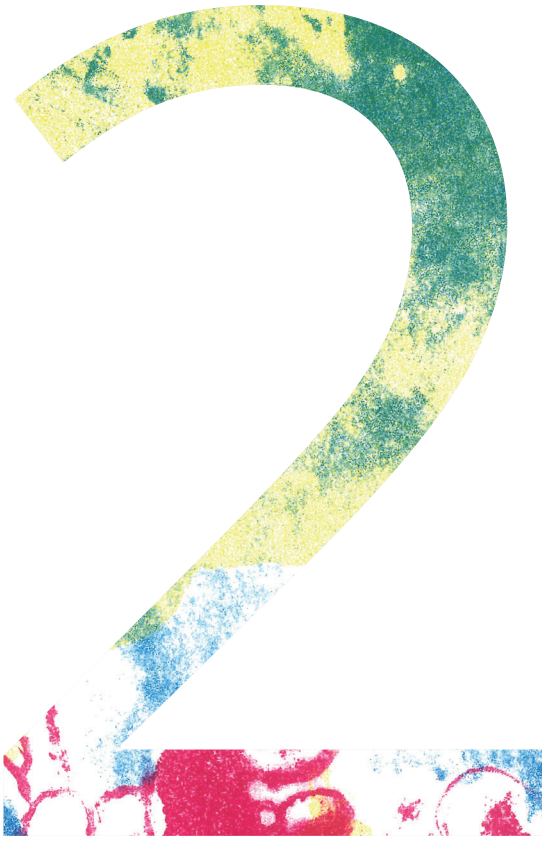
Ecology and genomics of Actinobacteria and their specialised metabolism

Bergeijk, D.A. van



Ecology and genomics of Actinobacteria: new concepts for natural product discovery

Doris A. van Bergeijk*, Barbara R. Terlouw*, Marnix H. Medema and Gilles P. van Wezel

* these authors contributed equally to this study

This chapter is published as:
van Bergeijk, D.A., Terlouw, B.R., Medema, M.H., van Wezel, G.P. Ecology and genomics of Actinobacteria: new concepts for natural product discovery. *Nat Rev Microbiol.* **18**, 546–558 (2020)

Abstract

Actinobacteria constitute a highly diverse bacterial phylum with an unrivalled metabolic versatility. They produce most of the clinically used antibiotics and a plethora of other natural products with medical or agricultural applications. Modern 'omics'-based technologies have revealed that the genomic potential of Actinobacteria greatly outmatches the known chemical space. In this Review, we argue that combining insights into actinobacterial ecology with state-of-the-art computational approaches holds great promise to unlock this unexplored reservoir of actinobacterial metabolism. This enables the identification of small molecules and other stimuli that elicit the induction of poorly expressed biosynthetic gene clusters, which should help reinvigorate screening efforts for their precious bioactive natural products.

Introduction

The phylum Actinobacteria represents one of the most diverse groups of microorganisms in nature. These Gram-positive bacteria have a high-GC content and show a remarkable range of morphologies, including unicellular cocci or rods (for example, members of the genera *Micrococcus* and *Mycobacterium)*, and morphologically complex multicellular bacteria (for example, members of the genera *Amycolatopsis, Frankia* and *Streptomyces*) [49]. Actinobacteria are widely distributed across both terrestrial and aquatic ecosystems, as well as in the microbiomes of higher eukaryotes [2]. This ecological diversity is reflected in their metabolic potential as Actinobacteria are extremely versatile producers of bioactive natural products. Notably, Actinobacteria produce two thirds of all known antibiotics used in the clinic today, but also a vast array of anticancer compounds, immunosuppressants, anthelmintics, herbicides and antiviral compounds, in addition to extracellular enzymes [9,50-52]. Therefore, these bacteria are attractive sources for clinical drugs [9,53].

