Chapter 7
Challenges and Opportunities in Discovery of Secondary Metabolites Using a Functional Metagenomic Approach

Alinne L.R. Santana-Pereira and Mark R. Liles

Abstract Bioprospecting for natural products via a metagenomic approach has been highly successful for enzyme discovery, yet methodological challenges have inhibited discovery of diverse secondary metabolites from environmental metagenomes. In this chapter, we discuss metagenomic approaches to identify and/or express secondary metabolites encoded from environmental DNA. The application of next-generation sequencing techniques has generated enormous metagenomic sequence databases for polyketide synthases. Isolated biosynthetic pathways can be introduced into multiple heterologous hosts, with some hosts engineered for expression of specific pathways. The goal of tapping into the extant diversity of secondary metabolites encoded by environmental metagenomes is being enabled by a combination of approaches, including advances in NGS technology, cloning methods, high-throughput screening, development of improved heterologous hosts, and pathway engineering.

7.1 Introduction

Natural products have been an important source of bioactive compounds thorough the history of humanity. Mankind has always taken advantage of molecules synthesized by other forms of life, remarkably plants I order to treat diseases, obtain pigments, and other useful activities. In the modern days, natural products still play a role as important as they used to play back then, especially bioactive compounds of microbiological origin.

Since the discovery of penicillin in 1929, a myriad of drugs have been developed based solely or inspired by metabolites synthesized by microorganisms. All the main classes of antibiotics—tetracyclines, cephalosporins, aminoglycosides, and

A.L.R. Santana-Pereira • M.R. Liles (✉)
Department of Biological Sciences, Auburn University, Auburn, AL 36849, USA
e-mail: lilesma@auburn.edu

T.C. Charles et al. (eds.), *Functional Metagenomics: Tools and Applications*,
DOI 10.1007/978-3-319-61510-3_7

Table 7.1 Total number of drugs for some relevant medical applications and number of natural product derived drugs

Activity	Total number of compounds	Natural compounds
Antibacterial	118	78
Antifungal	29	4
Antiviral	110	51
Antiparasitic	14	9
Anticancer	128	87
Antidiabetic	37	25
Immunosuppressant	12	12
Hypocholesterolemic	13	11

macrolides—were isolated from microbes, mainly from actinomycetes (Berdy 2005). In addition, microbial metabolites have been shown to have all sorts of useful medical applications and have been used as anticancer drugs, antivirals, antiparasitic, immunosuppressive, lipid control agents, and antidiabetics, revolutionizing medicine (Li and Vederas 2009; Newman and Cragg 2012). The antibiotic activities also make microorganisms an important source of pesticides and herbicides and therefore are also important in agriculture (Berdy 2005). A list of the main uses of natural products can be seen in Table 7.1.

However the rate of discovery of novel bioactive compounds has decreased significantly in the past decade, mainly due to two reasons: the pharmaceutical environment lack of investment in natural product discovery and technical limitations in identifying new compounds. The lack of novelty in the compounds isolated due to high rates of re-isolation has led the pharmaceutical industry to turn to synthetic chemistry and drop its investments in screening endeavors (Li and Vederas 2009).

For a long time, natural product discovery has relied on culture-dependent methods for isolation of bioactive compounds; however, because of the high rate of redundant isolation, these approaches have been becoming less efficient as means of drug discovery (Tulp and Bohlin 2005). Less than 1% of soil microorganisms can be readily cultured in a laboratory using nutrient-rich media. The percentage of extant bacteria that can be recovered on media will vary from environment to environment, with higher culture counts achieved from eutrophic environments such as gut microbiomes. Novel culture methods that utilize lower nutrient levels and extracts from natural environments can be used to obtain cultures from bacterial phyla that are underrepresented in culture collections, particularly when culturing on solid media (Schoenborn et al. 2004). The development of in situ incubation culturing methods has allowed cultivation of diverse bacteria (Nichols et al. 2010), some of which produce novel secondary metabolites such as teixobactin capable of inhibiting the growth of multidrug-resistant pathogens (Ling et al. 2015). While cultivation methods are increasingly allowing access to a greater diversity of microbial genomes, metabolisms, and metabolites, this approach is inherently

limited in the degree to which it can sample environmental microbial diversity. A large percentage of microorganism cannot be maintained in culture (Nichols et al. 2010), and these approaches are time-consuming and require focused attention on specific strains to optimize them for natural product discovery. These approaches continue to yield novel phylogenetic and chemical diversity and are complementary to culture-independent approaches.

In this scenario, metagenomic libraries combined with next-generation sequencing (NGS) and heterologous expression techniques are a powerful tool to access the microbial "dark matter" and exploit this reservoir of bioactive compounds (Cragg and Newman 2013) while having a unique set of biases and limitations. In this chapter, we will discuss the use of a metagenomic approach in mining and expressing secondary metabolites from environmental samples and how these inherent biases and limitations can be mitigated and overcome to provide access to novel natural products.

7.2 Secondary Metabolites

Microorganisms are the main source of novel bioactive compounds, and the majority of these compounds are classified as secondary metabolites (Helfrich et al. 2014). Those are produced at the stationary phase and share characteristics such as (O'Brien and Wright 2011):

- They are not essential for growth nor reproduction.
- Their formation is dependent on growth conditions.
- They are produced as a group of closely related compounds.

The characteristics of secondary metabolites and their overwhelming diversity and complexity point toward a system that has long been adapted to produce molecules that are able to stop or slow down competitors in complex environments where the producer is under constant stress (Challis and Hopwood 2003) or as mediators of cross-species mutualism with other microorganisms, plants, and animals (O'Brien and Wright 2011).

Secondary metabolites can be classified in many different ways, according to biological activity, chemical structure, or genetic organization (Berdy 2005). From a metagenomic point of view, a classification based on biosynthetic gene clusters (BGCs) is more appropriate, since secondary metabolites will be dealt with in a genetic level. Thus, this classification can give information on genetic organization and active sites and even predict final product structures that can after guide screening and heterologous expression projects (Cimermancic et al. 2014; Weber et al. 2015).

There are several families of BGCs that can often share genes or even biosynthetic pathways leading to a remarkably complex network (Cimermancic et al. 2014). Here we discuss the main families of secondary metabolite BGCs.

Table 7.2 Bacteriocin classification

Classification	Characteristics	Group name
Class I		Lantibiotics
IA	Linear, rigid	
IB	Globular, flexible	
IC	Multicomponent	
Class II		Non-lantibiotics
IIa	Pediocin-like	
IIb	Miscellaneous	
IIc	Multicomponent	
Class III		Large heat stable
IIIa	Bacteriolytic	
IIIb	Non-bacteriolytic	
Class IV		Circular

Adapted from (Snyder and Worobo 2014)

7.3 Bacteriocins

Bacteriocins are ribosomally synthesized peptides with bactericidal activity, produced by a wide range of organisms, most notably lactic acid bacteria (LAB). Due to their bactericidal properties, they are widely used as preservatives in food industry (Snyder and Worobo 2014). Bacteriocins differentiate from traditional antibiotics as they are also produced during log phase and usually are narrow spectrum being active only against closely related strains (Zacharof and Lovitt 2012), even though some have a wider spectrum such as circular bacteriocins (Gabrielsen et al. 2014).

