

REVIEW ARTICLE

Mining prokaryotes for antimicrobial compounds: from diversity to function

Vittorio Tracanna^{1,†}, Anne de Jong^{2,†}, Marnix H. Medema^{1,*}
and Oscar P. Kuipers^{2,*}

¹Bioinformatics Group, Wageningen University, Droevendaalsesteeg 1, Radix West, Building 107, 6708 PB Wageningen, The Netherlands and ²Molecular Genetics, University of Groningen, Nijenborgh 7, 9726AG Groningen, The Netherlands

*Corresponding authors: Marnix H. Medema: 1Bioinformatics Group, Wageningen University, Droevendaalsesteeg 1, Radix West, Building 107, 6708 PB Wageningen, The Netherlands. Tel: +31317484706; E-mail: marnix.medema@wur.nl; Oscar P. Kuipers: Molecular Genetics, University of Groningen, Nijenborgh 7, 9726AG Groningen, The Netherlands. Tel: +31 503632093; E-mail: o.p.kuipers@rug.nl

[†]Equal contribution.

One sentence summary: Mining prokaryotes for antimicrobial compounds: from diversity to function.

Editor: Ehud Banin

ABSTRACT

The bacterial kingdom provides a major source of antimicrobials that can either be directly applied or used as scaffolds to further improve their functionality in the host. The rapidly increasing amount of bacterial genomic, metabolomic and transcriptomic data offers unique opportunities to apply a variety of approaches to mine for existing and novel antimicrobials. Here, we discuss several powerful mining approaches to identify novel molecules with antimicrobial activity across structurally diverse natural products, including ribosomally synthesized and posttranslationally modified peptides, nonribosomal peptides and polyketides. We not only discuss the direct mining of genomes based on identification of biosynthetic gene clusters, but also describe more advanced and integrative approaches in ecology-based mining, functionality-based mining and mode-of-action-based mining. These efforts are likely to accelerate the discovery and development of novel antimicrobial drugs.

Keywords: antimicrobial; bacteria; biosynthetic; gene clusters; genome; mining; evolution; NRPS; RiPP; PKS

INTRODUCTION

Microbial specialized metabolites are the major source of antimicrobials currently used in the clinic, in agriculture and in food manufacturing. Due to the rapid development and spread of resistance against these molecules, there is an urgent need for novel compounds that can supplement the current arsenal. Since the 'golden age of antibiotics' in the sixties and seventies, there has been a steady decrease of novel antibiotic molecules entering the market. However, the recent development of computational genomic approaches to natural product discovery is replenishing hopes that this trend can be turned around: in

prokaryotic genome sequences, tens of thousands of biosynthetic gene clusters (BGCs) have been identified (Cimermanic *et al.* 2014; Doroghazi *et al.* 2014; Skinnider *et al.* 2016). Thousands of these BGCs are likely to encode the biosynthesis of thus far unknown molecules (Dejong *et al.* 2016). To uncover these, many innovative approaches are being developed to link them to metabolomic data with high throughput (Kersten *et al.* 2011, 2013; Medema *et al.* 2014; Mohimani *et al.* 2014) or to refactor synthetic versions of them for heterologous expression (Chang *et al.* 2013; Shao *et al.* 2013; Yamanaka *et al.* 2014; Kang, Charlop-Powers and Brady 2016; Montalbán-López and Kuipers 2016; van

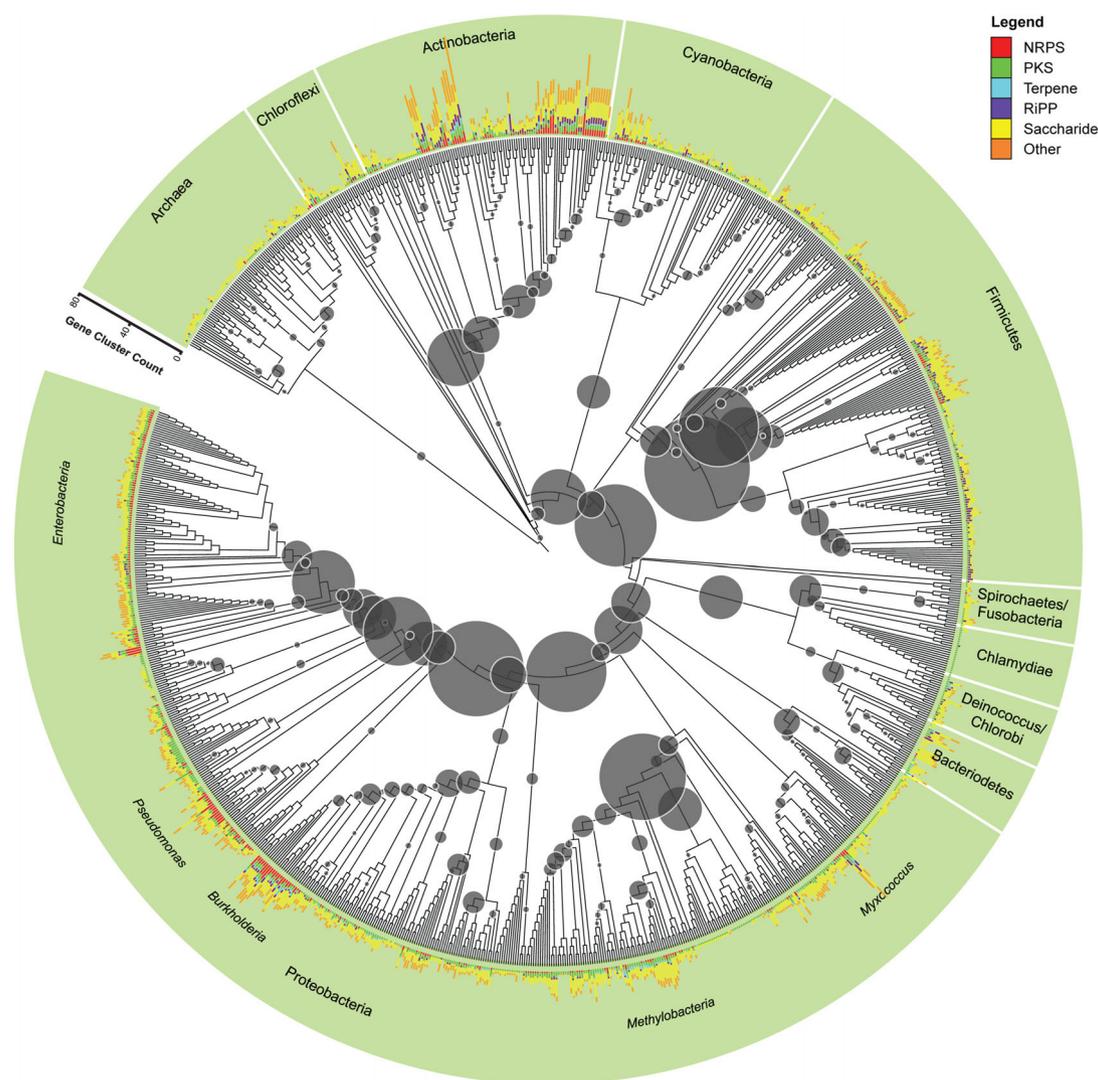


Figure 1. Taxonomic diversity of BGCs across bacteria and archaea. Bar plots indicating gene cluster counts as detected by antiSMASH+ClusterFinder (Cimermanic et al. 2014; Weber et al. 2015) are plotted onto prokaryotic taxonomy. Circles on the tree indicate the amount of within-taxon variation, estimated using the quadratic entropy index. Figure adapted from Cimermanic et al. (2014) with permission.