The introduction of antibiotics in the 20th century greatly contributed to the extension of human life span, and has saved millions of lives worldwide. However, with the increase in antibiotic resistance, we now face a huge challenge in treating infections by multidrug-resistant bacteria [54]. This coincides with a dramatic decrease in the success of traditional drug development through high-throughput screening [17,55]. Indeed, the chance of finding new antibiotics via traditional methods in randomly chosen Actinobacteria has been estimated at less than one per million [16,56]. However, advances in genome sequencing have unveiled a vast reservoir of biosynthetic gene clusters (BGCs) for natural products in microbial genomes, even in those that had been studied extensively for decades [20,21,57]. The OSMAC strategy (one strain many compounds) [58] of extensive mining of individual strains still yields promising new molecules [59,60]. However, the rate of success has decreased dramatically since the golden years of drug discovery, primarily due to replication [16]. The apparent failure to uncover the full potential of natural product-producing microorganisms is likely due the fact that we lack the understanding that is required to activate the expression of their BGCs in the laboratory. During industrial screening, bacteria and fungi are typically grown in isolation with ample nutrients and resources, which is in sharp contrast to their natural complex and rapidly changing habitat. Antibiotics are believed to be important for survival as mediators of resource competition in a competitive environment [61,62], and microbial interactions have a key role in their activation [8,63,64]. Hence, to increase the success of natural product-based drug discovery, we need to elucidate the triggers and cues that activate the expression of BGCs [29,37].

Bioactive metabolites mediate important ecological functions, which are as diverse as their chemical structures. Siderophores enhance iron uptake in environments where the

bioavailability of iron is limited [65], pigments provide protection against UV-radiation and have antioxidant activity [66], and compatible solutes protect against osmotic stress [67]. However, the most obvious ecological purpose is biological weaponry to outcompete other organisms for resource acquisition [8,61]. As producers of various bioactive molecules, Actinobacteria are attractive to eukaryotic hosts as symbionts. For instance, *Streptomyces* spp. that live in the antennal glands of beewolf digger wasps produce antibiotics that protect the wasp larvae from various pathogenic fungi and bacteria [68], and endophytic Actinobacteria protect their host against phytopathogens [2,69].

In this Review, we discuss why Actinobacteria excel as natural product producers, and how their specialised metabolism and the regulatory mechanisms governing this metabolism have evolved in the context of ecology and genomic structure. Finally, we explore how ecological insights can be translated into approaches for computational and experimental genome mining strategies that yield novel bioactive molecules, in particular antibiotics.

A mycelial lifestyle

The propensity to produce bioactive molecules, and the richness of bacterial genomes in terms of BGC diversity, have been correlated to key organismal features such as multicellularity, endospore formation and genome size [70,71]. Bacteria can be divided broadly into two groups based on their adaptability: specialists and generalists, each with their own environmental niches. Specialists are dedicated to life in specific environments and therefore require less extensive metabolism and, accordingly, smaller genomes [72,73]. *Mycoplasma genitalium* is a well-known example of an organism with a small genome of around 580 kb and less than 500 genes [74]. The smallest genomes known to date belong to parasites and symbionts, with the beetle symbiont *Stammera spp.* as a remarkable example of an organism with a small genome of around 270 kb and 250 genes [75]. By contrast, generalists usually have larger genomes and a complex morphology, such as multicellularity and the formation of endospores [72,73]. Their ability to use multiple nutrient sources enables them to adapt to diverse environments and growth conditions, which requires complex metabolic regulation. Therefore, it is not surprising that this group of bacteria includes the most important producers of natural products such as Actinobacteria, Cyanobacteria and Myxobacteria [73,76]. Hallmark features of multicellularity and development include intraspecies communication, morphological differentiation, and programmed cell death (PCD) [77].

The linkage between morphological and chemical differentiation is best explained using *Streptomyces* as an example. Streptomycetes are mycelial organisms that reproduce by sporulation, with a lifecycle similar to that of filamentous fungi. When the conditions are

favourable, a single and uninucleoid spore will germinate and the hyphae grow out by a combination of apical growth and branching, which results in a complex mycelial network [50,78]. Exo-enzymes are released to break down natural polymers like cellulose, mannan and chitin, thereby providing nutrients. The vegetative hyphae are compartmentalised by occasional semi-permeable cross-walls to form large multinucleoid cells. The next step in the developmental programme is the formation of new sporogenic aerial hyphae, which eventually differentiate into chains of uninucleoid spores [50,78]. To fuel the onset of the developmental programme, old vegetative mycelia are autolytically degraded through PCD to liberate the necessary nutrients for the new biomass [79] (Figure 1). Eventually, reproductive aerial hyphae differentiate into long chains of spores. The onset of development is controlled by the *bld* (bald) genes, so-called because mutants fail to produce the fluffy aerial mycelium [80]. Genes that control the distinct steps leading towards the maturation of aerial hyphae and subsequent sporulation are called *whi* (white) genes, referring to the white appearance of mutants due to their failure to produce grey-pigmented spores [81].