Bacteriocins are divided in four classes as shown in Table 7.2. They are usually coded in regulons containing genes for the bacteriocin peptide, secretion machinery (either transmembrane proteins or transporters), peptidases for bacteriocin activation, and an immunity peptide usually coded downstream the bacteriocin gene (Snyder and Worobo 2014). Regulon conformation shows significant variation from species to species, but despite the limited homology between genes of the BGC, they share features that make their identification possible such as multiple putative membrane-spanning domains and ATPases in the case of circular bacteriocins (van Belkum et al. 2011).

7.4 Non-ribosomal Peptide (NRP)

Non-ribosomal peptides are usually 2–48 amino acids in length and derive their diversity from the incorporation of more than 300 different amino acids including unusual ones such as D-amino acids and *N*-methylated amino acids (Du and Lou 2010). They work as mega-synthases, composed of at least three modules: (1) adenylation domain, recognizes and activates the entering amino acid; (2) peptidyl

carrier protein (PCP) domain, covalently bonds to the amino acid; and (3) condensation domain, catalyzes the elongation of the peptide; in addition, different domains may be present at the modules acting as tailoring enzymes (Rausch et al. 2007). The termination is performed by a thioesterase domain usually located at the end of the modular "assembly line"; however, complete NRP synthesis is not always linear and may involve intra- and intermolecular domain interactions (Sundlov et al. 2012).

7.5 Polyketides

Many secondary metabolites produced by microorganisms are classified as polyketides, compounds synthesized by polyketide synthases (PKSs) that show a remarkable diversity in structure and bioactivities. Currently, polyketide compounds have been developed for multiple applications such as chemotherapy (e.g., rapamycin), anticholesterol (e.g., lovastatin), and a number of antibiotics including erythromycin, rifamycin, and amphotericin B (Staunton and Weissman 2001). Recently, metagenomic studies have led to the characterization of polyketides with antibiotic activity against methicillin-resistant *Staphylococcus aureus* (MRSA) and vancomycin-resistant *Enterococcus faecalis* (Feng et al. 2011).

Polyketides are produced via successive decarboxylative Claisen thioester condensations of an activated acyl starter unit and extension by the addition of malonyl-CoA units or homologs (Hertweck 2009; Khosla et al. 2014). The elongation is performed by modules that are commonly composed of an acyl transferase (AT) that activates the acyl starter and loads it on the acyl carrier protein (ACP) and a β-ketoacylsynthase (KS) that catalyzes the carbon–carbon condensation between molecules in two adjacent modules. Modules can contain other optional sites like ketoreductases and/or dehydrases, while the newly synthesized compounds can be later processed by methyltransferases/esterases, for example, to further contribute to the diversity of the compounds produced (Hertweck 2009). PKSs can be divided into three main categories that differ significantly in the arrangement of synthetic modules. Type I PKSs use diverse modules in an assembly-line fashion to synthesize the final compound, and each module performs only one elongation step. Type II PKSs use the same ACP molecule throughout the entire elongation process, and the modules function in an iterative way to produce highly oxidized aromatic compounds (Hertweck 2009). Type III PKSs are homodimeric enzymes that act directly on the Acyl-CoA substrate in an iterative manner to catalyze the condensations without the need of an ACP module (Shen 2003).

PKS pathways can also be found associated with other biosynthetic pathways such as non-ribosomal peptide synthases (NRPS) and fatty acid synthases (FAS) (Ansari et al. 2004; Masschelein et al. 2013; Milano et al. 2013). Interestingly, all of these biosynthetic pathways share striking homology in their basic chemistry allowing their modules to interact to produce hybrid molecules that can incorporate unusual amino acids other than the 20 classically incorporated by ribosomes (Du et al. 2001).

Given the organization and nature of PKS pathways, they can be genetically engineered in the laboratory and serve as a tool for drug production and novel bioactive compound development (Siezen and Khayatt 2008). Nevertheless, the lack of knowledge of the extant diversity of PKS pathway diversity in environmental metagenomes has limited the polyketide diversity that can be obtained from pathway engineering approaches. There are many unknowns even for the existing polyketide synthases in understanding spatial interactions between enzymes, stereochemistry control, or the correct flux of reagents and products between the active sites (Khosla et al. 2014). Given these current constraints in our knowledge, the cloning and expression of environmental PKS pathways could provide insights into PKS/NRPS biochemistry and genetics as well as provide new bioactive compounds of medicinal value.

7.6 Shotgun Metagenomics

The development and lowering cost of next-generation sequencing (NGS) methods have enabled massively parallel sequencing and sampling of diverse microbial assemblages (Suenaga 2012). In a "shotgun metagenomic" approach, the isolated genomic DNA is sequenced directly, removing the need for library construction. The use of shotgun sequencing has the advantages of speed and lower cost compared to a cloning approach and has become a powerful tool in understanding and characterizing microbial assemblages in diverse environments (Suenaga 2012). The metagenomic sequences generated can provide information on taxonomic composition and functional capabilities for an entire microbiome.

However, the application of shotgun metagenomics to access biosynthetic pathways for secondary metabolite synthesis is limited due to short sequence read lengths and, in complex microbiomes, the inability to assemble large, contiguous genomic regions. While the use of shotgun sequencing has been effective in assembly of bacterial genomes in extreme environments, such as the Iron Mountain acidophilic biofilms (Tyson et al. 2004), the assembly of microbial genomes is very problematic in non-extreme habitats. For example, a massive degree of shotgun sequencing was used for a single sample of a native prairie soil from Iowa, resulting in 256 Gb of sequence data, but assembly de novo of this sequence dataset required the adoption of a digital normalization method, and the largest contig from this assembly was less than 10 kb (Howe et al. 2014). The assembly of NGS raw sequence data using de novo algorithms can also lead to the formation of chimeric contigs, especially from repetitive genomic regions (Treangen and Salzberg 2012). As an additional challenge, the degree of genome coverage for diverse microbial assemblages is not sufficient to fully assemble the number of genomes predicted in a given sample, especially for bacterial taxa that are at a low relative abundance (Ladoukakis et al. 2014).

Advances in NGS technology are addressing the challenges faced by shotgun metagenomics such as short read lengths and appropriate genome coverage.

Increasingly a significant bottleneck is the bioinformatics analyses of the sequence data obtained by this approach. While there are now powerful tools available for metagenomic analyses (Darling et al. 2014; Meyer et al. 2008), the in silico prediction of secondary metabolite biosynthetic pathways, even with an excellent bioinformatics tool like antiSMASH3.0 (Weber et al. 2015), will be ineffective without complete or nearly complete clusters. Despite these limitations, shotgun sequencing of metagenomes has provided interesting insights in microbiome taxonomical composition and dynamics, even though especial attention has to be given to biasing toward high abundance species (Brooks et al. 2015; Ladoukakis et al. 2014). Furthermore, targeting conserved domains of biosynthetic pathways, such as the B-ketosynthase (KS) domains of PKS pathways or the condensation (C) domains of NRPS pathways, using NGS sequencing of bar-coded domain amplicons or mining these from shotgun sequencing data, can provide information on the diversity of these pathways present in an environmental sample (Cacho et al. 2014; Foerstner et al. 2008; Wawrik et al. 2005). The further development of NGS technologies will undoubtedly one day afford the ability to generate more complete genomic sequences directly from environmental samples that can be mined for complete biosynthetic clusters.