Heel et al. 2016). However, detailed biochemical characterization of biosynthetic pathways and their products is still painstakingly slow and laborious, and many BGCs encode the production of natural products without any (useful) antimicrobial activity. Therefore, targeted approaches are needed to selectively mine genomes for natural products with antimicrobial properties, and to narrow down from tens of thousands of potentially interesting BGCs to manageable numbers that can be tested in the laboratory. Here, we outline and discuss several currently emerging computational and experimental strategies to this end, based on the analysis of chemical diversity, ecology and evolution, organismal function and/or modes of action.

DIVERSITY-BASED MINING FOR ANTIMICROBIALS

Genome-based identification of biosynthetic pathways for antimicrobials

Genomic information has become more and more important in the process of identifying novel biosynthetic pathways for

the production of antimicrobials. To identify the potential of a bacterium to produce bioactive natural products, mining for BGCs is particularly useful. A wide range of bioinformatic tools (e.g., antiSMASH (Weber et al. 2015), BAGEL3 (van Heel et al. 2013) and PRISM (Skinnider et al. 2016)) are available to identify these, mostly based on shared properties among known classes of biosynthetic pathways (Boddy 2014; Medema and Fischbach 2015; Pi et al. 2015; Ziemert, Alanjary and Weber 2016). For instance, the modification enzymes for the production of lanthipeptides are well conserved (Knerr and van der Donk 2012); as such, they can be used as ‘anchors’ or ‘signatures’ for genome mining.

Taxonomically, there are significant differences in biosynthetic richness and diversity across various branches of the tree of life (Fig. 1). Traditionally rich sources of natural products derived from *Streptomyces*, *Bacillus* (Zhao and Kuipers 2016) and *Pseudomonas* are now becoming supplemented with newly discovered genera. A great example of taxonomic diversity-based mining for antimicrobials is provided by *Entotheonella*, a group of sponge-microbiota-derived organisms that can produce extremely highly modified peptides with over six

different posttranslational modifications, including D-amino acids, methylated amino acids and dehydrated amino acids and dozens of residues modified (Wilson et al. 2014). Several other microbial taxa with high biosynthetic potential, such as *Clostridium*, *Burkholderia*, *Pseudonocardia*, *Photorhabdus*, *Xenorhabdus*, *Chitinophaga*, *Herpetosiphon* and *Planctomyces*, are emerging as novel target genera (Challinor and Bode 2015). Most of these genera are characterized by large genome sizes, which has been shown to correlate with an increased percentage of genomic capacity devoted to specialized metabolism (Cimermancic et al. 2014); however, a very large percentage of extant biosynthetic diversity in the biosphere is still found in the smaller genomes of the majority of bacteria, which should not be discarded in the search for antimicrobials; e.g. the biosynthesis of the recently discovered antibiotic non-ribosomal peptide lugdunin (Zipperer et al. 2016) is encoded in the genome of the human commensal *Staphylococcus lugdunensis*, which harbors just four other BGCs (as identified in antiSMASH-DB, Blin et al. 2017). The wide diversity of BGCs found in large as well as small genomes emphasizes the almost boundless chemical diversity of nature, and it is not at all unlikely that new classes of antimicrobials will be discovered from many novel sources in the next decades.

Even though natural product BGCs can thus be accurately identified and quantified, the question still remains which of these are most likely to encode the production of potent antimicrobials. Not only the taxonomic origins, but also the chemical structures of antimicrobials are highly diverse. Indeed, known antimicrobial compounds represent a cross-section of several chemical and biosynthetic classes, such as ribosomally synthesized and posttranslationally modified peptides (RiPPs), non-ribosomal peptides, polyketides, terpenoids and even oligosaccharides (see Table 1 for a comprehensive schematic overview). Although the percentage of antimicrobials is probably higher among RiPPs than among saccharides, for example, each chemical class of natural products comprises both many antimicrobials and many compounds with different biological activities. To prioritize specifically for antimicrobials, it is necessary to go beyond the genome sequences and couple the genomic information to ecological and functional data.

ENVIRONMENTAL AND ECOLOGY-BASED MINING FOR ANTIMICROBIALS

Using metagenomics to chart biosynthetic diversity

From soil to ocean, from plant roots to animal guts, the ecosystems in which natural products are found are highly diverse. Also within these ecosystems, the diversity is enormous: a gram of soil is estimated to contain hundreds to thousands of different species (Curtis, Sloan and Scannell 2002) that form an extremely intertwined society. The metabolic potential hidden in those communities is immense, and systematic analysis of soils across the globe shows very little overlap between the secondary metabolite repertoire of similar soils (Charlop-Powers et al. 2014). Potentially, understanding of ecology and microbial communities can be used to chart this variation and prioritize BGCs that are likely to encode the synthesis of molecules that function as potent antimicrobials. Metagenomics is a key technology that allows surveying BGCs and their abundances across varying communities (Wilson and Piel 2013; Charlop-Powers, Milshteyn and Brady 2014).

In the hunt for novel antimicrobials from the environment, there are two important strategies: searching for novel chemical scaffolds and searching for novel congeners. Compounds with

novel scaffolds are more difficult to discover (also due to rediscovery of known molecules for which no BGCs have yet been characterized), and predicting chemical structures and biological activities from BGC sequence data alone is very challenging. BGCs encoding the biosynthesis of potential congeners, variants upon an existing (and often extensively studied) molecular scaffold, however, can easily be identified based on sequence homology. While sharing the biosynthetic origin with their well-known counterparts, the small differences in the structures of some congeners can have a major effect on different characteristics of the molecule. Notably, those changes may affect the potency, toxicity profile, or the target of the compound, or even the resistance of microorganisms against it. Phylogenetic studies on novel variants of known BGCs can potentially be used to infer the substrates involved and the final products synthesized by the encoded pathway. Regardless of the product type and its discovery method, engineering expression of identified BGCs in a native or heterologous host is frequently necessary for both novel scaffolds and congeners, and is currently a rate-limiting step in antibiotic discovery. Another bottleneck lies in the charting of biosynthetic diversity, as this is often limited to organisms that are easy to culture in the laboratory.

