Exploration and exploitation of the environment for novel specialized metabolites

Loureiro, C., Medema, M. H., van der Oost, J., & Sipkema, D.

This is a "Post-Print" accepted manuscript, which has been published in "Current Opinion in Biotechnology"

Please cite this publication as follows:

Loureiro, C., Medema, M. H., van der Oost, J., & Sipkema, D. (2018). Exploration and exploitation of the environment for novel specialized metabolites. Current Opinion in Biotechnology, 50, 206-213. DOI: 10.1016/j.copbio.2018.01.017

You can download the published version at:

https://doi.org/10.1016/j.copbio.2018.01.017

Current Opinion in Biotechnology **50**, 203-213 (2018) [Author's copy]

Exploration and Exploitation of the Environment for novel Specialized Metabolites

Catarina Loureiro[1], Marnix H. Medema[2], John van der Oost[1], Detmer Sipkema[1*]

[1]Wageningen University & Research, Laboratory of Microbiology, Stippeneng 4, 6708 WE Wageningen, The Netherlands
[2]Wageningen University & Research, Bioinformatics Group, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands
[*]Corresponding author: detmer.sipkema@wur.nl

Abstract

Microorganisms are Nature's little engineers of a remarkable array of bioactive small molecules that represent most of our new drugs. The wealth of genomic and metagenomic sequence data generated in the last decade has shown that the majority of novel biosynthetic gene clusters (BGCs) is identified from cultivation-independent studies, which has led to a strong expansion of the number of microbial taxa known to harbour BGCs. The large size and repeat sequences of BGCs remain a bioinformatic challenge, but newly developed software tools have been created to overcome these issues and are paramount to identify and select the most promising BGCs for further research and exploitation. While heterologous expression of BGCs has been the greatest challenge until now, a growing number of polyketide synthase (PKS) and non-ribosomal peptide synthetase (NRPS)-encoding gene clusters have been cloned and expressed in bacteria and fungi based on techniques that mostly rely on homologous recombination. Finally, combining ecological insights with state-of-the-art computation and molecular methodologies will allow for further comprehension and exploitation of microbial specialized metabolites.

Introduction

Microorganisms are unparalleled with respect to the chemical diversity of specialized metabolites they produce. These encompass many chemical classes including polyketides (PKs), non-ribosomal peptides (NRPs), ribosomally synthesized and post translationally modified peptides (RiPPs), terpenes, saccharides and alkaloids [1]. Until the 1950s the majority of microbial metabolites were overlooked or merely regarded as waste products from primary metabolism. In contrast to a general set of primary metabolites, specialized metabolites are often specific to a restricted taxonomic range where they facilitate dedicated physiological, social or predatory functions [2]. Moreover, such metabolites have been found to possess a wide range of biological activities, making them useful for the development of antimicrobials, anticancer agents and immunosuppressants for pharmaceutical, agricultural and food manufacturing applications [3–6].

The majority of specialized metabolites result from metabolic pathways, each of which encoded by a suite of genes at the same chromosomal locus, generally known as biosynthetic gene clusters (BGCs). These BGCs are frequently "silent" in common laboratory media, whereas their expression is triggered by specific environmental cues [7–9]. Recent developments in genomics and computational biology, hand in hand with a vastly increasing number of sequenced metagenomes and metatranscriptomes, have led to the discovery of thousands of BGCs [10,11].

Modular assembly lines such as PK synthases (PKS) and NRP synthetases (NRPS) constitute two of the most important and diverse classes of specialized metabolites that can theoretically code for a near infinite diversity of unique molecular architectures [12–14]. Recent analyses based on retro-biosynthesis, i.e. the computational breakdown of PK and NRP chemical molecules and reversal of their assembly lines to predict their parent PKS/NRPS BGCs, allow linking BGCs from publicly available databases to known natural products and define clusters encoding new products. Such efforts have shown that thousands of BGCs are likely to be responsible for the production of novel molecules [10].

To prevent replication of previous research and yet discover specialized metabolites from microbes with novel applicable biological activities, it is important to shift attention to environments and microbial phyla that have so far been largely neglected. Moreover, advanced bioinformatics analyses must be applied that can

quickly assess the novelty of the gene clusters found and link them to predicted chemical structures and biological activities.

In this opinion paper, we highlight state-of-the-art developments regarding discovery, characterization and exploitation of microbial specialized metabolites, with a focus on PKS and NRPS. In addition, we identify environments, bioinformatics approaches and expression strategies that we consider most promising for future development of the field [Figure 1].