7.7 Cloning Metagenomic DNA

The only current methodological route to accomplishing the goal of obtaining complete biosynthetic clusters from metagenomic sources, without a cultivation step, is in the direct cloning of genomic fragments. Prior to constructing a metagenomic library, it is important to first select the source DNA, isolation method, vector, insert size, screening method, and expression host so that the output of the screening match the desired project goals (Kakirde et al. 2010). In particular, the host and cloning vector are critical for success. Different hosts have different metabolic capabilities and can be more or less feasible to culture and transform. Whereas the *E. coli* strain DH10B is readily transformable by electroporation, its capacity for secondary metabolite synthesis is marginal compared to *Streptomyces coelicolor*, which would conversely make a very poor host for constructing libraries. Because of these host differences in the relative degree of transformability and metabolic capacity, the development of broad host range shuttle vectors permits the construction of a metagenomic library in *E. coli* followed by its transfer to another heterologous host (Aakvik et al. 2009; Courtois et al. 2003; Craig et al. 2010; Kakirde et al. 2010). The resultant libraries may be conjugally transferred into a recipient host and can replicate autonomously as a plasmid (Kakirde et al. 2010) or can be integrated into the recipient host chromosome (Heil et al. 2012).

Generally, when the insert a vector can accommodate is larger, this reduces transformation efficiency and makes library construction more difficult. A clear exception to this rule is in the use of fosmid libraries (see below) which enhances transformation efficiency using phage packaging. On the other hand, vectors with

Table 7.3 Average sizes of biosynthetic gene clusters based on the curated AntiSMASH database (Blin et al. 2016)

Metabolite Type	Average BGC size (kb)
Bacteriocin	14.2
Polyketides:	
PKS I	61.9
PKS II	49.3
PKS III	44.9
NRPS	63.3

high efficiency of transformation but smaller insert sizes will fail to include the contiguity of the targeted pathways and will therefore be limiting for novel pathway and secondary metabolite discovery (Table 7.3).

7.8 High Molecular Weight DNA Cloning

When recovering PKS pathways from the environment for future heterologous expression, insert size is a key variable (Kakirde et al. 2010). The characteristic modular architecture of Type I polyketide synthases and NRPS's pathway are often coded as large assembly line pathways more than 20 kb long with some exceeding 100 kb (Fu et al. 2012). Even though the cluster size is smaller for Type II PKSs, which are commonly polycyclic, a large insert is still desirable in order to accommodate genes for tailoring enzymes that are not present in the core PKS pathway (Hertweck 2009). Therefore, metagenomic libraries have to be able to overcome insert size constraints in order to successfully clone complete biosynthetic gene clusters. Common large-insert cloning strategies for the construction of metagenomic libraries include the use of cosmid, fosmid, or bacterial artificial chromosome (BAC) vectors.

Cosmid vectors are plasmids containing the *cos* site of λ phage and an origin of replication allowing ligated sequences to be packed in λ phages and transduced into the host where the insert is maintained as a plasmid (Haley 1988). Because of the small size of the plasmid and in vitro packaging restraints, this technique can only successfully transform the host with inserts of usually 40 kb. The lack of controlling mechanisms on plasmid copy number and recombination between plasmids often decrease the stability of cosmid vectors. Functional screening of a cosmid metagenomic library using six different *Proteobacteria* hosts showed that not all hosts performed equally well and that phenotypes were rarely observed twice in different hosts (Craig et al. 2010). Thus, their work stresses the need of a vector system that can be used along different surrogate hosts. Cosmid vectors perform well in a narrow range of hosts, but its range can be widened via genetic manipulation. Broad host-range cosmid vectors pJC8 and pJC24 were designed by cleverly adding the Gateway® homology site *attL1* and *attL2* to the cosmid along with the RK-2 *otiT* (Cheng et al. 2014). This cosmid vectors were able to transfer the insert to a Gateway® plasmid via recombination, and the RK-2 *oriT* allows the transfer of the

plasmid via conjugation to a number of other hosts including Gram-positive bacteria and even yeasts like *Saccharomyces cerevisiae*. These enhances in cosmid vectors increase the efficiency of screening techniques and allow higher rates of positive hits through heterologous expression.

Fosmid vectors are based on the same principles of cosmid vectors but use an F-plasmid origin of replication. F-plasmids have advantages such as existing as single copy plasmids and having a tight fertility control which decreases host intolerance to repetitive sequences and also limits homologous recombination between copies. Therefore, they are more stable and desirable when constructing genomic libraries (Quail et al. 2011) and have been used for many studies of microbial ecology and natural product discovery (Felczykowska et al. 2014; Parsley et al. 2011). Nevertheless, fosmid vectors have the same restrictions on host range faced by cosmid vectors, but similar approaches can be used to create broad host-range fosmid vectors. Broad host-range fosmid vectors were designed by adding RK-2 *oriV* and an *oriT* to the commercial fosmid vector pCC1FOS (Aakvik et al. 2009). By doing so, the fosmid vector can be transferred to other hosts from *E. coli* via conjugation, and the additional *oriV* allows for plasmid copy number regulation in hosts other than *E. coli* that doesn't recognize the original *ori2* site.

Even though cosmid or fosmid libraries provide a more efficient method to generate large-insert metagenomic libraries, because of the necessity to use phage packaging, these vector systems are inherently limited in size and are ineffective at recovering larger biosynthetic clusters, such as Type I PKS pathways. In contrast, BAC vectors can be used to stably clone insert DNA fragments as large as 300 kbp (Shizuya et al. 1992), which is sufficient to contain the largest of known biosynthetic clusters. Recently, metagenomic libraries containing inserts exceeding 100 kbp on average have been reported (Monsma et al. manuscript in preparation). These vectors provide means to have inserts big enough to accommodate complete secondary metabolite pathways, but because the transformation efficiency is lower, they are more labor intensive and therefore are not a suitable and/or preferred vector for all applications (Quail et al. 2011).

The first BAC vectors, such as pBELOBAC11, were kept as single copy plasmids in the host cell, granting high stability but decreasing the amount of product obtained by heterologous expression. Therefore, larger scale *E. coli* cultures were necessary (Wild et al. 2002). To increase plasmid yield and also potentially benefit heterologous expression, BAC vectors were engineered to have an inducible copy system in which the BAC vector carries an additional origin of replication *oriV* from RK2 that is dependent upon the expression of the TrfA protein. By expressing the *trfA* gene under the control of an arabinose-dependent promoter, the BAC vector is then conditionally copy-induced only in the presence of arabinose, permitting stable maintenance of libraries in single copy and then using copy induction for screening or DNA isolation purposes (Kakirde et al. 2010; Wild et al. 2002). The broad host-range vector pGNS-BAC has been shown to replicate in diverse Gram-negative bacteria and can be conjugally transferred from *E. coli* (Kakirde et al. 2010). Therefore, this vector is well suited for metagenomic library construction and screening,

especially if the DNA isolated from an environmental sample is biased toward Gram-negative bacteria (Liles et al. 2003). Other derivatives of this vector have been generated that have the capacity to be introduced and stably integrated into Gram-positive hosts as well (Monsma et al. manuscript in preparation).