Metagenomics can overcome culture restrictions by sampling material directly from the environment of interest. Alternatively, functional amplicon sequencing approaches can achieve high-sequencing depth for BGCs by targeting the shared conserved regions of, e.g. polyketide synthases or non-ribosomal peptide synthetases with degenerate primers based on previously characterized genes. The molecule type for which the biosynthesis is encoded in an underlying BGCs can be predicted by assessing the sequence similarity between the sequence of the amplicon in question and curated data aggregation platforms such as MIBiG (Medema et al. 2015) and eSNaPD (Reddy et al. 2014). Development of primer pairs based on BGC classes that specifically encode the biosynthesis of antimicrobial compounds can be used to specifically target BGCs for potential congeners of known antimicrobials. The enormous cost decrease of high-throughput sequencing technologies now also allows generation of the immense quantity of data necessary to assemble BGCs directly from microbial communities. In shotgun metagenomics, environmental DNA, eDNA, is extracted from a community sample and sequenced with short read NGS technology. Sequence information can be assembled with metagenomics-specific assemblers as metaSPAdes (Nurk et al. 2016) or Ray Meta (Boisvert et al. 2012). Additional sequencing platforms such as PacBio and more recent Oxford Nanopore can be used to produce long sequence reads, which can aid in assembling contigs long enough to harbor complete BGCs, even in complex communities. Perhaps even more powerful to this end are artificial long read technologies: due to their high throughput, platforms as 10X Genomics and TruSeq, may start a new era in metagenomics studies. These synthetic long read platforms, can be used to reconstruct large numbers of long eDNA stretches with low error rates from complex metagenomes. eDNA is first digested in high molecular weight fragments, which are sorted and barcoded in different pools. Standard shotgun sequencing of each pool is then followed by an assembly of each high molecular weight DNA molecule. Although some species bias is introduced during library preparation, the approach has been shown to enable the assembly of synthetic long reads from relatively rare microorganisms in soil (Sharon et al. 2015). TruSeq has the read length necessary to assemble whole-length BGCs even in complex metagenomes, aided by specialized algorithms such as TruSPAdes (Bankevich and Pevzner 2016).

Table 1. Overview of known antimicrobial compound classes. For information on the biosynthetic diversity of all these compounds the reader is referred to several excellent recent reviews (Sanchez, Chiang and Wang 2008; Strieker and Marahiel 2009; Arnison et al. 2013; Fisch 2013; McCranie and Bachmann 2014; Yim et al. 2014; Dickschat 2016; Helfrich and Piel 2016; Ortega and van der Donk, Wilfred 2016).

Class	Example compound	Simplified structure
Class I: RiPP (<10 kDa)		
1. lanthipeptides	Type 1: nisin	
	Type 2: mersacidin	
2. head to tail cyclized	AS-48	
3. sactipeptides	subtilosin	
4. linear azole containing (LAP)	microcin B17	
5. thiopeptides	thiocillin	
6. glycocins	sublancin	
7. lassopeptides	microcin J25	
8. polytheonamides	proteusin	
Class II: unmodified ribosomal peptides (<10kDa)		
1. pediocin-like	pediocin	
2. other	lactococcin A	
Class III: large antimicrobial peptides (>10 kDa)		
Colicins and megacins	colicin B	
NRPS		
Lipopeptides	daptomycin	
Glycopeptides	vancomycin	
Pyrrolobenzodiazepines	anthramycin	
Lincosamides	lincomycin	
Cyclododecapeptides	valinomycin	

Table 1 (continued).

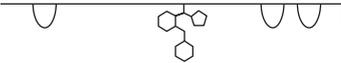
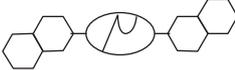
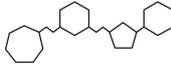
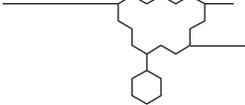
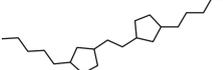
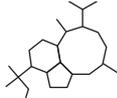
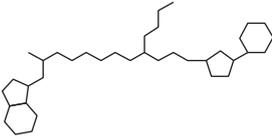
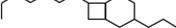
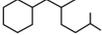
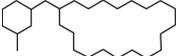
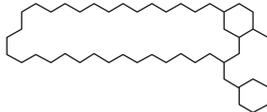
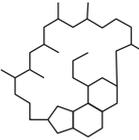
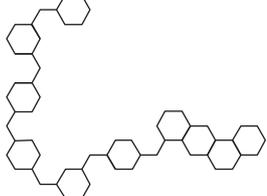
Class	Example compound	Simplified structure
Bleomycin-like antitumor antibiotics	bleomycin	
Quinomycin antibiotics	echinomycin	
Capuramycin-type antibiotics	capuramycin	
Tuberculostatic antibiotics (viomycin-like)	viomycin	
Cyclic diamidines	congocidine	
Indolactams	pendolmycin	
Pyrothine antibiotics	thiolutin	
Uridyl peptide antibiotics	pacidamycin	
Beta-lactam antibiotics	cephamycin C	
Chloramphenicol	chloramphenicol	
PKS		
Macrolides, macrolactams, polyethers, polyenes	erythromycin (macrolide)	
	vicenistatin (macrolactam)	
	amphotericin B (polyene)	
Ansamycins	rifamycin	
	landomycinA (angucycline)	

Table 1 (continued).

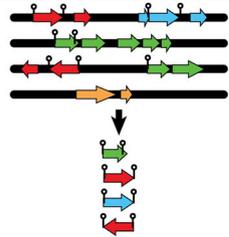
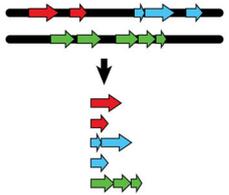
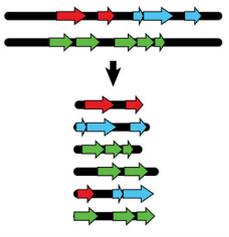
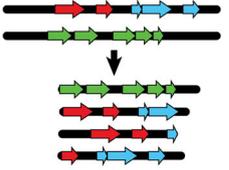
Class	Example compound	Simplified structure
Anthracyclines	daunorubicin	
Enediyines	neocarzinostatin	
Benzoisochromanequinones	actinorhodin	
Tetracyclines	tetracycline	
Others		
Nucleoside antibiotics	tunicamycin	
Carbapenems	thienamycin	
Aminoglycosides / aminocyclitols	spectinomycin	
Aminocoumarins	novobiocin	
Antimicrobial terpenoids	pentalenolactone	
Phosphonate antibiotics	fosfomicin	
Phenazine antibiotics	phenazine	
Indolocarbazoles	staurosporine	
Quinolones	quinolone	

Metagenome sequencing is able to find novel BGCs regardless of their conservation or representation across known genomes. The main advantage of these techniques compared to amplicon-based approaches lies in their unbiased nature (Table 2). The main disadvantage of this approach lies in the data complexity. However, a plethora of different solutions are becoming available to tackle this. One direct solution that allows untangling of the community is represented by single-cell sequencing. Although the protocols vary and evolve over time, the scope is unchanged: a single cell is

isolated, its DNA extracted and amplified to undergo a PCR screen or genome sequencing step. There are examples of successful application of single-cell sequencing to identify important secondary metabolite gene clusters such as the apratoxin pathway from a filamentous cyanobacterium (Grindberg *et al.* 2011).

Important steps are being made to apply the latest chromatin fixation techniques, such as Hi-C and 3C, to improve metagenomic assemblies. As of now, those techniques were successfully used to aid the assembly of synthetic metagenomes

Table 2. Strong and weak points of the different sequencing methods are described in this table. In the figures, arrows represent genes that are part of a BGC, and genes with the same colors originate from the same operon. The pins in the top right figure indicate conserved stretches targeted by custom primers.