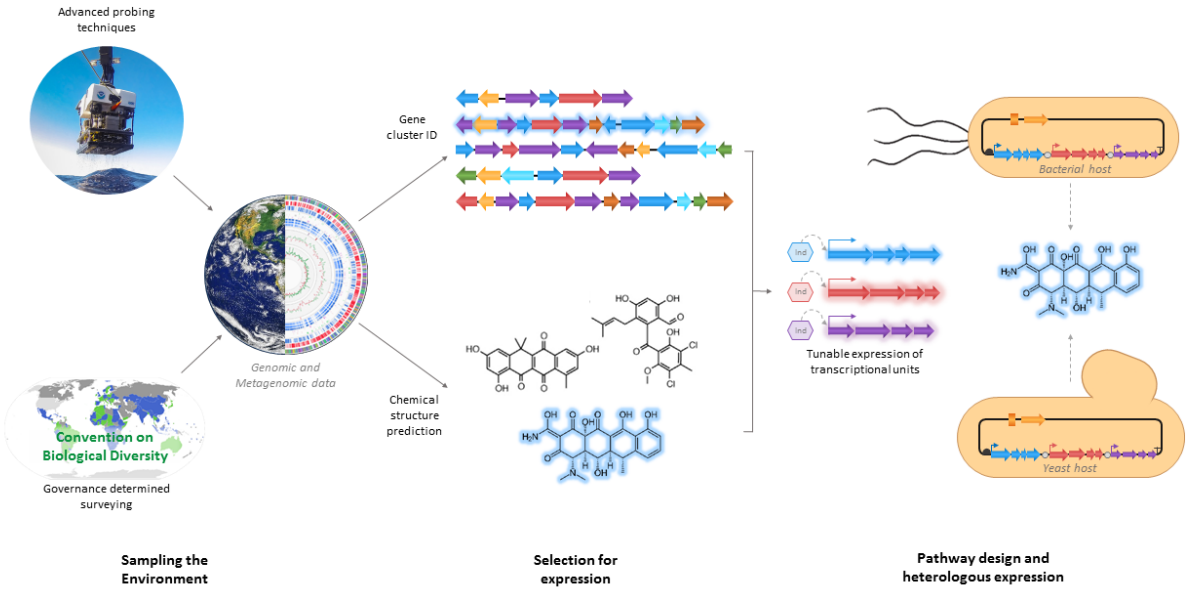

Figure 1. Approach for specialized metabolite discovery.
Microbial specialized metabolites are of great value, and in order to boost their discovery, exploration of scarcely screened environments is key. Technological advances in sampling tools and techniques play an important role in allowing researchers to access such locations.. At the same time, governmental constraints also dictate which regions will be favoured for exploration and exploitation of microbial bioactives.
Newly developed computational methodologies enable mining of genomic and metagenomic data for detection of potentially new classes of biosynthetic gene clusters (BGCs).These algorithms are optimized to conduct identification of BGCs and predict their chemical structures, and are crucial to identify and select the most promising BGCs for further research and exploitation.
The next step in unlocking and systematically exploiting these BGCs involves their controlled expression. Large DNA molecule manipulation involves assembly and cloning methods often based on homologous recombination mechanisms in both yeast and bacteria. Furthermore, advances in synthetic biology allowing customisation of transcriptional units' expression stoichiometry for production of complex chemicals, play an important role in the creation of automated production platforms.

Environmental sources of specialized metabolites

Nature has provided mankind with numerous bioactive compounds for medical purposes for thousands of years, and even in modern times most drugs are derived from natural sources [15]. Bacteria and fungi that are responsible for the production of small bioactive molecules have been found in widely diverse environmental niches, such as soil, sediment and aquatic environments, either as free-living microorganisms or in symbiosis with plants and animals [15,16]. Soil-dwelling cultivable Actinobacteria, and members of the genus *Streptomyces* in particular, have been in the limelight as proliferous sources of specialized bioactive metabolites, as witnessed by the discoveries of the antibiotics actinomycin, streptomycin and chloramphenicol in the 1940s, and the antiparasitic agent ivermectin [17–19]. Also soil-derived isolates from other bacterial genera, such as *Bacillus* [20] and *Pseudomonas* [12,21] are traditionally rich sources of specialized

metabolites. Interestingly, there appear to be important differences in biosynthetic potential between taxonomic groups within these genera, according to their ecological specializations [5,22]. Fungi, historically also mainly isolated from soils, represent a sometimes overlooked, but prolific source of bioactive molecules (e.g. antibiotics such as penicillin) [5,23]. A recently published study explored the environmental factors that drive changes in PKS and NRPS encoding BGC diversity across geographically distinct soil environments, and found changes in biosynthetic domain composition to correlate most consistently with variations in latitude[24].

However, cultivation-independent methods have shown that the uncultivated majority of the microorganisms encode many more BGCs (quantitatively and qualitatively) than the ones we know from isolates, a terra incognita with major potential for applications [4,5]. In addition, the use of these cultivation-independent methods shows that the traditional focus on Actinobacteria needs a shift towards other microorganisms such as marine fungi [25], Cyanobacteria[26,27], Proteobacteria[28] and the novel candidate phylum Tectomicrobia [29,30]. For example, the latter, represented by a newly discovered uncultivated marine sponge symbiont genus Candidatus Entotheonella which has the genetic capacity to produce over 40 natural compounds and is widely distributed in taxonomically diverse sponges [30]. Other microbial taxa including *Clostridium*, *Planctomycetes*, *Burkholderia* and *Xenorhabdus/Photorhabdus* are also emerging important targets with high biosynthetic potential [5,22].

While the terrestrial environment is by no means exhausted of novel bioactive molecules, a recent large metagenomics study of the ocean water revealed that a stunning 90% of the genes detected at a depth of 600 m did not have a match in public databases [31]. Although the ocean metagenome appeared to be rich in BGCs, we propose that the majority of BGCs in ocean water remains undetected as only the fraction <3 µm was considered in the aforementioned study, excluding small particles that are colonised by a community of microorganisms. These in turn are more likely to produce specialized metabolites of interest required for short-range molecular interactions. The same may be expected for marine sediments and biofilms (e.g., on macroalgae) that have been poorly investigated for their potential to produce specialized metabolites [32,33]. In addition, marine invertebrates display species-specific symbioses with microorganisms facilitated by unique metabolites, some of which may be valuable bioactive small molecules [34]. Particularly sponges, the biomass of which may be almost equally divided between host and symbionts, have been identified as one of the most promising natural source for future antibiotics [35,36].