7.9 Functional Screening of Metagenomic Libraries

The research groups of Prof. Jo Handelsman and Prof. Robert Goodman in their seminal paper on soil metagenomics demonstrated that clone libraries could be screened for desired bioactivities by expressing metagenomic clones within an *E. coli* host and resulted in discovery of various enzymatic activities (Rondon et al. 2000). Since this method had been proposed, different groups have turned to functional screening to identify clones exhibiting a desired activity. High-throughput function-based screenings often rely on indicators to rapidly identify the activity of interest, such as inhibition halos, color change in the media or colonies, and degradation of media (Coughlan et al. 2015).

A high-throughput screening method for antibiotic activity was developed based on a sequential two antibiotic selection schemes (Brady and Clardy 2000). First clones are selected for kanamycin resistance in media containing the antibiotic. Then antibiotic activity is verified by an overlay of kanamycin-resistant *Bacillus subtilis*. Positive hits are recovered directly from the assay plate and then streaked in media containing ampicillin killing *B. subtilis*. This screening method have led to the characterization of a number of compounds including several long-chain *N*-acyl amino acid antibiotics (Brady et al. 2004; Brady and Clardy 2000), isocyanide antibiotic (Brady and Clardy 2005), antibiotic polyketides (Craig et al. 2009), novel metalloproteases, serine proteases, and lipolytic compounds (Iqbal et al. 2014).

One interesting way to track activities of interest that may not be readily accessible is the use of reporter genes. *S*ubstrate-*i*nduced *g*ene *ex*pression (SIGEX) screening allows the screening of catabolic genes by using a green fluorescent protein (GFP) as a reporter. Since the expression of catabolic pathways is usually triggered by the presence of the substrate, it was rationalized that clones containing catabolic pathways would be activated in the presence of specific substrates (Uchiyama et al. 2005). The vector has *gfp* genes that are translationally dependent on the expression of the eDNA. Thus, if a catabolic pathway is activated in a clone, the colonies get green fluorescence that can be sorted by fluorescence-activated cell sorting for high-throughput screening.

However the technique is unable to identify constitutively expressed catabolic genes and fails to identify genes inserted in the opposite orientation of *gfp* (Yun and Ryu 2005). PIGEX is an improvement to SIGEX, where the expression of *gfr* is triggered by the presence of a particular catabolic product (Uchiyama and Miyazaki 2010). By doing so, PIGEX is not sensitive to insert orientation and is able to detect constitutively expressed catabolic pathways. In the same work, the authors reported the discovery of three novel amidases (Uchiyama and Miyazaki 2010), and diverse groups have also been successful in isolating catabolic genes of interest (Coughlan et al. 2015) (Fig. 7.1).

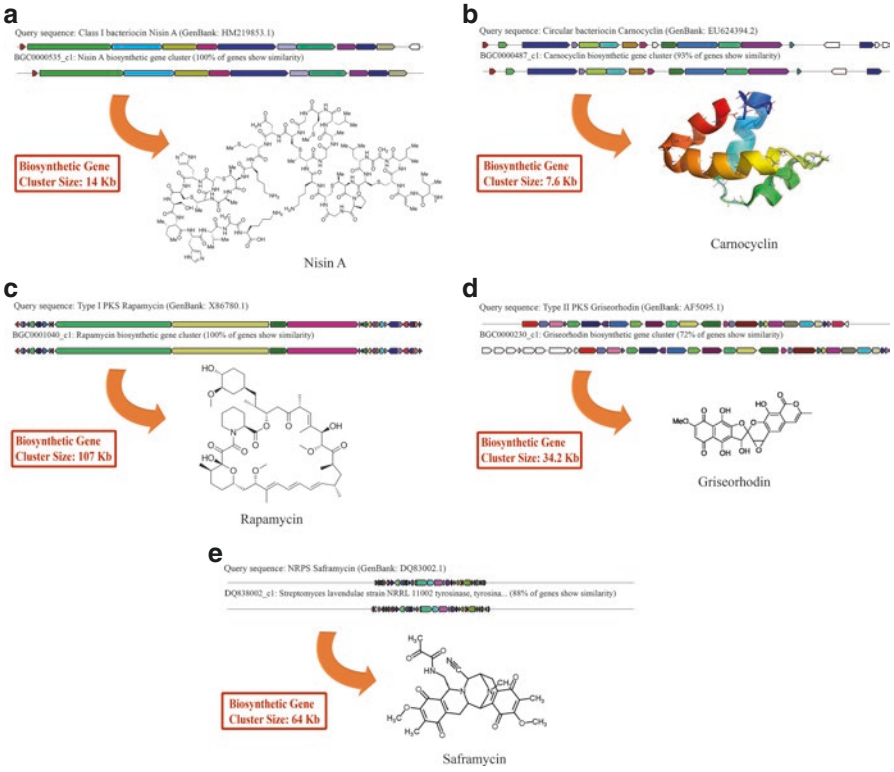

Fig. 7.1 Annotation of biosynthetic gene clusters (BCGs) of the main secondary metabolite groups using the program antiSMASH (Weber et al. 2015) using BGC sequences obtained from GenBank. Compounds from each group were chosen along with their chemical structures, illustrating the chemical diversity of secondary metabolites. (**a**) Nissin A (Class I bacteriocin). (**b**) Carnocyclin circular bacteriocin, (**c**) rapamycin (Type I PKS), (**d**) griseorhodin (Type II PKS), and (**e**) saframycin (NRPS peptide)

An interesting way to enrich clones containing not only PKS but also NRPS pathways is to utilize the biochemical characteristics of these pathways (Charlop-Powers et al. 2013). The ability of ACP/PCP to bind to the growing polymer chain relies on a phosphopantetheine prostetic group added by PPTases as part of the post-translational modification of the carrier proteins. In the absence of phosphopantetheine, neither ACP nor PCP is functional. Iron uptake is also dependent on NRPS-derived peptides, and in *E. coli* this function is performed by the siderophore enterobactin. The enterobactin gene cluster contains the PPTase EntD which is crucial to the peptide biosynthesis, thus *entD* mutants cannot grow in iron-limiting conditions. Therefore, *entD* mutants could have the phenotype reverted and be viable in low limiting conditions if the insert cloned coded for PPTase which is commonly associated with PKS and NRPS pathways. This approach provides a fast way to select promising clones within a metagenomic library in an easy, fast and low

cost manner. However, there are still limitations intrinsic to heterologous expression, especially when using an *E. coli* host as proposed by Charlop-Powers et al. (2013).

Nevertheless, there are several limitations on heterologous expression of environmental DNA that can result in a subset of the cloned genes being transcribed and translated or their protein product(s) being modified and active (National Research Council (US) Committee on Metagenomics, 2007). There is also evidence for an under-abundance of strong rpoD consensus sequences within a human microbiome sample, suggesting that the loss of AT-rich sequences may be due to bias in favor or less actively transcribed genes (Lam and Charles 2015). Conversely, expression of a heterologous sigma factor, for example, from *Acidobacteria* or *Lactobacillus* taxa, has been shown to enhance transcription from metagenomic clones (Gaida et al. 2015; Sabree et al. 2006).

Therefore, because of these potential issues with transcription and translation of metagenome-derived genes in a particular heterologous host, there have been efforts to expand the range of heterologous hosts for metagenomic library expression (Courtois et al. 2003; Craig et al. 2010). With specific regard to secondary metabolite producers, hosts such as *Streptomyces* spp. and *Pseudomonas* spp. and other *Proteobacteria* taxa have been a particular emphasis. Development of new heterologous hosts that better reflect the phylogenetic origin of cloned biosynthetic clusters and further engineering of existing hosts will be important for the field of functional metagenomics for the goal of novel secondary metabolite discovery.