The onset of morphological development correlates temporally with the production of antibiotics [82,83]. The production of antibiotics is likely to provide a line of defence to protect the PCD-released nutrients against motile saprophytic bacteria, while at the same time, antibiotic-mediated lysis of these competitors may serve as an alternative food source [50,77,84,85]. Interestingly, prodiginines, which have DNA-damaging properties that are not secreted and hence damage the DNA of the producer, apparently facilitate PCD in *Streptomyces coelicolor* [86]. This suggests that some antibiotics may even drive development via the initiation of PCD.

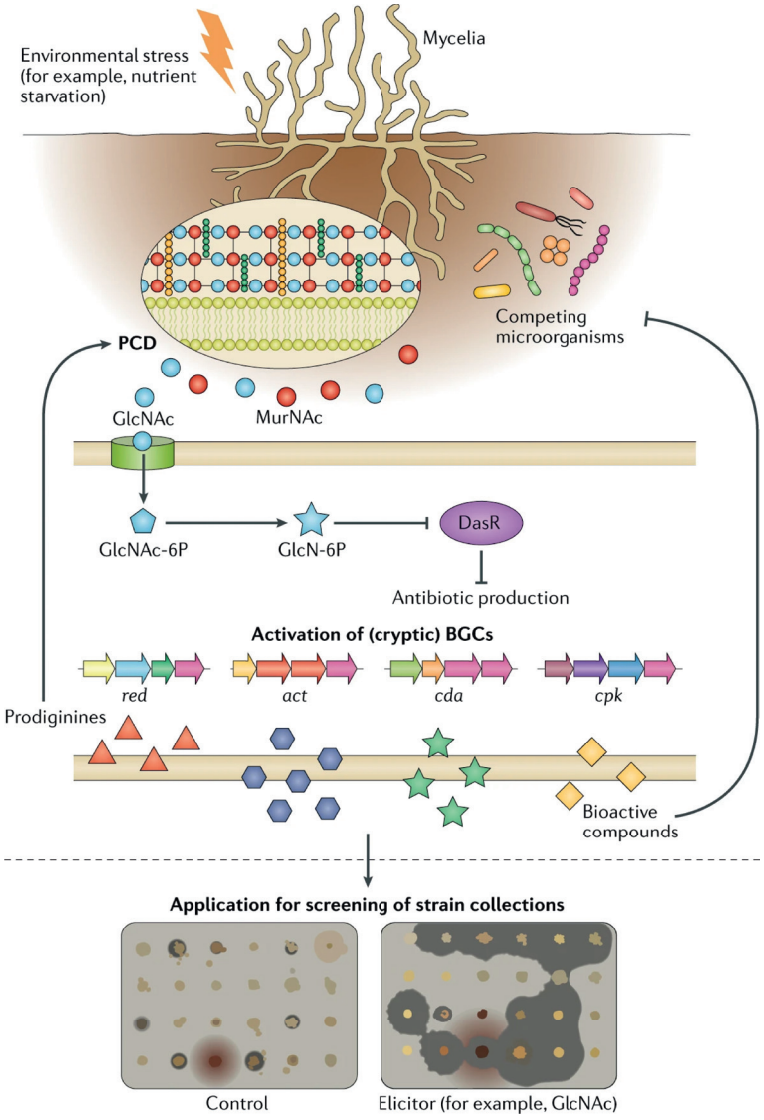


Figure 1. From biological understanding to elicitation of BGCs. Upon environmental stress, such as nutrient starvation, programmed cell death (PCD) leads to the autolytic degradation of the old mycelia, which liberates the necessary nutrients to fuel the onset of development, and for the new biomass. The onset of morphological development correlates temporally with the production of antibiotics. Specifically, cell-wall peptidoglycan is recycled to release the aminosugars *N*-acetylmuramic acid (MurNAc) and *N*-acetylglucosamine (GlcNAc). GlcNAc is internalised as GlcNAc-6P and converted to GlcN-6P, which functions as a ligand for the global regulator DasR. Binding of GlcN-6P inactivates the repressor DasR, which results in the de-repression of pathway-specific activators of antibiotic biosynthetic gene clusters (BGCs) [87]. The resulting production of antibiotics (such as calcium-dependent antibiotic (Cda), coelimycin P1 (Cpk) and actinorhodin (Act)) is likely to provide a line of defence to protect the nutrients that have been released by PCD against motile saprophytic bacteria, whereas the synthesis of DNA-degrading prodiginines (undecylprodigiosin (Red)) promotes PCD. GlcNAc elicits the transcription of BGCs that are not expressed under standard laboratory conditions and can thus be used in screening regimes for drug discovery.

