Jeannin P, Simonet P (2001) Quantification of bacterial subgroups in soil: comparison of DNA extracted directly from soil or from cells previously released by density gradient centrifugation. Environ Microbiol 3:431–439
23. Craig JW, Chang FY, Brady SF (2009) Natural products from environmental DNA hosted in *Ralstonia metallidurans*. ACS Chem Biol 4:23–28. doi:10.1021/cb8002754
24. Craig JW, Chang FY, Kim JH, Obiajulu SC, Brady SF (2010) Expanding small-molecule functional metagenomics through parallel screening of broad-host-range cosmid environmental DNA libraries in diverse proteobacteria. Appl Environ Microbiol 76:1633–1641. doi:10.1128/AEM.02169-09
25. Cueto M, Jensen PR, Kauffman C, Fenical W, Lobkovsky E, Clardy J (2001) Pestalone, a new antibiotic produced by a marine fungus in response to bacterial challenge. J Nat Prod 64:1444–1446
26. Davidson SK, Allen SW, Lim GE, Anderson CM, Haygood MG (2001) Evidence for the biosynthesis of bryostatins by the bacterial symbiont "Candidatus Endobugula sertula" of the bryozoan *Bugula neritina*. Appl Environ Microbiol 67:4531–4537
27. Davidson SK, Haygood MG (1999) Identification of sibling species of the bryozoan *Bugula neritina* that produce different anticancer bryostatins and harbor distinct strains of the bacterial symbiont "Candidatus Endobugula sertula". Biol Bull 196:273–280
28. Donia MS, Ruffner DE, Cao S, Schmidt EW (2011) Accessing the hidden majority of marine natural products through metagenomics. ChemBioChem 12:1230–1236. doi:10.1002/cbic.201000780
29. Edlund A, Loesgen S, Fenical W, Jensen PR (2011) Geographic distribution of secondary metabolite genes in the marine actinomycete *Salinispora arenicola*. Appl Environ Microbiol 77:5916–5925. doi:10.1128/AEM.00611-11
30. Engel K, Pinnell L, Cheng J, Charles TC, Neufeld JD (2012) Nonlinear electrophoresis for purification of soil DNA for metagenomics. J Microbiol Methods 88:35–40. doi:10.1016/j.mimet.2011.10.007
31. Feng Z, Kallifidas D, Brady SF (2011) Functional analysis of environmental DNA-derived type II polyketide synthases reveals structurally diverse secondary metabolites. Proc Natl Acad Sci USA 108:12629–12634. doi:10.1073/pnas.1103921108
32. Feng Z, Kim JH, Brady SF (2010) Fluostatins produced by the heterologous expression of a TAR reassembled environmental DNA derived type II PKS gene cluster. J Am Chem Soc 132:11902–11903. doi:10.1021/ja104550p
33. Ferrari BC, Binnerup SJ, Gillings M (2005) Microcolony cultivation on a soil substrate membrane system selects for previously uncultured soil bacteria. Appl Environ Microbiol 71:8714–8720. doi:10.1128/AEM.71.12.8714-8720.2005
34. Fleming A (2001) On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae. 1929. Bull World Health Organ 79(8):780–790
35. Freeman MF, Gurgui C, Helf MJ, Morinaka BI, Uria AR, Oldham NJ, Sahl HG, Matsunaga S, Piel J (2012) Metagenome mining reveals polytheonamides as posttranslationally modified ribosomal peptides. Science 338:387–390. doi:10.1126/science.1226121
36. Gabor EM, Alkema WB, Janssen DB (2004) Quantifying the accessibility of the metagenome by random expression cloning techniques. Environ Microbiol 6:879–886. doi:10.1111/j.1462-2920.2004.00640.x
37. Garcia JA, Fernandez-Guerra A, Casamayor EO (2011) A close relationship between primary nucleotides sequence structure and the composition of functional genes in the genome of prokaryotes. Mol Phylogenet Evol 61:650–658. doi:10.1016/j.ympev.2011.08.011
38. Giovannoni SJ, Britschgi TB, Moyer CL, Field KG (1990) Genetic diversity in Sargasso Sea bacterioplankton. Nature 345:60–63. doi:10.1038/345060a0