Expression of eDNA can also be influenced by the primary metabolism of the host; thus, the existing host metabolic pathways and their interactions have to be taken in consideration when inducing expression of an environmental biosynthetic pathway. PKS pathways are a good example of these interactions. PKS pathways and fatty acid synthesis show remarkable similarities and are thought to have a common evolutionary origin. Both pathways can share the precursors acetyl-CoA; thus, high expression of one metabolic pathway could inhibit the other by depleting the pool of acetyl-CoA in the cell (Cronan and Thomas 2009). In fact, compounds that block fatty acid metabolism are shown to enhance pigment production in *Streptomyces coelicolor*, demonstrating how primary and secondary metabolisms are linked together and stressing the significance of competition over precursors between metabolic pathways (Craney et al. 2012).

To overcome this limitations in secondary metabolite expression, different strains of *S. coelicolor* have been genetically engineered, with the strain M145 having deletions in four endogenous antibiotic gene clusters, thereby removing competing sinks for carbon and nitrogen, increasing precursor availability, and enhancing heterologous expression (Gomez-Escribano et al. 2012). The pathway deletions also help by simplifying the metabolic profile of *S. coelicolor*, making mass spectrometry analysis easier with a lower metabolite background and facilitating the identification of compounds encoded by the exogenous insert.

Another path to enhancing secondary metabolite expression in a host is to manipulate its transcriptional apparatus, especially when screening for antibiotics. To increase *S. coelicolor* antibiotic expression, point mutations were introduced in

the *rpoB* and *rpsL* genes that encode the RNA polymerase β-subunit and the ribosomal protein S12, respectively. The mutations conferred increased host resistance to rifampicin, streptomycin, and paromomycin and increased antibiotic production 30- to 40-fold when exogenous antibiotic genes were cloned into the modified strain (Gomez-Escribano and Bibb 2014; Hu et al. 2002).

7.10 Sequence-Based Screening of Metagenomic Libraries

Function-based screening has been an important way to mine metagenomic libraries for molecules and/or processes of interest; however, one of its main limitations is redundant isolation (Tulp and Bohlin 2005). Redundant isolation can be overcome by prior bioinformatics analyses of inserts selected by sequence-based screening, making the process more efficient (Chang and Brady 2013a). By first identifying a specific gene target, this method allows the identification and characterization of a pathway prior to functional screening, thereby reducing the number of clones subjected to functional screening, saving time and effort, and increasing the potential for heterologous expression (Culligan et al. 2014). Secondary metabolite pathways can be targeted by PCR-, hybridization-, or homology-based methods in order to identify pathway-containing clones, utilizing conserved domains and/or regions.

PKS clusters represent an ideal target for pathways encoding diverse secondary metabolites, given their well-characterized architecture and conserved domains and their presence in diverse marine and terrestrial organisms (Fieseler et al. 2007; Ginolhac et al. 2004; Muller et al. 2015). A minimal polyketide synthase module is composed of an acyl-transferase (AT), a ketosynthase (KS), and an acyl-carrying protein, with the ketosynthase being the most conserved element of the module since it catalyzes the carbon condensation in polyketide synthesis (Khosla et al. 2014). Fortunately, there are conserved sequences within KS domains, and their DNA sequence is an ideal target for PCR amplification (Wawrik et al. 2005). Therefore, clones containing polyketide pathway can be identified by homology-based approaches by targeting KS and other conserved domains (Banik and Brady 2010; Chang and Brady 2013b). Nevertheless, there is great sequence diversity among KS domain sequences, and it is probable that there are KS domains present within environmental metagenomes that would not be amplified using the available degenerate primer sets, resulting in a biased set of sequences most similar to those from previously characterized pathways.

Another powerful approach is the utilization of next-generation sequence and bioinformatics approaches to mine for conserved motifs. In this approach, all of the clones in the library are sequenced, and computational analysis tools are used to identify clones containing the target sequences. The availability of a complete PKS pathway sheds light on modular organization of the pathway, and this can be used for in silico prediction of polyketide structure, even though there are limitations in predicting the final structure based only on sequence data (Hertweck 2009; Zerikly and Challis 2009). In addition, PKS domain sequences cluster together in func-

tional groups, and therefore, a molecular phylogenetic analysis can be used as a guide to identify pathways that encode structurally related polyketide products (Metsa-Ketela et al. 2002). The use of NGS as a means of identifying NRPS, PKS, and other clusters has now been demonstrated using a large-insert soil metagenomic BAC library using a pooled strategy (rows, columns, and plates) and resulting in the discovery of diverse and novel pathways (Monsma et al. manuscript in preparation).

7.11 Future Perspectives

Secondary metabolites produced by as-yet-uncultured microorganisms could be a rich resource for bioactive compounds. Metagenomic approaches are one of the powerful tools that can be used to unravel the diversity of secondary metabolites present in nature, and the combination of tools now available is finally starting to bear fruit.

The rapid development of NGS platforms provides a means to sequence large amounts of DNA with increasing throughput, accuracy, and longer read lengths. Despite the rapid technological evolution of NGS platforms, read lengths currently are not sufficient to allow shotgun metagenomics to access larger biosynthetic pathways, and there is increasing computational demands to handle this deluge of sequencing data. The combination of new developments in NGS technology and bioinformatics approaches promises to make it one day possible to access more complete genomes and pathways directly from an environmental sample to mine these for their secondary metabolite encoding potential.

The advances in NGS technology and data processing have also revolutionized the way that screening is performed. The use of PCR to perform gene-targeted screens is still in use, but increasingly, NGS allows the discovery of a greater diversity of gene sequences and avoids the primary disadvantage of PCR in that it is more likely to identify genes that are most similar to previously known sequences. Thus, NGS is a much more desirable approach for gene targeting as bioinformatics analyses of the sequences may reveal novel genes and/or pathways configuration that were cryptic to PCR techniques. As NGS becomes cheaper and easier to process, more and more research projects are expected to shift from PCR-based screening to sequencing-based screening. Furthermore, the use of NGS as a first step in screening a library can avoid the significant biases inherent in functional screening, so that identified pathways of interest can then be selectively targeted for engineering toward enhanced expression.

When expressing secondary metabolite pathways for desired bioactivities, it is very important that environmental genes, operons, and/or pathways have their integrity preserved when cloning. For many biosynthetic clusters, such as Type I PKS, NRPS, or hybrid pathways, the use of BAC vectors constitutes the best option. Not only can BAC vectors carry inserts of greater than 100 kb, there are new derivatives engineered for copy induction, inducible expression of cloned inserts, and introduc-

tion into broad hosts. These features can be important in reducing host metabolic stress and allowing the cloning of potentially toxic pathways. These advances have set the stage for the isolation and expression of entire biosynthetic pathways that have been previously inaccessible, now permitting the mining of previously unexplored reservoirs of bioactive compounds. Complete pathway cloning is an important advance in the bioprospecting of active compounds via metagenomics, combined together with pathway discovery via NGS (Monsma et al. manuscript in preparation).