In addition to differences in the resource potential of particular natural environments, the governmental situation may dictate which regions will be favoured for exploration and exploitation of microbial bioactives. Compliance with the Nagoya Protocol requires the explorer to legally acquire any genetic resource, prove due diligence through traceability, risk assessment and risk mitigation procedures, and enable inspections by national authorities. Each signatory state may either determine its own access policy, or provide free access to its genetic resources and associated traditional knowledge (www.cbd.int/abs). However, concepts of biological diversity that are the foundation of the Nagoya Protocol are not directly applicable for microorganisms that do not abide the same patterns of endemism as plants and animals [37]. For example, *Streptomyces carpaticus* strains isolated from coastal habitats in four different continents all produced the same cytotoxic specialized metabolite (Ikarugamycin)[38]; to a large degree, ‘everything is everywhere’ where the environment selects for the same molecular functions [39]. In addition, structurally very similar polyketides have been obtained from bacterial symbionts from either insects or sponges [40]. Therefore, countries that have a more open attitude and lower administrative burden towards scientific exploration and commercial application of microbial specialized metabolites will likely be more attractive for scientists and industries.

Rapid identification and prioritizing specialized metabolites

In recent years, genome mining for BGCs has become a key approach for identification of new molecules and corresponding novel products. For compounds produced by PKSs and NRPSs, their biosynthetic pathways and product structures can be predicted using a range of computational tools and approaches [Table 1] [41–47].

The ability to detect potentially new classes of BGCs, including those prevailing in the uncultured majority of microorganisms, is a valuable endeavour as these will most likely code for molecules with new chemical scaffolds [51,52]. Tools such as ClusterFinder [11,53] and EvoMining [48] have been developed for this purpose. The former queries genome sequences for BGC-like regions based on the presence of broad

Pfam protein domains associated with enzyme families commonly recycled in diverse specialized metabolic pathways. The latter exploits the notion that enzymes involved in specialized metabolism are paralogs of primary ones, which have undergone sequence and functional divergence, and utilizes phylogenetic analyses to detect these outliers [47,54]. Recent developments in high-throughput, single cell and long-read next-generation sequencing technologies are leading us to an era of fast, affordable sequencing and assembly of genomes from microbial isolates/consortia. Thus, it is becoming increasingly feasible to access culturable bacterial taxa and obtain high-quality genomes from these strains, despite the presence of repetitive genomic regions such as those including BGCs encoding NRPS and PKS enzymes [22,47,54–57].

Table 1. Tools for identification of BGCs

Tool	Approach	Reference
antiSMASH	Identification or signature genomic and protein domains that are hallmarks of biosynthetic pathways. Usually making use of profile Hidden Markov Models (HMM), BLAST and both general databases like Pfam and specialized PKS/NRPS databases for annotation and protein identification.	[46]
SMURF		[41]
PRISM		[43]
NP.searcher		[45]
CLUSEAN		[44]
EvoMining	Exploits the notion that enzymes involved in specialized metabolism are paralogs of primary ones, which have undergone sequence and functional divergence to acquire functions in specialized metabolism. Utilizes phylogenetic mining to detect these outliers.	[48]
ClusterFinder	Queries genome sequences for BGC-like regions based on the presence of Pfam protein domain frequencies associated with enzyme families that are indicative of diverse specialized metabolic pathways.	[11]
GRAPE	Retro-biosynthesis, i.e. computational deconstruction of PK and NRP chemical structures to predict their parent PKS/NRPS, producing assembly line monomers and tailoring enzymes.	[49]
Bandage	Tool for visualizing de Bruijn assembly graphs, allows for a deeper analysis of *de novo* assemblies which is not accessible through study of individual contigs.	[50]

Moreover, through direct capture of environmental DNA from microbiomes of macroorganisms, metagenomics allows efficiently moving biosynthetic diversity from the environment into the drug discovery space [57]. PCR-based sequence tag approaches that screen metagenomic libraries for biosynthetic novelty are considered well established technologies [4,57,58]. However, despite being plagued by issues related to acquiring highly contiguous assemblies of BGCs, sequencing and assembly of environmental DNA by shotgun metagenomics constitutes a much more unbiased approach to profile biosynthetic diversity [22]. Whereas artificial long-read technologies offer valuable improvements in assembly quality [59–61], PK and NRP BGCs are usually still hard to assemble and often remain fragmented across multiple contigs. Nevertheless, contigs generated by De Bruijn Graph assembly algorithms are not islands on their own, but in fact are connected to other contigs in an assembly graph. Although information contained in the assembly graph is lost in the way sequence assemblies are usually studied, the assembly graph files themselves can be analyzed with visualization software tools like Bandage[50]. By performing BLAST similarity searches on such a graph, one can often derive which BGC fragments belong to the same gene cluster. Based on this, clusters can potentially be reconstructed by finding the most plausible path through the assembly graph based on homology inference (as recently done for the bananamide BGC in a fragmented draft genome of *Pseudomonas fluorescens* BW11P2 [62] [Figure 2] [21]) or otherwise by designing primers to amplify and Sanger-sequence the gaps between the contigs. Alternatively, long read nanopore sequencing can also be used[63]. Additionally, binning metagenomes into metagenome-assembled genomes (MAGs) based on differential coverage and oligonucleotide frequencies, and subsequently re-assembling/finishing of high-quality MAGs allows increasing the contiguity of the assembly for particular organisms within a microbial community [64].