39. Gontang EA, Gaudencio SP, Fenical W, Jensen PR (2010) Sequence-based analysis of secondary-metabolite biosynthesis in marine actinobacteria. Appl Environ Microbiol 76:2487–2499. doi:10.1128/AEM.02852-09
40. Gregory MA, Till R, Smith MC (2003) Integration site for Streptomyces phage phiBT1 and development of site-specific integrating vectors. J Bacteriol 185:5320–5323
41. Hamada T, Matsunaga S, Fujiwara M, Fujita K, Hirota H, Schmucki R, Guntert P, Fusetani N (2010) Solution structure of polytheonamide B, a highly cytotoxic nonribosomal polypeptide from marine sponge. J Am Chem Soc 132:12941–12945. doi:10.1021/ja104616z
42. Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM (1998) Molecular biological access to the chemistry of

unknown soil microbes: a new frontier for natural products. Chem Biol 5:R245–R249
43. Hildebrand M, Waggoner LE, Liu H, Sudek S, Allen S, Anderson C, Sherman DH, Haygood M (2004) bryA: an unusual modular polyketide synthase gene from the uncultivated bacterial symbiont of the marine bryozoan *Bugula neritina*. Chem Biol 11:1543–1552. doi:10.1016/j.chembiol.2004.08.018
44. Hori M, Saito S, Shin YZ, Ozaki H, Fusetani N, Karaki H (1993) Mycalolide-B, a novel and specific inhibitor of actomyosin ATPase isolated from marine sponge. FEBS Lett 322:151–154
45. Ikeda H, Ishikawa J, Hanamoto A, Shinose M, Kikuchi H, Shiba T, Sakaki Y, Hattori M, Omura S (2003) Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. Nat Biotechnol 21:526–531. doi:10.1038/nbt820
46. Iqbal HA, Feng Z, Brady SF (2012) Biocatalysts and small molecule products from metagenomic studies. Curr Opin Chem Biol 16:109–116. doi:10.1016/j.cbpa.2012.02.015
47. Ji HF, Li XJ, Zhang HY (2009) Natural products and drug discovery. Can thousands of years of ancient medical knowledge lead us to new and powerful drug combinations in the fight against cancer and dementia? EMBO Rep 10:194–200. doi:10.1038/embor.2009.12
48. Kakirde KS, Wild J, Godiska R, Mead DA, Wiggins AG, Goodman RM, Szybalski W, Liles MR (2011) Gram negative shuttle BAC vector for heterologous expression of metagenomic libraries. Gene 475:57–62. doi:10.1016/j.gene.2010.11.004
49. Kallifidas D, Brady SF (2012) Reassembly of functionally intact environmental DNA-derived biosynthetic gene clusters. Methods Enzymol 517:225–239. doi:10.1016/B978-0-12-404634-4.00011-5
50. Kallifidas D, Kang HS, Brady SF (2012) Tetarimycin A, an MRSA-active antibiotic identified through induced expression of environmental DNA gene clusters. J Am Chem Soc 134:19552–19555. doi:10.1021/ja3093828
51. Kampa A, Gagunashvili AN, Gulder TA, Morinaka BI, Daolio C, Godejohann M, Miao VP, Piel J, Andresson O (2013) Metagenomic natural product discovery in lichen provides evidence for a family of biosynthetic pathways in diverse symbioses. Proc Natl Acad Sci USA 110:E3129–E3137. doi:10.1073/pnas.1305867110
52. Kang HS, Brady SF (2013) Arimetamycin A: improving clinically relevant families of natural products through sequence-guided screening of soil metagenomes. Angew Chem 52:11063–11067. doi:10.1002/anie.201305109
53. Kang HS, Brady SF (2014) Arixanthomycins A–C: phylogeny-guided discovery of biologically active eDNA-derived pentangular polyphenols. ACS Chem Biol 9:1267–1272. doi:10.1021/cb500141b
54. Kang HS, Brady SF (2014) Mining soil metagenomes to better understand the evolution of natural product structural diversity: pentangular polyphenols as a case study. J Am Chem Soc 136:18111–18119. doi:10.1021/ja510606j
55. Kato Y, Fusetani N, Matsunaga S, Hashimoto K, Fujita S, Furuya T (1986) The bioactive marine metabolites. 16. calyculin-a, a novel antitumor metabolite from the marine sponge discodermia-calyx. J Am Chem Soc 108:2780–2781. doi:10.1021/Ja00270a061