Despite these advances in cloning and NGS screening methods, the heterologous expression of cloned pathways has serious limitations related to the ability of the host to correctly and efficiently express the foreign DNA inserted. Some hosts have stricter promoter recognition and deviant codon usage or lack the metabolic ability to synthesize a particular product. Heterologous hosts can be successfully genetically engineered to increase their efficiency for secondary metabolite pathway expression, and there is value in attaining a deeper understanding of host physiology in order to rationally address expression limitations to facilitate better expression systems.

Given the vast metagenomic diversity represented in clone libraries, there is also a benefit in utilizing multiple heterologous hosts, including both Gram-negative hosts such as *E. coli* and other bacteria phylogenetically related to the origin of cloned genes and selected to have desired characteristics for expression and metabolic support. There is, therefore, great value in continued pursuit of novel microbial cultures for their metabolites and their use as hosts, which are representative of the microbial "dark matter" previously tapped into purely by culture-independent methods (Rinke et al. 2013).

Realizing the promise of functional metagenomics for secondary metabolite discovery, as articulated 17 years ago by Jo Handelsman and colleagues (Handelsman et al. 1998), has required concomitant development of many tool sets. Only through the wise combination of methods to clone, identify, and express the diverse pathways encoded by environmental microorganisms can this dream be fulfilled.

References

Aakvik T, Degnes KF, Dahlsrud R, Schmidt F, Dam R, Yu L et al (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(2):149–158. doi:10.1111/j.1574-6968.2009.01639.x

Ansari MZ, Yadav G, Gokhale RS, Mohanty D (2004) NRPS-PKS: a knowledge-based resource for analysis of NRPS/PKS megasynthases. Nucleic Acids Res 32(Web Server Issue):W405–W413. doi:10.1093/nar/gkh359

Banik JJ, Brady SF (2010) Recent application of metagenomic approaches toward the discovery of antimicrobials and other bioactive small molecules. Curr Opin Microbiol 13(5):603–609. doi:10.1016/j.mib.2010.08.012

Berdy J (2005) Bioactive microbial metabolites. J Antibiot (Tokyo) 58(1):1–26. doi:10.1038/ja.2005.1

Blin K, Medema MH, Kottmann R, Lee SY, Weber T (2016). The antiSMASH database, a comprehensive database of microbial secondary metabolite biosynthetic gene clusters. Nucleic acids research, gkw960. doi:10.1093/nar/gkw960

Brady SF, Clardy J (2000) Long-chain N-acyl amino acid antibiotics isolated from heterologously expressed environmental DNA. J Am Chem Soc 122(51):12903–12904. doi:10.1021/ja002990u

Brady SF, Clardy J (2005) Cloning and heterologous expression of isocyanide biosynthetic genes from environmental DNA. Angew Chem Int Ed Engl 44(43):7063–7065. doi:10.1002/anie.200501941

Brady SF, Chao CJ, Clardy J (2004) Long-chain N-acyltyrosine synthases from environmental DNA. Appl Environ Microbiol 70(11):6865–6870. doi:10.1128/AEM.70.11.6865-6870.2004

Brooks JP, Edwards DJ, Harwich MD Jr, Rivera MC, Fettweis JM, Serrano MG et al (2015) The truth about metagenomics: quantifying and counteracting bias in 16S rRNA studies. BMC Microbiol 15(1):66. doi:10.1186/s12866-015-0351-6

Cacho RA, Tang Y, Chooi YH (2014) Next-generation sequencing approach for connecting secondary metabolites to biosynthetic gene clusters in fungi. Front Microbiol 5:774. doi:10.3389/fmicb.2014.00774

Challis GL, Hopwood DA (2003) Synergy and contingency as driving forces for the evolution of multiple secondary metabolite production by Streptomyces species. Proc Natl Acad Sci U S A 100(Suppl 2):14555–14561. doi:10.1073/pnas.1934677100

Chang F-YY, Brady SF (2013a) Discovery of indolotryptoline antiproliferative agents by homology-guided metagenomic screening. Proc Natl Acad Sci U S A 110(7):2478–2483. doi:10.1073/pnas.1218073110

Chang FY, Brady SF (2013b) Discovery of indolotryptoline antiproliferative agents by homology-guided metagenomic screening. Proc Natl Acad Sci U S A 110(7):2478–2483. doi:10.1073/pnas.1218073110

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(1):138–143. doi:10.1021/cb3004918

Chatterjee C, Paul M, Xie L, van der Donk WA (2005) Biosynthesis and mode of action of lantibiotics. Chem Rev 105(2):633–684. doi:10.1021/cr030105v

Cheng J, Pinnell L, Engel K, Neufeld JD, Charles TC (2014) Versatile broad-host-range cosmids for construction of high quality metagenomic libraries. J Microbiol Methods 99:27–34. doi:10.1016/j.mimet.2014.01.015

Cimermancic P, Medema MH, Claesen J, Kurita K, Wieland Brown LC, Mavrommatis K et al (2014) Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. Cell 158(2):412–421. doi:10.1016/j.cell.2014.06.034

Coughlan LM, Cotter PD, Hill C (2015) Biotechnological applications of functional metagenomics in the food and pharmaceutical industries. Front Microbiol 6:672. doi:10.3389/fmicb.2015.00672

Courtois S, Cappellano CM, Ball M, Francou FX, Normand P, Helynck G et al (2003) Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. Appl Environ Microbiol 69(1):49–55. doi:10.1128/AEM.69.1.49-55.2003

Cragg GM, Newman DJ (2013) Natural products: a continuing source of novel drug leads. Biochim Biophys Acta 1830(6):3670–3695. doi:10.1016/j.bbagen.2013.02.008

Craig JW, Chang F-Y, Brady SF (2009) Natural products from environmental DNA hosted in Ralstonia metallidurans. ACS Chem Biol 4(1):23–28. doi:10.1021/cb8002754

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(5):1633–1641. doi:10.1128/AEM.02169-09

Craney A, Ozimok C, Pimentel-Elardo SM, Capretta A, Nodwell JR (2012) Chemical perturbation of secondary metabolism demonstrates important links to primary metabolism. Chem Biol 19(8):1020–1027. doi:10.1016/j.chembiol.2012.06.013

Cronan JE, Thomas J (2009) Bacterial fatty acid synthesis and its relationships with polyketide synthetic pathways. Methods Enzymol 459:395–433. doi:10.1016/S0076-6879(09)04617-5

Culligan EP, Sleator RD, Marchesi JR, Hill C (2014) Metagenomics and novel gene discovery: promise and potential for novel therapeutics. Virulence 5(3):399–412. doi:10.4161/viru.27208

Darling AE, Jospin G, Lowe E, Matsen FA 4th, Bik HM, Eisen JA (2014) PhyloSift: phylogenetic analysis of genomes and metagenomes. PeerJ 2:e243. doi:10.7717/peerj.243

Du L, Lou L (2010) PKS and NRPS release mechanisms. Nat Prod Rep 27(2):255–278. doi:10.1039/b912037h

Du L, Sanchez C, Shen B (2001) Hybrid peptide-polyketide natural products: biosynthesis and prospects toward engineering novel molecules. Metab Eng 3(1):78–95. doi:10.1006/mben.2000.0171

Ennahar S, Sashihara T, Sonomoto K, Ishizaki A (2000) Class IIa bacteriocins: biosynthesis, structure and activity. FEMS Microbiol Rev 24(1):85–106. doi:10.1016/S0168-6445(99)00031-5