Missing signals and cryptic antibiotics

The genome sequences of extant Actinobacteria 'document' their present and (recent) past ecology. They specify various functionalities used to cope with challenges that they face in their specific niches, including the natural products they produce. Not surprisingly, genome sequencing has had a huge impact on natural product discovery. This is exemplified by the specialised metabolism of *S. coelicolor* A3(2), the model organism for antibiotic biosynthesis. For a long time, this organism was known to produce four antibiotics, namely actinorhodin (Act), calcium-dependent antibiotic (Cda) and undecylprodigiosin (Red) and the plasmid-encoded methylenomycin (Mmy). When the genome of *S. coelicolor* was sequenced, many more BGCs were uncovered than originally anticipated, and the same was true for other model streptomycetes [20,21,88]. Renewed efforts for drug discovery led to the surprising discovery of yet a fifth antibiotic in *S. coelicolor*, called coelimycin P1 [87] in a model organism that had been studied by thousands of scientists in hundreds of laboratories around the world. Moreover, a novel branch of the biosynthetic pathway of the model polyketide actinorhodin was shown to be activated by co-cultivation with the fungus *Aspergillus niger* [89], and mass spectral imaging revealed substantial changes in the secreted metabolome during interaction of *S. coelicolor* with five other Actinobacteria [64]. These examples illustrate the concept of silent biosynthetic pathways, namely those that have been identified in the genome but whose cognate natural products are not synthesised under laboratory conditions (Box 1). This concept has revolutionised the field of antibiotic research [90], leading to the current era of genomics-based drug discovery [3,4].

Nowadays, next-generation genome sequencing enables scientists to explore large numbers of microorganisms in search of novel BGCs. However, there are several challenges. A first challenge lies in prioritising (cryptic) BGCs in terms of their potential for chemical novelty and/or clinical relevance to optimally exploit the wealth of available biological, genomic and metabolomics data. Second, the bulk of the microorganisms in soil and marine environments resist cultivation under laboratory conditions, and thus represent a huge 'white space' of biochemical diversity [91,92]. Third, many BGCs are not expressed during laboratory cultivation. To unlock this potential, we need to better understand the ecological context in which Actinobacteria live, as this will provide clues on the mechanisms that activate the biosynthetic pathways of natural products. To leverage genome information to this end, it is of major importance to understand how such ecological forces shape actinobacterial genomes and how biosynthetic diversity evolves.

Box 1: Silent biosynthetic gene clusters

What we have learned from genome sequencing efforts it is that many of the biosynthetic gene clusters (BGCs) are not accounted for in the corresponding metabolomes. These BGCs are referred to as 'silent' or 'cryptic'. Although these terms are sometimes interchanged, they have different meanings. A BGC is 'cryptic' when it has been identified but cannot yet be linked to a product, activity, or phenotype. To establish if a cluster is 'silent', experimental validation is required, at the level of gene expression or metabolomics (in case the compound is known). Silence of a BGC may reflect the fact that we do not yet understand the environmental conditions that are required for their expression. Alternatively, a BGC may have been silenced as a first step towards loss of the complete cluster. A transcriptome study of *Salinispora* strains revealed that more than half of all of their BGCs were expressed at levels that should facilitate discovery of the compounds they produce [93]. As transcriptome data of many BGCs is lacking, this suggests that most of the unexplored BGCs may be cryptic —that is, not linked to their products —rather than silent. For these, our failure to detect the corresponding metabolites is linked to other factors, such as translation efficiency and extraction methods.

Although the importance of 'omics'-based natural product discovery is clear [94], a proportion of BGCs remains transcriptionally silent in the laboratory. Transcriptome analysis revealed that biosynthetic potential is more complicated than the presence or absence of a BGC alone, as a substantial proportion of the BGCs were differentially expressed between strains [93]. An argument often heard during discussions between scientists in the field is that if a few different isolates of the same species are analysed, only one of these may express the BGC in question. True transcriptional inactivity of BGCs may be due to a mutation in a structural or regulatory gene, although the former is less logical as the entire biosynthetic machinery would be produced in vain. An intriguing concept is that regulatory genes may sustain a single (frame-shift) mutation, rendering the gene inactive, which may be easily restored by a compensatory mutation as a strategy for bet-hedging in a community, as seen for isolates of *Streptomyces lunaelactis* [95]. Such 'light-switch silencing' may be a primary mechanism by which strains are able to maintain large numbers of BGCs, as natural product biosynthesis is an energy-intensive process. Still, it is reasonable to assume that for many of the silent BGCs the right conditions for their transcriptional activation are lacking when the strain is grown in the laboratory.