These and other computational methodologies are now making it possible to assemble many complete biosynthetic gene clusters from relatively complex metagenomes.The prediction of natural product structures from gene clusters is still challenging as deviations in gene order and enzyme modularity occur frequently [49], and predicting the regioselectivity of tailoring reactions is very complicated. Nevertheless, computational dereplication strategies based on sequence similarity [1] or retro-biosynthesis [49] make it possible to reliably identify BGCs that are likely to be involved in the production of novel chemical scaffolds. Moreover, target-

based genome mining based on the detection of resistance genes within BGCs [65] makes it possible to pinpoint 'low-hanging fruits' that are likely to be responsible for the production of molecules that bind to cellular targets of interest, as the resistance genes often constitute paralogous copies of these molecular targets that are insensitive to the product of the BGC. Based on such and other criteria, at least a sub-set of BGCs can be intelligently shortlisted for expression studies.

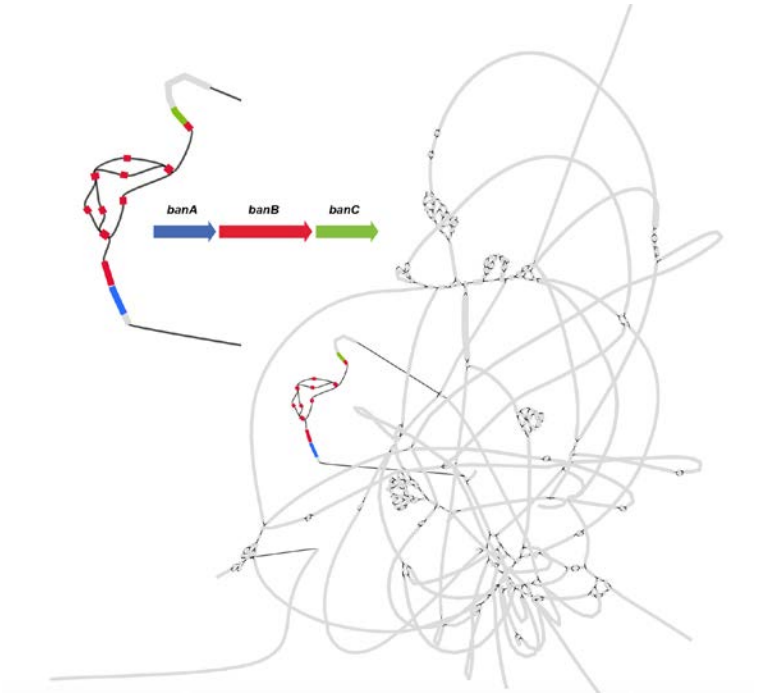

Figure 2. Assembly graph of a fragmented draft genome of *Pseudomonas fluorescens* BW11P2[54], assembled by SPAdes, containing the reconstructed bananamide BGC. In the graph, the grey lines represent nodes (contigs) and black lines paths that represent possible connections between contigs. The upper left corner of the panel depicts a zoom-in visualization for the BLAST result of the genes in this BGC, blue, red and green represent the BLAST hits for gene banA, banB and banC correspondingly. Co-localization on the same node of banA and part of banB indicates proximity of these genes on the genome.

Heterologous expression strategies for specialized metabolites

Biodiversity profiling of different environmental niches provides an outline of the phylogenetic composition of the corresponding communities, and demonstrates that uncultured species outnumber their cultured counterparts. Therefore, the quest to functionally express BGCs is currently the most urgent issue to unlock and exploit these gene clusters. However, this is not a straightforward undertaking. Firstly, because many BGCs are found in non-model organisms, often with rather distinct codon usage to general production hosts such as *E. coli*. Secondly, they are often encoded by clusters that can span over 100 kb of DNA, possibly including complex regulatory mechanisms [66]. Nevertheless, several methods have been developed allowing PKS and NRPS gene clusters to be successfully cloned and expressed in bacteria and fungi [67–74].

DNA assembly methods, such as transformation-associated recombination (TAR) cloning are powerful tools for manipulating large DNA molecules. TAR makes use of homologous recombination in yeast and it has been successfully applied to clone and express the 73kb gene cluster encoding the antibiotic taromycin A, originating from a marine actinomycete [72]. Furthermore, a number of direct cloning methods allow integration of gene clusters at specific sites in the production host's chromosome, mainly via standard recombination methods. Direct cloning via Red/ET recombineering is based on *E. coli* linear plus linear homologous recombination [75], and has been sucessfully used to express large biosynthetic pathways such as the NPRS clusters coding for edeine (48.3kb) and bacillomycin (37.3kb) [70].