Felczykowska A, Dydecka A, Bohdanowicz M, Gasior T, Sobon M, Kobos J et al (2014) The use of fosmid metagenomic libraries in preliminary screening for various biological activities. Microb Cell Fact 13(1):105. doi:10.1186/s12934-014-0105-4

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 U S A 108(31):12629–12634. doi:10.1073/pnas.1103921108

Fieseler L, Hentschel U, Grozdanov L, Schirmer A, Wen G, Platzer M et al (2007) Widespread occurrence and genomic context of unusually small polyketide synthase genes in microbial consortia associated with marine sponges. Appl Environ Microbiol 73(7):2144–2155. doi:10.1128/AEM.02260-06

Foerstner KU, Doerks T, Creevey CJ, Doerks A, Bork P (2008) A computational screen for type I polyketide synthases in metagenomics shotgun data. PLoS One 3(10):e3515. doi:10.1371/journal.pone.0003515

Foulston LC, Bibb MJ (2010) Microbisporicin gene cluster reveals unusual features of lantibiotic biosynthesis in actinomycetes. Proc Natl Acad Sci U S A 107(30):13461–13466. doi:10.1073/pnas.1008285107

Fu J, Bian X, Hu S, Wang H, Huang F, Seibert PM et al (2012) Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. Nat Biotechnol 30(5):440–446. doi:10.1038/nbt.2183

Gabrielsen C, Brede DA, Nes IF, Diep DB (2014) Circular bacteriocins: biosynthesis and mode of action. Appl Environ Microbiol 80(22):6854–6862. doi:10.1128/AEM.02284-14

Gaida SM, Sandoval NR, Nicolaou SA, Chen Y, Venkataramanan KP, Papoutsakis ET (2015) Expression of heterologous sigma factors enables functional screening of metagenomic and heterologous genomic libraries. Nat Commun 6:7045. doi:10.1038/ncomms8045

Ginolhac A, Jarrin C, Gillet B, Robe P, Pujic P, Tuphile K et al (2004) Phylogenetic analysis of polyketide synthase I domains from soil metagenomic libraries allows selection of promising clones. Appl Environ Microbiol 70(9):5522–5527. doi:10.1128/AEM.70.9.5522-5527.2004

Gomez-Escribano JP, Bibb MJ (2014) Heterologous expression of natural product biosynthetic gene clusters in Streptomyces coelicolor: from genome mining to manipulation of biosynthetic pathways. J Ind Microbiol Biotechnol 41(2):425–431. doi:10.1007/s10295-013-1348-5

Gomez-Escribano JP, Song LJ, Fox DJ, Yeo V, Bibb MJ, Challis GL (2012) Structure and biosynthesis of the unusual polyketide alkaloid coelimycin P1, a metabolic product of the cpk gene cluster of Streptomyces coelicolor M145. Chem Sci 3(9):2716–2720. doi:10.1039/c2sc20410j

Haley JD (1988) Cosmid library construction. Methods Mol Biol 4:257–283. doi:10.1385/0-89603-127-6:257

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(10):R245–R249. doi:10.1016/S1074-5521(98)90108-9

Heil JR, Cheng J, Charles TC (2012) Site-specific bacterial chromosome engineering: PhiC31 integrase mediated cassette exchange (IMCE). J Vis Exp (61). doi:10.3791/3698

Helfrich EJ, Reiter S, Piel J (2014) Recent advances in genome-based polyketide discovery. Curr Opin Biotechnol 29:107–115. doi:10.1016/j.copbio.2014.03.004

Hertweck C (2009) The biosynthetic logic of polyketide diversity. Angew Chem Int Ed Engl 48(26):4688–4716. doi:10.1002/anie.200806121

Hill AM (2006) The biosynthesis, molecular genetics and enzymology of the polyketide-derived metabolites. Nat Prod Rep 23(2):256–320. doi:10.1039/b301028g

Hopwood DA (1997) Genetic contributions to understanding polyketide synthases. Chem Rev 97(7):2465–2498. doi:10.1021/cr960034i

Howe AC, Jansson JK, Malfatti SA, Tringe SG, Tiedje JM, Brown CT (2014) Tackling soil diversity with the assembly of large, complex metagenomes. Proc Natl Acad Sci U S A 111(13):4904–4909. doi:10.1073/pnas.1402564111

Hu H, Zhang Q, Ochi K (2002) Activation of antibiotic biosynthesis by specified mutations in the rpoB gene (encoding the RNA polymerase beta subunit) of Streptomyces lividans. J Bacteriol 184(14):3984–3991. doi:10.1128/JB.184.14.3984-3991.2002

Iqbal HA, Craig JW, Brady SF (2014) Antibacterial enzymes from the functional screening of metagenomic libraries hosted in Ralstonia metallidurans. FEMS Microbiol Lett 354(1):19–26. doi:10.1111/1574-6968.12431

Kakirde KS, Parsley LC, Liles MR (2010) Size does matter: application-driven approaches for soil metagenomics. Soil Biol Biochem 42(11):1911–1923. doi:10.1016/j.soilbio.2010.07.021

Khosla C, Herschlag D, Cane DE, Walsh CT (2014) Assembly line polyketide synthases: mechanistic insights and unsolved problems. Biochemistry 53(18):2875–2883. doi:10.1021/bi500290t

Ladoukakis E, Kolisis FN, Chatziioannou AA (2014) Integrative workflows for metagenomic analysis. Front Cell Dev Biol 2:70. doi:10.3389/fcell.2014.00070

Lal R, Kumari R, Kaur H, Khanna R, Dhingra N, Tuteja D (2000) Regulation and manipulation of the gene clusters encoding type-I PKSs. Trends Biotechnol 18(6):264–274. doi:10.1016/S0167-7799(00)01443-8

Lam KN, Charles TC (2015) Strong spurious transcription likely contributes to DNA insert bias in typical metagenomic clone libraries. Microbiome 3:22. doi:10.1186/s40168-015-0086-5

Li JW, Vederas JC (2009) Drug discovery and natural products: end of an era or an endless frontier? Science 325(5937):161–165. doi:10.1126/science.1168243

Li L, Deng W, Song J, Ding W, Zhao QF, Peng C et al (2008) Characterization of the saframycin A gene cluster from Streptomyces lavendulae NRRL 11002 revealing a nonribosomal peptide synthetase system for assembling the unusual tetrapeptidyl skeleton in an iterative manner. J Bacteriol 190(1):251–263. doi:10.1128/JB.00826-07

Liles MR, Manske BF, Bintrim SB, Handelsman J, Goodman RM (2003) A census of rRNA genes and linked genomic sequences within a soil metagenomic library. Appl Environ Microbiol 69(5):2684–2691. doi:10.1128/Aem.69.5.2684-2691.2003

Ling LL, Schneider T, Peoples AJ, Spoering AL, Engels I, Conlon BP et al (2015) A new antibiotic kills pathogens without detectable resistance. Nature 517(7535):455–459. doi:10.1038/nature14098

Masschelein J, Mattheus W, Gao LJ, Moons P, Van Houdt R, Uytterhoeven B et al (2013) A PKS/NRPS/FAS hybrid gene cluster from Serratia plymuthica RVH1 encoding the biosynthesis of three broad spectrum, zeamine-related antibiotics. PLoS One 8(1):e54143. doi:10.1371/journal.pone.0054143