	Pros	Cons	Target	Representation
Functional amplicon sequencing	+ Inexpensive + Data has low complexity	- Incomplete coverage of biosynthetic diversity - Suitable conserved sequence regions are required - Only targets part of the BGC	Individual signature genes or domains	
Metatranscriptomics	+ Allows prioritization of BGCs based on biological responses + Allows better connection of genomic with metabolomic and phenotypic data	- Silent BGCs are not sequenced - Metagenome data required to reconstruct multi-operon BGCs	Operons	
Conventional metagenomics	+ Targets entire BGC + Relatively unbiased	- Assemblies will be fragmented in complex communities - Complex data analysis	Complete BGCs	
HiC / TruSeq / 10X metagenomics	+ Long scaffolds: many BGCs can be completely assembled	- Complex sample preparation - Some species bias - Expensive	Complete BGCs	

(Beitel et al. 2014; Burton et al. 2014). Potentially, fixation techniques can become powerful tools to validate metagenomics assemblies, as they provide an extra layer of information that is not used by current binning tools. In addition, information on contiguity or genomic distance between contigs will allow to reconstruct longer and more complete clusters from raw sequence data.

Metagenomics-derived BGCs are not easy to revive in heterologous hosts. For instance, proper amounts of physical DNA that contains the BGC of interest is often not readily available; hence, the isolation of the natural producer or synthetic refactoring of the gene cluster is required. Also, regulation of the transcription and the precursors required for the encoded pathway are sometimes not functionally available in classic hosts. Synthetic DNA costs have decreased substantially in the last years, opening the path to high-throughput expression of BGCs through synthesis and refactoring in order to match the impressive output of metagenomics analysis. Refactored BGCs are designed to achieve a better control over expression levels in the heterologous host, by replacing native regulation by synthetic promoters, ribosome-binding sites and terminators (Medema et al. 2011; Smanski et al. 2016). Also, codon usage can be re-designed to match the host to increase mRNA translation speed. Further manipulation of refactored BGCs is much easier compared to the original, greatly reducing the development time for

BGC-derived products. Key challenges that need to be overcome here are DNA synthesis costs, tuning stoichiometry of gene expression and avoiding the introduction of unexpected functions into synthetic DNA.

Ecology-based prioritization for antimicrobial function

To understand bacteria and harness their metabolic potential, it is important to consider them within a microbial community framework. Indeed, bacteria that thrive under physicochemical conditions in which they have an elaborate interaction network with other species are excellent targets for identification of secondary metabolites. Accordingly, complex communities are reported to be more resistant to invasion from alien species (van Elsas et al. 2012). Multiple factors can influence community resilience, including the production of secondary metabolites with a negative interspecies interaction function, i.e. antibiotics (Cordero et al. 2012).

Given the ubiquitous nature of potential target communities, the selection of promising candidates is vital during the experimental design phase. Biologists can use knowledge on an unexpected phenotype to deduce the presence of potent antimicrobial compounds. Marine sponges are a great example of prime candidates when hunting for novel and potent antibacterial compounds. Sponges have a strong endosymbi-

otic relation with bacteria to the point that up to 40% of their bio mass is composed by bacterial cells (Friedrich et al. 2001). The microbial community inside the sponge is radically different from the surrounding water. Lately, sponge communities were successfully targeted for identification of novel natural products with antibacterial properties, such as polytheonamides (Trindade-Silva et al. 2012), which might play a role in protecting the sponge host against predators. Suppressive soils are another great example of ecology-inspired mining. These soils provide protection to plants against specific pathogens. Conducive soils, which do not restrain the development of a disease, can be gradually transformed into suppressive soils by infecting plants that grow in it in multiple cycles (Berendsen, Pieterse and Bakker 2012). Hence, by investigating the BGCs that are abundant in or expressed by a community in specific conditions (e.g. treatment of suppressive soil with a pathogen compared to control), candidates responsible for the phenotype can be prioritized: this type of strategy has led to the discovery of the lipopeptide thanamycin, which suppresses fungal root pathogens (Mendes et al. 2011; Watrous et al. 2012).

Principled approaches to prioritize environmental BGCs based on ecology require more than just bare metagenome sequences. Specifically, metatranscriptomics and/or environmental metadata can be used to map and understand differences in BGC abundance and expression, in order to prioritize for those that are most likely to function as antimicrobials. Also, metadata on environmental and physicochemical conditions pertaining microbial communities can potentially be used to direct the search of BGC hotspots. Localized nutritional hotspots, such as organic particles in the ocean, may create highly competitive environmental niches, where microbes utilize antimicrobials to secure resources against competitors. Therefore, metadata on, e.g., nutrient availability could potentially be exploited to identify priority targets. When metatranscriptomics data are generated for diverse samples for which metadata are also recorded, one could even identify BGCs that are expressed specifically under such conditions; these would then have an elevated probability to be involved in generating antimicrobial activity.

Several tools and databases are available to study metagenomes in the context of metadata: e.g. the eSNaPD webserver (Owen et al. 2013) is a metagenomics atlas that holds BGC distributions in soil across the globe. Different physicochemical characteristics and secondary metabolite repertoires are aggregated for over 100 samples of different biomes. The tools integrated in the webserver allow a quick and intuitive inspection of the data. Similarly, the Tara ocean webserver (<https://www.embl.de/tara-oceans/>) hosts metagenomic information and metadata on a wide range of marine environments. Particular attention was directed towards the different sampling depths as they are related to the light intensity, an essential driver of community composition. For the human microbiome, which also hosts a wide range of BGCs (including antimicrobials, see Donia et al. 2014), several datasets are available with rich clinical metadata, such as the Belgian Flemish Gut Flora Project (Falony et al. 2016) and the Dutch LifeLines-DEEP study (Zhernakova et al. 2016).

As more and more metagenomes become available with better assemblies (and including more and more metagenome-assembled genomes), richer metadata and rapidly rising amounts of comprehensive (time-series) metatranscriptomics data, the opportunities for ecology-based antibiotic discovery are likely to increase drastically.

FUNCTION-BASED MINING FOR ANTIMICROBIALS

Use of protein domains for genome mining based on predicted function

In any strategy for genome-based mining for antimicrobials, a key step is the identification of genes involved in their biosynthesis. A robust method to achieve this, especially for proteins that do not have high similarity to proteins with known function, is through the detection of one or more conserved protein domains using curated models. Certain protein domains or domain combinations are indicative of biochemical functions that are specific to certain biosynthetic pathways; hence, they can be used as 'anchors' or 'signatures' to identify certain classes of BGCs. The gene cluster identification algorithms in tools such as antiSMASH, BAGEL3 and PRISM are based on this principle. As we learn more about the enzymology of novel classes of natural products, this will allow the addition of more 'domain markers' that can be used for new mining strategies. Recently, such strategies have been used to systematically mine genomes for biosynthetic pathways encoding cyanobactins (Leikoski et al. 2013), thiazole/oxazole-modified microcins (Cox, Doroghazi and Mitchell 2015) and enediynes (Shen et al. 2015).

Identifying natural products by phenotypic and metabolic profiling

Several types of high-throughput profiling methods have become available to screen large collections of organisms for their natural-product-producing potential. Besides (meta)genomic information, these also leverage phenotypic and metabolomic data.

Traditionally, phenotypic information has been key in the discovery of antimicrobials, as strain collections were screened for activity against various pathogens. Besides conventional growth inhibition assays, an interesting emerging technique is cytological profiling, which allows 'function-first' mining of natural products with desired biological activities by predicting the mechanisms of action based on microscopy analysis of the cellular responses of target strains (Nonejuie et al. 2013; Potts et al. 2013; Schulze et al. 2013; Woehrmann et al. 2013). When applied onto microbial extracts, this technique can be used to specifically search for natural products with certain functional profiles (i.e. activity profiles that match those of molecules with known modes of action) (Ochoa et al. 2015). A proof-of-principle study in *Bacillus subtilis* showed that through activity-guided purification, molecules with multiple mechanisms of action can even be separated based on this technology (Nonejuie et al. 2016). Moreover, connecting molecules to activities can even be automated through a procedure called Compound Activity Mapping (Zhang et al. 2016), which applies networking analysis to correlate specific molecules to each activity.