56. Kim JH, Feng Z, Bauer JD, Kallifidas D, Calle PY, Brady SF (2010) Cloning large natural product gene clusters from the environment: piecing environmental DNA gene clusters back together with TAR. Biopolymers 93:833–844. doi:10.1002/bip.21450
57. King RW, Bauer JD, Brady SF (2009) An environmental DNA-derived type II polyketide biosynthetic pathway encodes the biosynthesis of the pentacyclic polyketide erdacin. Angew Chem 48:6257–6261. doi:10.1002/anie.200901209
58. Kisselev AF (2008) Joining the army of proteasome inhibitors. Chem Biol 15:419–421. doi:10.1016/j.chembiol.2008.04.010
59. Kisselev AF, van der Linden WA, Overkleeft HS (2012) Proteasome inhibitors: an expanding army attacking a unique target. Chem Biol 19:99–115. doi:10.1016/j.chembiol.2012.01.003
60. Lambalot RH, Gehring AM, Flugel RS, Zuber P, LaCelle M, Marahiel MA, Reid R, Khosla C, Walsh CT (1996) A new enzyme superfamily—the phosphopantetheinyl transferases. Chem Biol 3:923–936
61. Laureti L, Song L, Huang S, Corre C, Leblond P, Challis GL, Aigle B (2011) Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in *Streptomyces ambofaciens*. Proc Natl Acad Sci USA 108:6258–6263. doi:10.1073/pnas.1019077108
62. Liles MR, Williamson LL, Rodbumrer J, Torsvik V, Goodman RM, Handelsman J (2008) Recovery, purification, and cloning of high-molecular-weight DNA from soil microorganisms. Appl Environ Microbiol 74:3302–3305. doi:10.1128/AEM.02630-07
63. Luo Y, Huang H, Liang J, Wang M, Lu L, Shao Z, Cobb RE, Zhao H (2013) Activation and characterization of a cryptic polycyclic tetramate macrolactam biosynthetic gene cluster. Nat Commun 4:2894. doi:10.1038/ncomms3894
64. MacNeil IA, Tiong CL, Minor C, August PR, Grossman TH, Loiacono KA, Lynch BA, Phillips T, Narula S, Sundaramoorthi R, Tyler A, Aldredge T, Long H, Gilman M, Holt D, Osburne MS (2001) Expression and isolation of antimicrobial small molecules from soil DNA libraries. J Mol Microbiol Biotechnol 3:301–308
65. Maplestone RA, Stone MJ, Williams DH (1992) The evolutionary role of secondary metabolites–a review. Gene 115:151–157
66. Matzanke BF, Muller GI, Raymond KN (1984) Hydroxamate siderophore mediated iron uptake in *E. coli*: stereospecific recognition of ferric rhodotorulic acid. Biochem Biophys Res Commun 121:922–930
67. Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R (2011) antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res 39:W339–W346. doi:10.1093/nar/gkr466
68. Milshteyn A, Schneider JS, Brady SF (2014) Mining the metabiome: identifying novel natural products from microbial communities. Chem Biol 21:1211–1223. doi:10.1016/j.chembiol.2014.08.006
69. Montiel D, Kang H-S, Chang F-Y, Charlop-Powers Z, Brady SF (2015) Yeast homologous recombination-based promoter engineering for the activation of silent natural product biosynthetic gene clusters. Proc Natl Acad Sci 112:8953–8958. doi:10.1073/pnas.1507606112
70. Moore JM, Bradshaw E, Seipke RF, Hutchings MI, McArthur M (2012) Use and discovery of chemical elicitors that stimulate biosynthetic gene clusters in Streptomyces bacteria. Methods Enzymol 517:367–385. doi:10.1016/B978-0-12-404634-4.00018-8
71. Nakano H, Omura S (2009) Chemical biology of natural indolocarbazole products: 30 years since the discovery of staurosporine. J Antibiot 62:17–26. doi:10.1038/ja.2008.4
72. Newman DJ, Cragg GM (2012) Natural products as sources of new drugs over the 30 years from 1981 to 2010. J Natl Prod 75:311–335. doi:10.1021/np200906s
73. Olano C, Garcia I, Gonzalez A, Rodriguez M, Rozas D, Rubio J, Sanchez-Hidalgo M, Brana AF, Mendez C, Salas JA (2014)