Advances in synthetic biology (including DNA construction tools, synthetic regulatory circuits and multiplexed genome engineering) enable the harnessing of metagenomic data for high-throughput molecular discovery, as well as pathway design for the production of complex chemicals [66,74,76]. In at least one instance, using a plug-and-play DNA assembly strategy to achieve full gene cluster refactoring in a single step manner has proved more effective than direct cloning and promoter insertion. This also made it possible to construct an automated platform with a high degree of flexibility for generating gene deletions or additions [77]. As a proof of principle for this approach, Luo and colleagues succeeded in expressing and characterizing a cryptic BCG encoding for the production of a polycyclic tretamate macrolactam PKS-NRPS hybrid [77]. With the continuous decrease in DNA synthesis cost, synthetic (codon-optimized) versions of many BGCs can be reconstructed in high-throughput using this technology.

One key issue that is difficult to address, especially for gene clusters for which the real molecular product is unknown beforehand, is that of cross-talk between the heterologously expressed pathway and the native pathways. A recent study by Zhang et al. showed that heterologous expression of the lyngbyatoxin gene cluster in three different streptomycete hosts lead to the generation of different natural product derivatives [78]. Because small variations in chemical structure can have a major impact on biological activity, expression studies in multiple hosts (or multiple versions of the same hosts with different native BGCs knocked out using e.g. CRISPR/Cas) are required to ascertain the true product of a synthetically reconstructed BGC for which the native product is unknown. In the more distant future, it might be worth considering to design 'orthogonal' heterologous expression strategies that isolate the heterologous pathway from native metabolism, e.g. through compartmentalization [79,80].

Outlook

Microbial specialized metabolites are a vast and exceptional resource that may contribute to solving the current antibiotic resistance crisis [19,81,82]. Based on several technological advances, it is now possible to reach and sample the most difficult-to-access places on Earth. Exploration of scarcely touched environments in combination with the revolutionary developments in metagenomics and computational biology has already led to an explosion in the number of known BGC sequences. Our greatest current challenge is to systematically use these sequences for the production of specialized metabolites and the discovery of their biological functions. Nonwithstanding, we have witnessed a growing number of success cases in the past decade, including the activation or heterologous expression of cognate BGCs from non-model organisms leading to the successful production of several previously unknown secondary metabolites. Ultimately, the implementation of multi-omics approaches that combine ecological insights with state-of-the-art computational and molecular genomics developments will lead to deep understanding and more efficient exploitation of microbial specialized metabolism.

Conflict of interest

None

Acknowledgements

C.L was funded by NWO-VLAG grant *Mare incognita*, M.H.M. was supported by VENI grant 863.15.002 from The Netherlands Organization for Scientific Research (NWO), J.v.d.O. is supported by the Netherlands Organization for Scientific Research (NWO) by a TOP grant (714.015.001) and D.S. acknowledges funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No. 679849 (SponGES).

References

1. • Medema MH: **The Minimum Information about a Biosynthetic Gene cluster (MIBiG) specification**. *Nat Chem Biol* 2015, **revised ve**:625–631.
2. Hartmann T: **The lost origin of chemical ecology in the late 19th century**. *Proc Natl Acad Sci U S A* 2008, **105**:4541–4546.
3. Medema MH, Fischbach MA: **HHS Public Access**. *Nat Chem Biol* 2016, **11**:639–648.
4. Milshteyn A, Schneider JS, Brady SF: **Mining the metabiome: Identifying novel natural products from microbial communities**. *Chem Biol* 2014, **21**:1211–1223.
5. Challinor VL, Bode HB: **Bioactive natural products from novel microbial sources**. 2015, **1354**:82–97.
6. Harvey AL, Edrada-Ebel R, Quinn RJ: **The re-emergence of natural products for drug discovery in the genomics era**. *Nat Rev Drug Discov* 2015, **14**:111–129.
7. Van der Meij A, Worsley SF, Hutchings MI, Van Wezel GP: **Chemical ecology of antibiotic production by**