Metabolic profiling can be done, for example, through large-scale mass spectrometric analysis of strain collections in order to identify all the molecular families (Watrous et al. 2012; Nguyen et al. 2013) that are produced under typical laboratory conditions. This is facilitated by public platforms such as Global Natural Products Social Molecular Networking (<http://gnps.ucsd.edu>) (Wang et al. 2016), which allow connecting metabolomic data from multiple sources. A recent study showed how comprehensive molecular networking analysis of hundreds of strains from the same genus (*Pseudomonas*) can effectively identify novel natural products by comparing all molecules observed across these

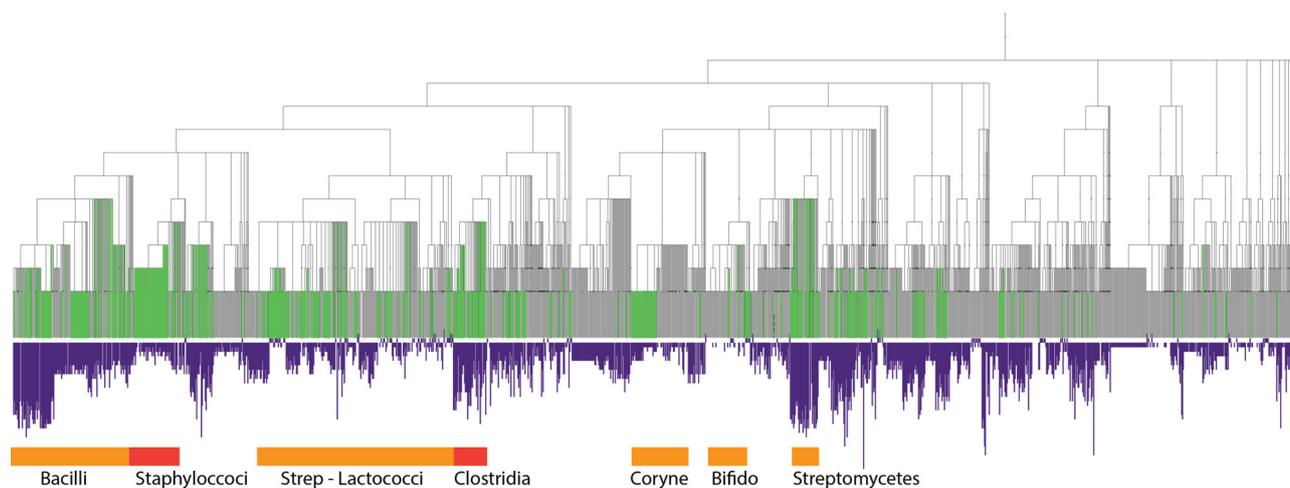


Figure 2. Correlation between the presence of lanthipeptide BGCs and the predicted proteolytic activity. On top, the taxonomic tree of Gram-positive bacteria and the green branches indicate the putative presence of lanthipeptide biosynthetic genes. The purple bars below the tree indicate the predicted abundance of proteolytic enzyme-coding genes. In gray the organism names which can be read in Table S1. The orange bars indicate the position of groups of bacteria.

strains with a set of reference molecules (Nguyen et al. 2016). For some compound types, such as RiPPs and NRPs, dereplication (i.e. identification of novelty) can also be automated with algorithms such as DEREPLICATOR (Mohimani et al. 2016). A disadvantage of metabolomics-based natural product discovery is that one can only observe molecules that are produced under the conditions tested; yet, this can be partially mitigated by performing molecular networking on co-cultivated strains (Traxler et al. 2013; Briand et al. 2016).

Although the throughput of metabolic and cytologic profiling methods is increasing, smart selection strategies for strains to be used as input are still very important. Genome-based analysis of biological features that correlate with natural product biosynthetic potential could be an interesting strategy for this. For example, bacteria that have a large capacity for the biosynthesis of posttranslationally modified (and thus more stable) ribosomally synthesized natural products (instead of unmodified bacteriocins) would be expected to show relatively high proteolytic activity. Hence, genomic profiling of proteolytic capacity could enable selection of taxonomic groups that are more likely to yield novel RiPPs, as the RiPP BGC types that can be detected with current bioinformatic methods are likely to be just the tip of the iceberg. Such an analysis could help identify bacteria that are good candidates for discovery of novel RiPP classes. Figure 2 shows that, indeed, for a number of taxa, the presence of proteases and peptidases in the genome shows a clear correlation to the presence of lanthipeptide BGCs. To analyze this, we selected 180 peptidase and 30 protease motifs from the PFAM database (Table S1, Supporting Information) and screened all complete bacterial genomes for the number of motifs present to get a rough indication of the proteolytic activity of each bacterium. One would expect highly proteolytic bacteria such as *Bacillales* and *Actinomycetes* to produce a high fraction of modified peptides relative to unmodified ones, to avoid proteolytic degradation. Based on simple matching, the correlation between the presence of predicted lantibiotic and high protease or peptidase abundance (number of motifs above the median) is 0.75 and 0.62, respectively. Although the expected correlation can be shown in several cases, it is not always consistent. Staphylococci and corynebacteria show low proteolytic enzyme-coding gene abundance, but do encode the biosynthesis of many putative lantibiotics. Clearly, there are other factors also involved in

the evolution of RiPP repertoires. One of them could be the nature of the environment the producing bacteria are thriving in, which could also be very proteolytic. Still, exploration of these and other genomic and phenotypic features that correlate with biosynthetic potential could be useful for strain selection purposes in large-scale genomic and metabolomic experiments.

MODE-OF-ACTION-BASED GENOME MINING FOR ANTIMICROBIALS

As indicated earlier, it is difficult to predict which BGCs encode the production of natural products with antimicrobial activity from sequence alone. Yet, there are at least two potential options to do so: resistance-based genome mining and mining for synergistic antibiotics.

Resistance-based mining (recently also reviewed in some detail by Ziemert, Alanjary and Weber 2016) utilizes the fact that many BGCs encoding the biosynthesis of antibiotics will also encode one or more genes that confer self-resistance to the molecule produced, in order to avoid suicide. These self-resistance genes are frequently the same or very similar to the resistance genes used by other bacteria to evade antibiotics, and are probably acquired by the latter through horizontal gene transfer (Mak et al. 2014; Ogawara 2016); they include transporters, drug-modifying enzymes and, most interestingly, paralogous genes encoding 'resistant' copies of housekeeping proteins that are targeted by the antibiotic. Hence, the identification of self-resistance genes in BGCs can be very effective predictors of antibiotic function of their products. Moreover, if these genes are resistant paralogs of housekeeping genes that are targeted by the antibiotic, they can even be used to predict the target. Recently, Tang et al. (2015) used this approach to identify the BGC for the thiotetronate antibiotic thiolactomycin in *Salinispora* genomes based on the presence of a resistant copy of a fatty acid synthase gene. Similarly, Yeh et al. (2016) identified the biosynthetic genes for the proteasome inhibitor fellutamide B through the identification of a proteasome subunit-encoding gene inside a gene cluster. Intriguingly, the strategy can also be used to predict natural products with new modes of action; Johnston et al. (2016) recently used

resistance-based mining to show that the telomycin family of natural products targets the phospholipid cardiolipin.