Evolution of biosynthetic repertoires

Vertical inheritance and horizontal gene transfer

The diversity of the environments that Actinobacteria inhabit is extraordinary [2,96]. Specialised metabolites have been isolated from marine organisms such as *Salinispora* species [97], arctic *Streptomyces nitrosporeus* [98], desert-dwelling *Streptomyces* species [41], and species that live in the microbiomes of animals, insects and plants [43,69,99] . This ecological diversity has major implications for the evolution of secondary metabolic repertoires, which are driven by a combination of vertical inheritance and horizontal gene transfer (HGT). In terms of vertical inheritance, *Salinispora* species show a strong relationship between phylogeny and BGC diversity, with close relatives sharing nearly all BGCs [100]. Similarly, in *Amycolatopsis*, BGCs are conserved within clades but not between them [101]. These BGCs may be more important for adaptation than for survival. Similarly, rapid demographic expansion of *Streptomyces* into previously uninhabited niches has led to differentiation into various species clusters, distinguishable from each other by ancestral homologous recombination events [102].

HGT is another important evolutionary driver of chemical complexity of natural products in *Streptomyces* [100,103]. In *Salinispora*, it has been estimated that up to 96% of its biosynthetic pathways may have been acquired through HGT [100]. It is unclear how frequent HGT occurs in Actinobacteria. Some argue that lateral acquisition and subsequent maintenance of complete BGCs is very rare [104]. Nevertheless, it is likely that HGT has a key role in shaping BGC repertoires. Many more ancient HGT events of complete BGCs will have occurred than those that can still be reliably inferred, because those BGCs will have diverged over time since the moment of transfer [105]. The large discordance between gene phylogenies of core biosynthetic genes and species phylogeny testifies to this [106-108]. When HGT leads to acquisition of BGCs with similar functions, there may be strong evolutionary pressure for BGC loss. Indeed, in the evolutionary history of the genus *Salinispora*, genes for the biosynthesis of the siderophore desferrioxamine were lost in three strains independently as a direct consequence of HGT of a functionally similar BGC. Such events may happen at a large scale, and therefore could mask a large proportion of historical HGT events [109]. In fact, rates of HGT may vary strongly between different types of BGCs, depending on their ecological roles; many species have a conserved ‘core’ set of BGCs, together with a strongly varying set of ‘accessory’ BGCs that is acquired or exchanged through HGT [101,110]. This is evident in *Amycolatopsis*, in which BGCs that were acquired through HGT largely localise to non-conserved genomic regions [101].

Chromosome structure

Genomic structure is intimately tied with genetic change and conservation in specialised metabolism, possibly both as cause and effect. Actinobacteria have either linear or circular chromosomes. In linear chromosomes, as seen in *Streptomyces* and *Rhodococcus*, strain-specific genes and BGCs that are incorporated through HGT tend to be localised in the unstable subtelomeric end regions of the chromosome [103,111-113]. Selective pressure might induce migration of such BGCs towards the chromosomal core, where they are likely more stably maintained through vertical inheritance, and thus display much higher levels of conservation across different actinobacterial species [111-113]. A recent heterologous expression study also showed that chromosomal location of BGCs has an effect on expression levels, with expression levels of a β-glucuronidase reporter gene measured being the highest in the central regions of the chromosomal arms [114]. In *Streptomyces*, more centralised BGCs tend to encode molecules like ectoines and siderophores (Figure 2), which are likely to be essential for survival of the genus. In the circular chromosomes of *Salinispora* and *Amycolatopsis*, species-specific BGCs are largely located on genomic islands relatively distant from the origin of replication [115]. Conversely, conserved BGCs tend to be localised in the 'core genome'. Like in *Streptomyces*, in *Amycolatopsis* these BGCs specify metabolites such as ectoines, siderophores and terpenes, whereas strain-specific BGCs, which make up as much as 67% of all BGCs in *Amycolatopsis*, localise largely away from the core genome to the genomic islands [101]. Although, to our knowledge, no studies have investigated BGC migration from peripheral regions to core regions, BGC migration between genomic islands has been inferred to be likely based on phylogenomic analyses, which supports the hypothesis that BGCs preferentially migrate towards genomic core regions [115].