actinomycetes. *FEMS Microbiol Rev* 2017, **41**:392–416.
8. Abdelmohsen UR, Grkovic T, Balasubramanian S, Kamel MS, Quinn RJ, Hentschel U: **Elicitation of secondary metabolism in actinomycetes**. *Biotechnol Adv* 2015, **33**:798–811.
9. Rutledge PJ, Challis GL: **Discovery of microbial natural products by activation of silent biosynthetic gene clusters**. *Nat Rev Microbiol* 2015, **13**:509–523.
10. •• Dejong CA, Chen GM, Li H, Johnston CW, Edwards MR, Rees PN, Skinnider MA, Webster ALH, Magarvey NA: **Polyketide and nonribosomal peptide retro-biosynthesis and global gene cluster matching**. *Nat Chem Biol* 2016, **12**:1007–1014.
11. •• Cimermancic P, Medema MH, Claesen J, Kurita K, Wieland Brown LC, Mavrommatis K, Pati A, Godfrey PA, Koehrsen M, Clardy J, et al.: **Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters**. *Cell* 2014, **158**:412–421.
12. Helfrich EJN: **Biosynthesis of polyketides by trans-AT polyketide synthases**. 2016, doi:10.1039/C5NP00125K.
13. Doroghazi JR, Metcalf WW: **Comparative genomics of actinomycetes with a focus on natural product biosynthetic genes.** *BMC Genomics* 2013, **14**:611.
14. •• Rees P, Johnston CW, Skinnider MA, Dejong CA, Rees PN, Chen GM, Walker CG, French S, Brown ED, Bérdy J, et al.: **Assembly and clustering of natural antibiotics guides target identification**. 2018, doi:10.1038/nchembio.2018.
15. Newman DJ, Cragg GM: **Natural Products As Sources of New Drugs over the 30 Years from 1981 to 2010**. 2012,
16. Pi B, Yu D, Dai F, Song X, Zhu C, Li H: **A Genomics Based Discovery of Secondary Metabolite Biosynthetic Gene Clusters in Aspergillus ustus**. 2015, doi:10.1371/journal.pone.0116089.
17. Barka EA, Vatsa P, Sanchez L, Gaveau-vaillant N, Jacquard C, Klenk H-P, Clément C, Ouhdouch Y, van Wezel GP: **Taxonomy, Physiology, and Natural Products of Actinobacteria**. *Microbiol Mol Biol Rev* 2016, **80**:1–43.
18. Campbell WC, Fisher MH, Stapley EO, Albers-Schönberg G, Jacob TA: **Ivermectin: A Potent New Antiparasitic Agent**. *Science (80-)* 1983, **221**:823–828.
19. • Ling LL, Schneider T, Peoples AJ, Spoering AL, Engels I, Conlon BP, Mueller A, Hughes DE, Epstein S, Jones M, et al.: **A new antibiotic kills pathogens without detectable resistance**. *Nature* 2015, **517**:455–459.
20. Zhao X, Kuipers OP: **Identification and classification of known and putative antimicrobial compounds produced by a wide variety of Bacillales species**. *BMC Genomics* 2016, doi:10.1186/s12864-016-3224-y.
21. • Nguyen DD, Melnik A V., Koyama N, Lu X, Schorn M, Fang J, Aguinaldo K, Lincecum TL, Ghequire MGK, Carrion VJ, et al.: **Indexing the *Pseudomonas* specialized metabolome enabled the discovery of poaeamide B and the bananamides.** *Nat Microbiol* 2016, **2**:16197.
22. Tracanna V, de Jong A, Medema MH, Kuipers OP: **Mining prokaryotes for antimicrobial compounds: From diversity to function**. *FEMS Microbiol Rev* 2017, **41**:417–429.
23. Alberti F, Foster GD, Bailey AM: **Natural products from filamentous fungi and production by heterologous expression**. *Appl Microbiol Biotechnol* 2017, **101**:493–500.
24. Lemetre C, Maniko J, Charlop-Powers Z, Sparrow B, Lowe AJ, Brady SF: **Bacterial natural product biosynthetic domain composition in soil correlates with changes in latitude on a continent-wide scale**. *Proc Natl Acad Sci* 2017, **114**:201710262.
25. Nicoletti R, Trincone A: *Bioactive compounds produced by strains of Penicillium and Talaromyces of marine origin*. 2016.
26. Banik JJ, Brady SF: **Recent application of metagenomic approaches toward the discovery of antimicrobials and other bioactive small molecules.** *Curr Opin Microbiol* 2010, **13**:603–9.
27. Long PF, Dunlap WC, Battershill CN, Jaspars M: **Shotgun cloning and heterologous expression of the patellamide gene cluster as a strategy to achieving sustained metabolite production**. *ChemBioChem* 2005, **6**:1760–1765.
28. Desriac F, Jégou C, Balnois E, Brillet B, Le Chevalier P, Fleury Y: **Antimicrobial peptides from marine proteobacteria**. *Mar Drugs* 2013, **11**:3632–3660.
29. Martins A, Vieira H, Gaspar H, Santos S: **Marketed marine natural products in the pharmaceutical and cosmeceutical industries: Tips for success**. *Mar Drugs* 2014, **12**:1066–1101.
30. •• Wilson MC, Mori T, Rückert C, Uria AR, Helf MJ, Takada K, Gernert C, Steffens U a E, Heycke N, Schmitt S, et al.: **An environmental bacterial taxon with a large and distinct metabolic repertoire.** *Nature* 2014, **506**:58–62.
31. Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, Djahanschiri B, Zeller G, Mende DR, Alberti A, et al.: **Structure and function of the global ocean microbiome**. *Science (80-)* 2015, **348**:1261359.
32. Patin N V., Schorn M, Aguinaldo K, Lincecum T, Moore BS, Jensen PR: **Effects of actinomycete secondary metabolites on sediment microbial communities**. *Appl Environ Microbiol* 2017, **83**:1–16.
33. Niederdorfer R, Besemer K, Battin TJ, Peter H: **Ecological strategies and metabolic trade-offs of complex environmental biofilms**. *npj Biofilms Microbiomes* 2017, **3**:21.
34. Blunt JW, Copp BR, Keyzers RA, Munro MHG, Prinsep MR: **Marine natural products**. *Nat Prod Rep* 2017, **34**:235–294.
35. Fortman JL, Mukhopadhyay A: **The Future of Antibiotics: Emerging Technologies and Stewardship**. *Trends Microbiol* 2017, **24**:515–517.
36. Sipkema D: **Marine biotechnology: diving deeper for drugs**. *Microb Biotechnol* 2017, **10**:7–8.
37. • Overmann J, Scholz AH: **Microbiological Research Under the Nagoya Protocol: Facts and Fiction**. *Trends Microbiol* 2017, **25**:85–88.
38. Overmann J: **Significance and future role of microbial resource centers**. *Syst Appl Microbiol* 2015, **38**:258–265.
39. De Wit R, Bouvier T: **"Everything is everywhere, but, the environment selects"; what did Baas Becking and Beijerinck really say?** *Environ Microbiol* 2006, **8**:755–758.
40. Piel J, Hui D, Wen G, Butzke D, Platzer M, Fusetani N, Matsunaga S: **Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge Theonella swinhoei.** *Proc Natl Acad Sci U S A* 2004, **101**:16222–16227.
41. Khaldi N, Seifuddin FT, Turner G, Haft D, Nierman WC, Wolfe KH, Fedorova ND: **SMURF: Genomic mapping of fungal secondary metabolite clusters**. *Fungal Genet Biol* 2010, **47**:736–741.
42. Starcevic A, Zucko J, Simunkovic J, Long PF, Cullum J, Hranueli D: **ClustScan: An integrated program package for the semi-automatic annotation of modular biosynthetic gene clusters and in silico prediction of novel chemical structures**. *Nucleic Acids Res* 2008, **36**:6882–6892.
43. Skinnider MA, Dejong CA, Rees PN, Johnston CW, Li H, Webster ALH, Wyatt MA, Magarvey NA: **Genomes to natural products PRediction Informatics for Secondary Metabolomes (PRISM)**. *Nucleic Acids Res* 2015, **43**:9645–9662.
44. Weber T, Rausch C, Lopez P, Hoof I, Gaykova V, Huson DH, Wohlleben W: **CLUSEAN: A computer-based framework for the automated analysis of bacterial secondary metabolite biosynthetic gene clusters**. *J Biotechnol* 2009, **140**:13–17.