Metsa-Ketela M, Halo L, Munukka E, Hakala J, Mantsala P, Ylihonko K (2002) Molecular evolution of aromatic polyketides and comparative sequence analysis of polyketide ketosynthase and 16S ribosomal DNA genes from various streptomyces species. Appl Environ Microbiol 68(9):4472–4479. doi:10.1128/Aem.68.9.4472-4479.2002

Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M et al (2008) The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. BMC Bioinformatics 9(1):386. doi:10.1186/1471-2105-9-386

Milano T, Paiardini A, Grgurina I, Pascarella S (2013) Type I pyridoxal 5′-phosphate dependent enzymatic domains embedded within multimodular nonribosomal peptide synthetase and polyketide synthase assembly lines. BMC Struct Biol 13(1):26. doi:10.1186/1472-6807-13-26

Muller CA, Oberauner-Wappis L, Peyman A, Amos GC, Wellington EM, Berg G (2015) Mining for nonribosomal peptide synthetase and polyketide synthase genes revealed a high level of

diversity in the Sphagnum bog metagenome. Appl Environ Microbiol 81(15):5064–5072. doi:10.1128/AEM.00631-15

Newman DJ, Cragg GM (2012) Natural products as sources of new drugs over the 30 years from 1981 to 2010. J Nat Prod 75(3):311–335. doi:10.1021/np200906s

Nichols D, Cahoon N, Trakhtenberg EM, Pham L, Mehta A, Belanger A et al (2010) Use of ichip for high-throughput in situ cultivation of "uncultivable" microbial species. Appl Environ Microbiol 76(8):2445–2450. doi:10.1128/AEM.01754-09

O'Brien J, Wright GD (2011) An ecological perspective of microbial secondary metabolism. Curr Opin Biotechnol 22(4):552–558. doi:10.1016/j.copbio.2011.03.010

Parsley LC, Linneman J, Goode AM, Becklund K, George I, Goodman RM et al (2011) Polyketide synthase pathways identified from a metagenomic library are derived from soil Acidobacteria. FEMS Microbiol Ecol 78(1):176–187. doi:10.1111/j.1574-6941.2011.01122.x

Quail MA, Matthews L, Sims S, Lloyd C, Beasley H, Baxter SW (2011) Genomic libraries: I. Construction and screening of fosmid genomic libraries. Methods Mol Biol 772:37–58. doi:10.1007/978-1-61779-228-1_3

Rausch C, Hoof I, Weber T, Wohlleben W, Huson DH (2007) Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. BMC Evol Biol 7(1):78. doi:10.1186/1471-2148-7-78

Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng JF et al (2013) Insights into the phylogeny and coding potential of microbial dark matter. Nature 499(7459):431–437. doi:10.1038/nature12352

Rondon MR, August PR, Bettermann AD, Brady SF, Grossman TH, Liles MR et al (2000) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. Appl Environ Microbiol 66(6):2541–2547

Sabree ZL, Bergendahl V, Liles MR, Burgess RR, Goodman RM, Handelsman J (2006) Identification and characterization of the gene encoding the Acidobacterium capsulatum major sigma factor. Gene 376(1):144–151. doi:10.1016/j.gene.2006.02.033

Schoenborn L, Yates PS, Grinton BE, Hugenholtz P, Janssen PH (2004) Liquid serial dilution is inferior to solid media for isolation of cultures representative of the phylum-level diversity of soil bacteria. Appl Environ Microbiol 70(7):4363–4366. doi:10.1128/AEM.70.7.4363-4366.2004

Shen B (2003) Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. Curr Opin Chem Biol 7(2):285–295. doi:10.1016/S1367-5931(03)00020-6

Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y, Simon M (1992) Cloning and stable maintenance of 300-Kilobase-pair fragments of human DNA in Escherichia-Coli using an F-factor-based vector. Proc Natl Acad Sci U S A 89(18):8794–8797. doi:10.1073/pnas.89.18.8794

Siezen RJ, Khayatt BI (2008) Natural products genomics. Microb Biotechnol 1(4):275–282. doi:10.1111/j.1751-7915.2008.00044.x

Snyder AB, Worobo RW (2014) Chemical and genetic characterization of bacteriocins: antimicrobial peptides for food safety. J Sci Food Agric 94(1):28–44. doi:10.1002/jsfa.6293

Staunton J, Weissman KJ (2001) Polyketide biosynthesis: a millennium review. Nat Prod Rep 18(4):380–416. doi:10.1039/a909079g

Suenaga H (2012) Targeted metagenomics: a high-resolution metagenomics approach for specific gene clusters in complex microbial communities. Environ Microbiol 14(1):13–22. doi:10.1111/j.1462-2920.2011.02438.x

Summers RG, Donadio S, Staver MJ, Wendt-Pienkowski E, Hutchinson CR, Katz L (1997) Sequencing and mutagenesis of genes from the erythromycin biosynthetic gene cluster of Saccharopolyspora erythraea that are involved in L-mycarose and D-desosamine production. Microbiology 143(10):3251–3262. doi:10.1099/00221287-143-10-3251

Sundlov JA, Shi C, Wilson DJ, Aldrich CC, Gulick AM (2012) Structural and functional investigation of the intermolecular interaction between NRPS adenylation and carrier protein domains. Chem Biol 19(2):188–198. doi:10.1016/j.chembiol.2011.11.013

Treangen TJ, Salzberg SL (2012) Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat Rev Genet 13(1):36–46. doi:10.1038/nrg3117

Tulp M, Bohlin L (2005) Rediscovery of known natural compounds: nuisance or goldmine? Bioorg Med Chem 13(17):5274–5282. doi:10.1016/j.bmc.2005.05.067

Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM et al (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature 428(6978):37–43. doi:10.1038/nature02340

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(21):7029–7035. doi:10.1128/AEM.00464-10

Uchiyama T, Abe T, Ikemura T, Watanabe K (2005) Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. Nat Biotechnol 23(1):88–93. doi:10.1038/nbt1048

van Belkum MJ, Martin-Visscher LA, Vederas JC (2011) Structure and genetics of circular bacteriocins. Trends Microbiol 19(8):411–418. doi:10.1016/j.tim.2011.04.004

Wawrik B, Kerkhof L, Zylstra GJ, Kukor JJ (2005) Identification of unique type II polyketide synthase genes in soil. Appl Environ Microbiol 71(5):2232–2238. doi:10.1128/AEM.71.5.2232-2238.2005

Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R et al (2015) antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res 43(W1):W237–W243. doi:10.1093/nar/gkv437

Wild J, Hradecna Z, Szybalski W (2002) Conditionally amplifiable BACs: switching from single-copy to high-copy vectors and genomic clones. Genome Res 12(9):1434–1444. doi:10.1101/gr.130502

Yun J, Ryu S (2005) Screening for novel enzymes from metagenome and SIGEX, as a way to improve it. Microb Cell Fact 4(1):8. doi:10.1186/1475-2859-4-8

Zacharof MP, Lovitt RW (2012) Bacteriocins produced by lactic acid bacteria a review article. APCBEE Procedia 2:50–56. doi:10.1016/j.apcbee.2012.06.010

Zerikly M, Challis GL (2009) Strategies for the discovery of new natural products by genome mining. ChemBioChem 10(4):625–633. doi:10.1002/cbic.200800389