Several good libraries of known self-resistance genes and detection models for them are now available (Jia et al. 2016). In particular, the ResFams database (Gibson, Forsberg and Dantas 2015) contains a rich set of profile Hidden Markov Models (pHMMs) that allow detection of resistance genes against a wide range of antibiotics; in their abovementioned study, Johnston et al. (2016) supplemented these pHMMs with an additional 91 models for additional resistance genes found in the literature. Additionally, phylogenetic analysis (reminiscent of the EvoMining approach recently described by Cruz-Morales et al. 2016) may make it possible to hunt for additional resistant paralogs of housekeeping genes, by identifying aberrant copy numbers of such genes in genomes, where the additional copy has long branch lengths and resides inside a BGC. By doing this systematically, while using inclusive and broad probabilistic BGC predictions as offered by algorithms such as ClusterFinder (Cimermancic et al. 2014), large numbers of both novel modes of action and novel antibiotic biosynthetic classes may be discovered.

The second possible bioinformatic strategy to prioritize BGCs for their potential to encode novel antibiotics is based on synergistic interactions. Many cases are known in which a combination of two bioactive compounds acts in a synergistic way to obtain higher antimicrobial efficacy or circumvent resistance mechanisms in the target strains. A famous example of this, that is still used in the clinic, is Augmentin[®], a combination of the beta-lactam antibiotic amoxicillin with the beta-lactamase inhibitor clavulanic acid. In theory, synergistic pairs of natural products have a major advantage in the battle against antimicrobial resistance, as it is more difficult for pathogens to develop resistance against both molecules. Synergistic pairs of natural products also occur in nature. In fact, Augmentin was modeled after the combination of cephamycin and clavulanic acid, which is naturally produced by *Streptomyces clavuligerus*. In the genome of *S. clavuligerus*, the BGCs for the two compounds are intertwined in a 'supercluster' configuration, allowing the co-ordinated regulation of both biosynthetic pathways (Ward and Hodgson 1993; Medema et al. 2010). Intriguingly, multiple other cases have been identified where two synergistic natural products are jointly encoded in such superclusters: examples include the synergistic antibiotics lankamycin and lankacidin (Suwa et al. 2000), which bind complementary sites on the large ribosomal subunit (Belousoff et al. 2011), and the synergistic antibiotics griseoviridin and viridogrisein (Xie et al. 2016). Hence, it is likely that intertwined supercluster configurations are indicative of synergistic interactions between natural products, which thus far has only been linked to antibiotic function. Automated comparative genomic analysis of exceptionally large 'hybrid' BGCs (encoding enzymes to generate multiple different scaffold types) predicted by tools such as antiSMASH, may make it possible to trace the evolutionary histories of their genes, and assess whether they encode 'superclusters' encoding the production of multiple different (and possibly synergistic) molecules, based on whether the BGC has relatively recently originated by merger of two previously independent gene clusters. Furthermore, analysis of regulatory motifs for transcription factor binding sites (Wolf et al. 2016) can potentially distinguish such superclusters from pairs of gene clusters which happen to be encoded right next to each other on the genome by chance, by assessing the likelihood that genes on both sides of the putative supercluster are in fact coregulated.

All in all, at least two strategies based on sequence analysis alone are available to mine the vast numbers of extant BGCs for

novel antibiotics. Yet, when sequence analysis is complemented by ecological information, e.g. from metagenomics, many more identification possibilities become available.

DISCUSSION AND OUTLOOK

Considering the wealth of opportunities offered by the ever-increasing sequencing power that allows studies ranging from community mining to elucidating expression circuitries and protein functional analyses, we can only expect an ever-increasing pile of information on known and novel antimicrobials. Big tasks ahead are the functional and structural characterization of these compounds, assessing their potential synergistic or antagonistic activities, and determining the roles of (cross)immunity against these compounds. As also completely novel biosynthetic routes and new peptide modifications are being discovered at a regular pace, we need to codevelop high-throughput methodologies for activity screening, employing elicitors of their production if necessary. The interplay of bacteria with their environment, with other organisms and with hosts (e.g. gut microbiota, plant microbiota) will certainly involve antimicrobials of all sorts, and assessing their role in population dynamics will be an important area of research in the coming decade. Moreover, synthetic biology and novel chemistry approaches will enable the production of new-to-nature compounds to fight the increasing and important problem of antibiotic resistance in major human pathogens. A combination of mining approaches at all levels, followed by biochemical and pharmaceutical studies, will be greatly beneficial to find the solutions for major societal problems ahead.

SUPPLEMENTARY DATA

Supplementary data are available at [FEMSRE](https://femsre.oup.com/femsre/article-abstract/41/3/417/3574665) online.

ACKNOWLEDGEMENTS

VT is supported by the research programme NWO-Groen, which is jointly funded by the Netherlands Organisation for Scientific Research (NWO), BASF SE and Baseclear BV, under project number ALWGR.2015.1. MHM is supported by VENI grant 863.15.002 from The Netherlands Organization for Scientific Research (NWO).

Conflict of interest. None declared.

REFERENCES

- Arnison PG, Bibb MJ, Bierbaum G et al. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat Prod Rep* 2013;**30**:108–60.
- Bankevich A, Pevzner PA. TruSPAdes: barcode assembly of TruSeq synthetic long reads. *Nat Methods* 2016;**13**:248–50.
- Beitel CW, Froenicke L, Lang JM et al. Strain- and plasmid-level deconvolution of a synthetic metagenome by sequencing proximity ligation products. *PeerJ* 2014;**2**:e415.
- Belousoff MJ, Shapira T, Bashan A et al. Crystal structure of the synergistic antibiotic pair, lankamycin and lankacidin, in complex with the large ribosomal subunit. *P Natl Acad Sci USA* 2011;**108**:2717–5.
- Berendsen RL, Pieterse CM, Bakker PA. The rhizosphere microbiome and plant health. *Trends Plant Sci* 2012;**17**:478–86.