45. Li MHT, Ung PMU, Zajkowski J, Garneau- S, Sherman DH: **Automated genome mining for natural products**. 2009, **10**:1–10.
46. Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, Lee SY, Fischbach MA, M??ller R, Wohlleben W, et al.: **AntiSMASH 3.0-A comprehensive resource for the genome mining of biosynthetic gene clusters**. *Nucleic Acids Res* 2015, **43**:W237–W243.
47. Medema MH, Fischbach MA: **Computational approaches to natural product discovery**. 2016, **11**:639–648.
48. Cruz-Morales P, Kopp JF, Mart??nez-Guerrero C, Y????ez-Guerra LA, Selem-Mojica N, Ramos-Aboites H, Feldmann J, Barona-G??mez F: **Phylogenomic Analysis of Natural Products Biosynthetic Gene Clusters Allows Discovery of Arseno-Organic Metabolites in Model Streptomycetes**. *Genome Biol Evol* 2016, **8**:1906–1916.
49. Dejong CA, Chen GM, Li H, Johnston CW, Edwards MR, Rees PN, Skinnider MA, Webster ALH, Magarvey NA: **Polyketide and nonribosomal peptide retro- biosynthesis and global gene cluster matching**. 2016, doi:10.1038/nchembio.2188.
50. Wick RR, Schultz MB, Zobel J, Holt KE: **Bandage: Interactive visualization of de novo genome assemblies**. *Bioinformatics* 2015, **31**:3350–3352.
51. Fischbach MA, Walsh CT: **Antibiotics for Emerging Pathogens**. *Science (80-)* 2009, **325**:1089–1093.
52. • Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J, Darling A, Malfatti S, Swan BK, Gies EA, et al.: **Insights into the phylogeny and coding potential of microbial dark matter**. *Nature* 2013, **499**:431–437.
53. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, et al.: **The Pfam protein families database**. 2012, **40**:290–301.
54. Ziemert N, Alanjary M, Weber T: **The evolution of genome mining in microbes – a review**. *Nat Prod Rep* 2016, **33**:988–1005.
55. Klassen JL, Currie CR: **Gene fragmentation in bacterial draft genomes: extent , consequences and mitigation**. 2012,
56. Epstein SS: **The phenomenon of microbial uncultivability**. *Curr Opin Microbiol* 2013, **16**:636–642.
57. Charlop-powers Z, Milshteyn A, Brady SF: **Metagenomic small molecule discovery methods**. *Curr Opin Microbiol* 2014, **19**:70–75.
58. Charlop-Powers Z, Pregitzer CC, Lemetre C, Ternei MA, Maniko J, Hover BM, Calle PY, McGuire KL, Garbarino J, Forgione HM, et al.: **Urban park soil microbiomes are a rich reservoir of natural product biosynthetic diversity.** *Proc Natl Acad Sci U S A* 2016, **113**:14811–14816.
59. • White RA, Bottos EM, Roy Chowdhury T, Zucker JD, Brislawn CJ, Nicora CD, Fansler SJ, Glaesemann KR, Glass K, Jansson JK: **Moleculo Long-Read Sequencing Facilitates Assembly and Genomic Binning from Complex Soil Metagenomes**. *mSystems* 2016, **1**:e00045-16.
60. Kuleshov V, Jiang C, Zhou W, Jahanbani F, Batzoglou S, Snyder M: **Synthetic long-read sequencing reveals intraspecies diversity in the human microbiome**. *Nat Biotechnol* 2015, **34**.
61. Moss E, Bishara A, Tkachenko E, Kang JB, Andermann TM, Wood C, Handy C, Ji H, Batzoglou S, Bhatt AS: **De novo assembly of microbial genomes from human gut metagenomes using barcoded short read sequences**. *bioRxiv* 2017, doi:10.1101/125211.
62. Nguyen DD, Melnik A V, Koyama N, Lu X, Schorn M, Fang J, Aguinaldo K, Lincecum Jr TL, Ghequire MGK, Carrion VJ, et al.: **Indexing the Pseudomonas specialized metabolome enabled the discovery of poaeamide B and the bananamides**. 2016, **2**:16197.
63. Loman NJ, Quick J, Simpson JT: **A complete bacterial genome assembled de novo using only nanopore sequencing data**. *Nat Meth* 2015, **12**:733–735.
64. Anantharaman K, Brown CT, Hug LA, Sharon I, Castelle CJ, Probst AJ, Thomas BC, Singh A, Wilkins MJ, Karaoz U, et al.: **Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system**. *Nat Commun* 2016, **7**:13219.