Activation and identification of five clusters for secondary metabolites in Streptomyces albus J1074. Microb Biotechnol 7:242–256. doi:10.1111/1751-7915.12116
74. Owen JG, Charlop-Powers Z, Smith AG, Ternei MA, Calle PY, Reddy BV, Montiel D, Brady SF (2015) Multiplexed metagenome mining using short DNA sequence tags facilitates targeted discovery of epoxyketone proteasome inhibitors. Proc Natl Acad Sci USA 112:4221–4226. doi:10.1073/pnas.1501124112
75. Owen JG, Reddy BV, Ternei MA, Charlop-Powers Z, Calle PY, Kim JH, Brady SF (2013) Mapping gene clusters within arrayed metagenomic libraries to expand the structural diversity of biomedically relevant natural products. Proc Natl Acad Sci USA 110:11797–11802. doi:10.1073/pnas.1222159110
76. Pel J, Broemeling D, Mai L, Poon HL, Tropini G, Warren RL, Holt RA, Marziali A (2009) Nonlinear electrophoretic response yields a unique parameter for separation of biomolecules. Proc Natl Acad Sci USA 106:14796–14801. doi:10.1073/pnas.0907402106
77. Pettit GR (1991) The bryostatins. Fortschritte der Chemie organischer Naturstoffe = Progress in the chemistry of organic natural products Progres dans la chimie des substances organiques naturelles 57:153–195
78. Rappe MS, Giovannoni SJ (2003) The uncultured microbial majority. Annu Rev Microbiol 57:369–394. doi:10.1146/annurev.micro.57.030502.090759
79. Reddy BV, Kallifidas D, Kim JH, Charlop-Powers Z, Feng Z, Brady SF (2012) Natural product biosynthetic gene diversity in geographically distinct soil microbiomes. Appl Environ Microbiol 78:3744–3752. doi:10.1128/AEM.00102-12
80. Reddy BV, Milshteyn A, Charlop-Powers Z, Brady SF (2014) eSNaPD: a versatile, web-based bioinformatics platform for surveying and mining natural product biosynthetic diversity from metagenomes. Chem Biol 21:1023–1033. doi:10.1016/j.chembiol.2014.06.007
81. Roesch LF, Fulthorpe RR, Riva A, Casella G, Hadwin AK, Kent AD, Daroub SH, Camargo FA, Farmerie WG, Triplett EW (2007) Pyrosequencing enumerates and contrasts soil microbial diversity. ISME J 1:283–290. doi:10.1038/ismej.2007.53
82. Rondon MR, August PR, Bettermann AD, Brady SF, Grossman TH, Liles MR, Loiacono KA, Lynch BA, MacNeil IA, Minor C, Tiong CL, Gilman M, Osburne MS, Clardy J, Handelsman J, Goodman RM (2000) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. Appl Environ Microbiol 66:2541–2547
83. Rondon MR, Raffel SJ, Goodman RM, Handelsman J (1999) Toward functional genomics in bacteria: analysis of gene expression in *Escherichia coli* from a bacterial artificial chromosome library of Bacillus cereus. Proc Natl Acad Sci USA 96:6451–6455
84. Seow KT, Meurer G, Gerlitz M, Wendt-Pienkowski E, Hutchinson CR, Davies J (1997) A study of iterative type II polyketide synthases, using bacterial genes cloned from soil DNA: a means to access and use genes from uncultured microorganisms. J Bacteriol 179:7360–7368
85. Shao Z, Rao G, Li C, Abil Z, Luo Y, Zhao H (2013) Refactoring the silent spectinabilin gene cluster using a plug-and-play scaffold. ACS synthetic biology 2:662–669. doi:10.1021/sb400058n
86. Starcevic A, Zucko J, Simunkovic J, Long PF, Cullum J, Hranueli D (2008) ClustScan: an integrated program package for the semi-automatic annotation of modular biosynthetic gene clusters and in silico prediction of novel chemical structures. Nucleic Acids Res 36:6882–6892. doi:10.1093/nar/gkn685
87. Stevens DC, Conway KR, Pearce N, Villegas-Penaranda LR, Garza AG, Boddy CN (2013) Alternative sigma factor over-expression enables heterologous expression of a type II polyketide biosynthetic pathway in *Escherichia coli*. PLoS One 8:e64858. doi:10.1371/journal.pone.0064858