- Blin K, Medema MH, Kottmann R et al. The antiSMASH database, a comprehensive database of microbial secondary metabolite biosynthetic gene clusters. *Nucleic Acids Res* 2017;**45**:D555–9.
- Boddy CN. Bioinformatics tools for genome mining of polyketide and non-ribosomal peptides. *J Ind Microbiol Biot* 2014;**41**:443–50.
- Boisvert S, Raymond F, Godzaridis E et al. Ray Meta: scalable de novo metagenome assembly and profiling. *Genome Biol* 2012;**13**:R122.
- Briand E, Bormans M, Gugger M et al. Changes in secondary metabolic profiles of *Microcystis aeruginosa* strains in response to intraspecific interactions. *Environ Microbiol* 2016;**18**:384–400.
- Burton JN, Liachko I, Dunham MJ et al. Species-level deconvolution of metagenome assemblies with Hi-C-based contact probability maps. *G3 (Bethesda)* 2014;**4**:1339–46.
- Challinor VL, Bode HB. Bioactive natural products from novel microbial sources. *Ann N Y Acad Sci* 2015;**1354**:82–97.
- Chang FY, Ternei MA, Calle PY et al. Discovery and synthetic refactoring of tryptophan dimer gene clusters from the environment. *J Am Chem Soc* 2013;**135**:17906–12.
- Charlop-Powers Z, Milshteyn A, Brady SF. Metagenomic small molecule discovery methods. *Curr Opin Microbiol* 2014;**19**:70–5.
- Charlop-Powers Z, Owen JG, Reddy BV et al. Chemical-biogeographic survey of secondary metabolism in soil. *P Natl Acad Sci USA* 2014;**111**:3757–62.
- Cimermancic P, Medema MH, Claesen J et al. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* 2014;**158**:412–21.
- Cordero OX, Wildschutte H, Kirkup B et al. Ecological populations of bacteria act as socially cohesive units of antibiotic production and resistance. *Science* 2012;**337**:1228–31.
- Cox CL, Doroghazi JR, Mitchell DA. The genomic landscape of ribosomal peptides containing thiazole and oxazole heterocycles. *BMC Genomics* 2015;**16**:778.
- Cruz-Morales P, Kopp JF, Martinez-Guerrero C et al. Phylogenomic analysis of natural products biosynthetic gene clusters allows discovery of arseno-organic metabolites in model streptomycetes. *Genome Biol Evol* 2016;**8**:1906–16.
- Curtis TP, Sloan WT, Scannell JW. Estimating prokaryotic diversity and its limits. *P Natl Acad Sci USA* 2002;**99**:10494–9.
- Dejong CA, Chen GM, Li H et al. Polyketide and nonribosomal peptide retro-biosynthesis and global gene cluster matching. *Nat Chem Biol* 2016;**12**:1007–14.
- Dickschat JS. Bacterial terpene cyclases. *Nat Prod Rep* 2016;**33**:87–110.
- Donia MS, Cimermancic P, Schulze CJ et al. A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. *Cell* 2014;**158**:1402–14.
- Doroghazi JR, Albright JC, Goering AW et al. A roadmap for natural product discovery based on large-scale genomics and metabolomics. *Nat Chem Biol* 2014;**10**:963–8.
- Falony G, Joossens M, Vieira-Silva S et al. Population-level analysis of gut microbiome variation. *Science* 2016;**352**:560–4.
- Fisch KM. Biosynthesis of natural products by microbial iterative hybrid PKS–NRPS. *RSC Adv* 2013;**3**:18228–47.
- Friedrich AB, Fischer I, Proksch P et al. Temporal variation of the microbial community associated with the Mediterranean sponge *Aplysina aerophoba*. *FEMS Microbiol Ecol* 2001;**38**:105–13.
- Gibson MK, Forsberg KJ, Dantas G. Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. *ISME J* 2015;**9**:207–16.
- Grindberg RV, Ishoey T, Brinza D et al. Single cell genome amplification accelerates identification of the apratoxin biosynthetic pathway from a complex microbial assemblage. *PLoS One* 2011;**6**:e18565.
- Helfrich EJ, Piel J. Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat Prod Rep* 2016;**33**:231–316.
- Jia B, Raphenya AR, Alcock B et al. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res* 2016;**45**:566–73.
- Johnston CW, Skinnider MA, Dejong CA et al. Assembly and clustering of natural antibiotics guides target identification. *Nat Chem Biol* 2016;**12**:233–9.
- Kang HS, Charlop-Powers Z, Brady SF. Multiplexed CRISPR/Cas9 and TAR-mediated promoter engineering of natural product biosynthetic gene clusters in yeast. *ACS Synth Biol* 2016;**5**:1002–10.
- Kersten RD, Yang YL, Xu Y et al. A mass spectrometry-guided genome mining approach for natural product peptidogenomics. *Nat Chem Biol* 2011;**7**:794–802.
- Kersten RD, Ziemert N, Gonzalez DJ et al. Glycogenomics as a mass spectrometry-guided genome-mining method for microbial glycosylated molecules. *P Natl Acad Sci USA* 2013;**110**:4407–16.
- Knerr PJ, van der Donk WA. Discovery, biosynthesis, and engineering of lantipeptides. *Annu Rev Biochem* 2012;**81**:479–505.
- Leikoski N, Liu L, Jokela J et al. Genome mining expands the chemical diversity of the cyanobactin family to include highly modified linear peptides. *Chem Biol* 2013;**20**:1033–43.
- McCranie EK, Bachmann BO. Bioactive oligosaccharide natural products. *Nat Prod Rep* 2014;**31**:1026–42.
- Mak S, Xu Y, Nodwell JR. The expression of antibiotic resistance genes in antibiotic-producing bacteria. *Mol Microbiol* 2014;**93**:391–402.
- Medema MH, Breitling R, Bovenberg R et al. Exploiting plug-and-play synthetic biology for drug discovery and production in microorganisms. *Nat Rev Microbiol* 2011;**9**:131–7.
- Medema MH, Fischbach MA. Computational approaches to natural product discovery. *Nat Chem Biol* 2015;**11**:639–48.
- Medema MH, Kottmann R, Yilmaz P et al. Minimum information about a biosynthetic gene cluster. *Nat Chem Biol* 2015;**11**:625–31.
- Medema MH, Paalvast Y, Nguyen DD et al. Pep2Path: automated mass spectrometry-guided genome mining of peptidic natural products. *PLoS Comput Biol* 2014;**10**:e1003822.
- Medema MH, Trefzer A, Kovalchuk A et al. The sequence of a 1.8-mb bacterial linear plasmid reveals a rich evolutionary reservoir of secondary metabolic pathways. *Genome Biol Evol* 2010;**2**:212–24.
- Mendes R, Kruijt M, de Bruijn I et al. Deciphering the rhizosphere microbiome for disease-suppressive bacteria. *Science* 2011;**332**:1097–100.
- Mohimani H, Gurevich A, Mikheenko A et al. Dereplication of peptidic natural products through database search of mass spectra. *Nat Chem Biol* 2016;**13**:30–7.
- Mohimani H, Kersten RD, Liu WT et al. Automated genome mining of ribosomal peptide natural products. *ACS Chem Biol* 2014;**9**:1545–51.
- Montalbán-López M, Kuipers OP. Posttranslational peptide-modification enzymes in action: key roles for leaders and glutamate. *Cell Chem Biol* 2016;**23**:318–9.
- Nguyen DD, Melnik AV, Koyama N et al. Indexing the *Pseudomonas* specialized metabolome enabled the