65. Alanjary M, Kronmiller B, Adamek M, Blin K, Weber T, Huson D, Philmus B, Ziemert N: **The Antibiotic Resistant Target Seeker (ARTS), an exploration engine for antibiotic cluster prioritization and novel drug target discovery**. *Nucleic Acids Res* 2017, **22**:43–51.
66. Smanski MJ, Zhou H, Claesen J, Shen B, Fischbach MA, Voigt CA: **Synthetic biology to access and expand nature's chemical diversity**. *Nat Rev Microbiol* 2016, **14**:135–149.
67. Du D, Wang L, Tian Y, Liu H, Tan H, Niu G: **Genome engineering and direct cloning of antibiotic gene clusters via phage φBT1 integrase-mediated site-specific recombination in *Streptomyces*.** *Sci Rep* 2015, **5**:8740.
68. Ishiuchi K, Nakazawa T, Ookuma T, Sugimoto S, Sato M, Tsunematsu Y, Ishikawa N, Noguchi H, Hotta K, Moriya H, et al.: **Establishing a New Methodology for Genome Mining and Biosynthesis of Polyketides and Peptides through Yeast Molecular Genetics**. *ChemBioChem* 2012, **13**:846–854.
69. Chiang Y-M, Oakley CE, Ahuja M, Entwistle R, Schultz A, Chang S-L, Sung CT, Wang CCC, Oakley BR: **An Efficient System for Heterologous Expression of Secondary Metabolite Genes in Aspergillus nidulans**. *J Am Chem Soc* 2013, **135**:7720–7731.
70. Liu Q, Shen Q, Bian X, Chen H, Fu J, Wang H, Lei P, Guo Z, Chen W, Li D, et al.: **Simple and rapid direct cloning and heterologous expression of natural product biosynthetic gene cluster in *Bacillus subtilis* via Red/ET recombineering**. *Sci Rep* 2016, **6**:34623.
71. Tu Q, Herrmann J, Hu S, Raju R, Bian X: **Genetic engineering and heterologous expression of the disorazol biosynthetic gene cluster via Red / ET recombineering**. *Nat Publ Gr* 2016, doi:10.1038/srep21066.
72. •• Yamanaka K, Reynolds KA, Kersten RD, Ryan KS, Gonzalez DJ, Nizet V, Dorrestein PC, Moore BS: **Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A.** *Proc Natl Acad Sci U S A* 2014, **111**:1957–62.
73. Shao Z, Zhao H, Zhao H: **DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways**. *Nucleic Acids Res* 2009, **37**:1–10.
74. Luo Y, Enghiad B, Zhao H: **New tools for reconstruction and heterologous expression of natural product biosynthetic gene clusters**. *Nat Prod Rep* 2016, **0**:1–9.
75. Muyrers JPP, Zhang Y, Stewart AF: **Techniques: Recombinogenic engineering–new options for cloning and manipulating DNA**. *Trends Biochem Sci* 2001, **26**:325–331.
76. Reports NP, Andrews S, Scienc I, Rese P: **The generation of " unNatural " products: Synthetic biology meets synthetic chemistry**. 2012, doi:10.1039/c2np00001f.
77. •• Luo Y, Huang H, Liang J, Wang M, Lu L, Shao Z, Cobb RE, Zhao H: **Activation and characterization of a cryptic polycyclic tetramate macrolactam biosynthetic gene cluster.** *Nat Commun* 2013, **4**:2894.

78. Zhang L, Hoshino S, Awakawa T, Wakimoto T, Abe I: **Structural Diversification of Lyngbyatoxin A by Host-Dependent Heterologous Expression of the tleABC Biosynthetic Gene Cluster**. *ChemBioChem* 2016, **17**:1407–1411.
79. Chen AH, Silver PA: **Designing biological compartmentalization**. *Trends Cell Biol* 2012, **22**:662–670.
80. Agapakis CM, Boyle PM, Silver PA: **Natural strategies for the spatial optimization of metabolism in synthetic biology**. *Nat Chem Biol* 2012, **8**:527–535.
81. Schaik W Van: **The human gut resistome**. *Phil Trans R Soc B* 2015, **370**.
82. Laxminarayan R, Duse A, Wattal C, Zaidi AKM, Wertheim HFL, Sumpradit N, Vlieghe E, Hara GL, Gould IM, Goossens H, et al.: **Antibiotic resistance—the need for global solutions**. *Lancet Infect Dis* 2013, **13**:1057–1098.