88. Torsvik V, Goksoyr J, Daae FL (1990) High diversity in DNA of soil bacteria. Appl Environ Microbiol 56:782–787
89. Torsvik V, Salte K, Sorheim R, Goksoyr J (1990) Comparison of phenotypic diversity and DNA heterogeneity in a population of soil bacteria. Appl Environ Microbiol 56:776–781
90. Trindade M, van Zyl LJ, Navarro-Fernandez J, Abd Elrazak A (2015) Targeted metagenomics as a tool to tap into marine natural product diversity for the discovery and production of drug candidates. Front Microbiol 6:890. doi:10.3389/fmicb.2015.00890
91. Tringe SG, von Mering C, Kobayashi A, Salamov AA, Chen K, Chang HW, Podar M, Short JM, Mathur EJ, Detter JC, Bork P, Hugenholtz P, Rubin EM (2005) Comparative metagenomics of microbial communities. Science 308:554–557. doi:10.1126/science.1107851
92. Uchiyama T, Miyazaki K (2010) Product-induced gene expression, a product-responsive reporter assay used to screen metagenomic libraries for enzyme-encoding genes. Appl Environ Microbiol 76:7029–7035. doi:10.1128/AEM.00464-10
93. Wakimoto T, Egami Y, Nakashima Y, Wakimoto Y, Mori T, Awakawa T, Ito T, Kenmoku H, Asakawa Y, Piel J, Abe I (2014) Calyculin biogenesis from a pyrophosphate protoxin produced by a sponge symbiont. Nat Chem Biol 10:648. doi:10.1038/NCHEMBIO.1573
94. Wang GY, Graziani E, Waters B, Pan W, Li X, McDermott J, Meurer G, Saxena G, Andersen RJ, Davies J (2000) Novel natural products from soil DNA libraries in a streptomycete host. Org Lett 2:2401–2404
95. Wang K, Li G, Yu SQ, Zhang CT, Liu YH (2010) A novel metagenome-derived beta-galactosidase: gene cloning, overexpression, purification and characterization. Appl Microbiol Biotechnol 88:155–165. doi:10.1007/s00253-010-2744-7
96. Wang Q, Wu H, Wang A, Du P, Pei X, Li H, Yin X, Huang L, Xiong X (2010) Prospecting metagenomic enzyme subfamily genes for DNA family shuffling by a novel PCR-based approach. J Biol Chem 285:41509–41516. doi:10.1074/jbc.M110.139659
97. Wexler M, Johnston AW (2010) Wide host-range cloning for functional metagenomics. Methods Mol Biol 668:77–96. doi:10.1007/978-1-60761-823-2_5
98. Wilson MC, Mori T, Ruckert C, Uria AR, Helf MJ, Takada K, Gernert C, Steffens UA, Heycke N, Schmitt S, Rinke C, Helfrich EJ, Brachmann AO, Gurgui C, Wakimoto T, Kracht M, Crusemann M, Hentschel U, Abe I, Matsunaga S, Kalinowski J, Takeyama H, Piel J (2014) An environmental bacterial taxon with a large and distinct metabolic repertoire. Nature 506:58–62. doi:10.1038/nature12959
99. Woo SS, Jiang J, Gill BS, Paterson AH, Wing RA (1994) Construction and characterization of a bacterial artificial chromosome library of Sorghum bicolor. Nucleic Acids Res 22:4922–4931
100. Woodhouse JN, Fan L, Brown MV, Thomas T, Neilan BA (2013) Deep sequencing of non-ribosomal peptide synthetases and polyketide synthases from the microbiomes of Australian marine sponges. ISME J 7:1842–1851. doi:10.1038/ismej.2013.65
101. Yamanaka K, Reynolds KA, Kersten RD, Ryan KS, Gonzalez DJ, Nizet V, Dorrestein PC, Moore BS (2014) Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. Proc Natl Acad Sci USA 111:1957–1962. doi:10.1073/pnas.1319584111
102. Yao J, Fan XJ, Lu Y, Liu YH (2011) Isolation and characterization of a novel tannase from a metagenomic library. J Agric Food Chem 59:3812–3818. doi:10.1021/jf104394m

103. Yap WH, Li X, Soong TW, Davies JE (1996) Genetic diversity of soil microorganisms assessed by analysis of *hsp70* (*dnaK)* sequences. J Ind Microbiol 17:179–184
104. Ziemert N, Podell S, Penn K, Badger JH, Allen E, Jensen PR (2012) The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. PLoS One 7:e34064. doi:10.1371/journal.pone.0034064