- discovery of poaeamide B and the bananamides. *Nat Microbiol* 2016;2:16197.
- Nguyen DD, Wu CH, Moree WJ et al. MS/MS networking guided analysis of molecule and gene cluster families. *P Natl Acad Sci USA* 2013;110:E2611–20.
- Nonejuie P, Burkart M, Pogliano K et al. Bacterial cytological profiling rapidly identifies the cellular pathways targeted by antibacterial molecules. *P Natl Acad Sci USA* 2013;110:16169–74.
- Nonejuie P, Trial RM, Newton GL et al. Application of bacterial cytological profiling to crude natural product extracts reveals the antibacterial arsenal of *Bacillus subtilis*. *J Antibiot* 2016;69:353–61.
- Nurk S, Meleshko D, Korobeynikov A et al. metaSPAdes: a new versatile de novo metagenomics assembler. 2016, [arXiv:1604.03071v2](https://arxiv.org/abs/1604.03071v2)
- Ochoa JL, Bray WM, Lokey RS et al. Phenotype-guided natural products discovery using cytological profiling. *J Nat Prod* 2015;78:2242–8.
- Ogawara H. Self-resistance in *Streptomyces*, with special reference to β -lactam. *Antibiot Mol* 2016;21:605.
- Ortega MA, van der Donk, Wilfred A. New insights into the biosynthetic logic of ribosomally synthesized and post-translationally modified peptide natural products. *Cell Chem Biol* 2016;23:31–44.
- Owen JG, Reddy BV, Ternei MA et al. Mapping gene clusters within arrayed metagenomic libraries to expand the structural diversity of biomedically relevant natural products. *P Natl Acad Sci USA* 2013;110:11797–802.
- Pi B, Yu D, Dai F et al. A genomics based discovery of secondary metabolite biosynthetic gene clusters in *Aspergillus ustus*. *PLoS One* 2015;10:e0116089.
- Potts MB, Kim HS, Fisher KW et al. Using functional signature ontology (FUSION) to identify mechanisms of action for natural products. *Sci Signal* 2013;6:ra90.
- Reddy BV, Milshteyn A, Charlop-Powers Z et al. eSNaPD: a versatile, web-based bioinformatics platform for surveying and mining natural product biosynthetic diversity from metagenomes. *Chem Biol* 2014;21:1023–33.
- Sanchez JF, Chiang Y, Wang CC. Diversity of polyketide synthases found in the *Aspergillus* and *Streptomyces* genomes. *Mol Pharm* 2008;5:226–33.
- Schulze CJ, Bray WM, Woerhmann MH et al. “Function-first” lead discovery: mode of action profiling of natural product libraries using image-based screening. *Chem Biol* 2013;20:285–95.
- Shao Z, Rao G, Li C et al. Refactoring the silent spectinabilin gene cluster using a plug-and-play scaffold. *ACS Synth Biol* 2013;2:662–9.
- Sharon I, Kertesz M, Hug LA et al. Accurate, multi-kb reads resolve complex populations and detect rare microorganisms. *Genome Res* 2015;25:534–43.
- Shen B, Yan X, Huang T et al. Eneidyne: Exploration of microbial genomics to discover new anticancer drug leads. *Bioorg Med Chem Lett* 2015;25:9–15.
- Skinnder MA, Johnston CW, Edgar RE et al. Genomic charting of ribosomally synthesized natural product chemical space facilitates targeted mining. *P Natl Acad Sci USA* 2016;113:E6343–51.
- Smanski MJ, Zhou H, Claesen J et al. Synthetic biology to access and expand nature’s chemical diversity. *Nat Rev Microbiol* 2016;14:135–49.
- Strieker M, Marahiel MA. The structural diversity of acidic lipopeptide antibiotics. *ChemBioChem* 2009;10:607–16.
- Suwa M, Sugino H, Sasaoka A et al. Identification of two polyketide synthase gene clusters on the linear plasmid pSLA2-L in *Streptomyces rochei*. *Gene* 2000;246:123–31.
- Tang X, Li J, Millan-Aguinaga N et al. Identification of thiotetronic acid antibiotic biosynthetic pathways by target-directed genome mining. *ACS Chem Biol* 2015;10:2841–9.
- Traxler MF, Watrous JD, Alexandrov T et al. Interspecies interactions stimulate diversification of the *Streptomyces coelicolor* secreted metabolome. *MBio* 2013;4:e00459-13.
- Trindade-Silva AE, Rua C, Silva GG et al. Taxonomic and functional microbial signatures of the endemic marine sponge *Arenosclera brasiliensis*. *PLoS One* 2012;7:e39905.
- van Elsas JD, Chiurazzi M, Mallon CA et al. Microbial diversity determines the invasion of soil by a bacterial pathogen. *P Natl Acad Sci USA* 2012;109:1159–64.
- van Heel AJ, de Jong A, Montalban-Lopez M et al. BAGEL3: Automated identification of genes encoding bacteriocins and (non-)bactericidal posttranslationally modified peptides. *Nucleic Acids Res* 2013;41:W448–53.
- van Heel AJ, Kloosterman TG, Montalban-Lopez M et al. Discovery, production and modification of five novel lantibiotics using the promiscuous nisin modification machinery. *ACS Synth Biol* 2016;5:1146–54.
- Wang M, Carver JJ, Phelan VV et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol* 2016;34:828–37.
- Ward JM, Hodgson JE. The biosynthetic genes for clavulanic acid and cephamycin production occur as a ‘super-cluster’ in three *Streptomyces*. *FEMS Microbiol Lett* 1993;110:239–42.
- Watrous J, Roach P, Alexandrov T et al. Mass spectral molecular networking of living microbial colonies. *P Natl Acad Sci USA* 2012;109:E1743–52.
- Weber T, Blin K, Duddela S et al. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res* 2015;43:W237–43.
- Wilson MC, Mori T, Rückert C et al. An environmental bacterial taxon with a large and distinct metabolic repertoire. *Nature* 2014;506:58–62.
- Wilson MC, Piel J. Metagenomic approaches for exploiting uncultivated bacteria as a resource for novel biosynthetic enzymology. *Chem Biol* 2013;20:636–47.
- Woerhmann MH, Bray WM, Durbin JK et al. Large-scale cytological profiling for functional analysis of bioactive compounds. *Mol BioSyst* 2013;9:2604–17.
- Wolf T, Shelest V, Nath N et al. CASSIS and SMIPS: promoter-based prediction of secondary metabolite gene clusters in eukaryotic genomes. *Bioinformatics* 2016;32:1138–43.
- Xie Y, Wang B, Liu J et al. Identification of the biosynthetic gene cluster and regulatory cascade for the synergistic antibacterial antibiotics griseoviridin and viridigrisein in *Streptomyces griseoviridis*. *ChemBioChem* 2016;13:2745.
- Yamanaka K, Reynolds KA, Kersten RD et al. Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *P Natl Acad Sci USA* 2014;111:1957–62.
- Yeh H, Ahuja M, Chiang Y et al. Resistance gene-guided genome mining: serial promoter exchanges in *Aspergillus nidulans* reveal the biosynthetic pathway for fellutamide B, a proteasome inhibitor. *ACS Chem Biol* 2016;11:2275.
- Yim G, Thaker MN, Koteva K et al. Glycopeptide antibiotic biosynthesis. *J Antibiot* 2014;67:31–41.
- Zhang G, Li J, Zhu T et al. Advanced tools in marine natural drug discovery. *Curr Opin Biotechnol* 2016;42:13–23.

- Zhao X, Kuipers OP. Identification and classification of known and putative antimicrobial compounds produced by a wide variety of Bacillales species. *BMC Genomics* 2016;**17**:882.
- Zhernakova A, Kurilshikov A, Bonder MJ et al. Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science* 2016;**352**:565–9.
- Ziemert N, Alanjary M, Weber T. The evolution of genome mining in microbes—a review. *Nat Prod Rep* 2016;**33**:988–1005.
- Zipperer A, Konnerth MC, Laux C et al. Human commensals producing a novel antibiotic impair pathogen colonization. *Nature* 2016;**535**:511–6.