One of the reasons that actinobacterial genomes are packed with BGCs is because there are well-established mechanisms in place to acquire and exchange BGCs. One common way of acquisition is through integrative and conjugative elements (ICEs). These ICEs are plasmids with the ability to integrate themselves into the chromosome. The prevalence of these ICEs seems to be partially dictated by their ecological background: Actinobacteria originating from soil, plants or aquatic environments contain a greater number of ICEs than species from other environments [116]. In addition to ICEs, giant linear plasmids have a high density of BGCs for antibiotics, which suggests an important role in the acquisition of bioactive metabolites throughout evolution [117-119].

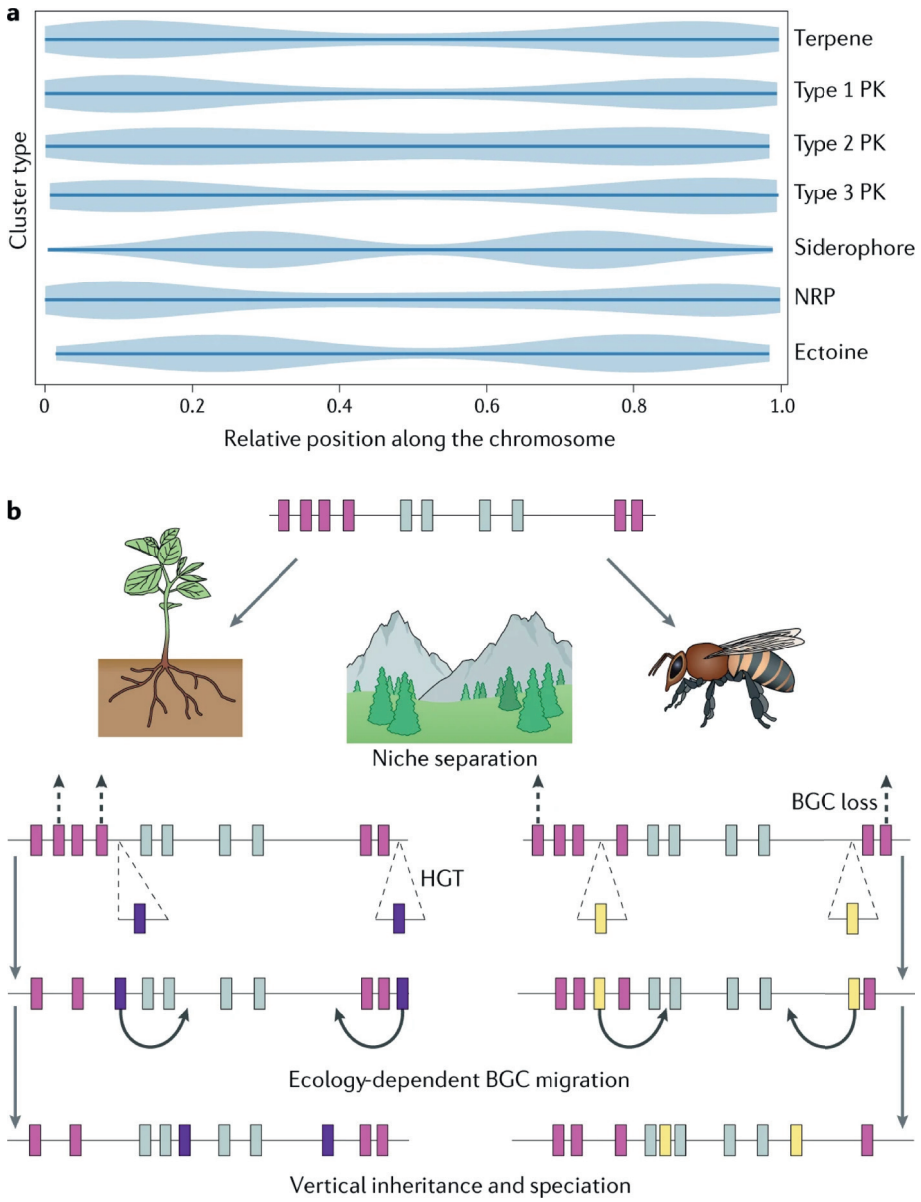


Figure 2. BGC distribution and evolution in *Streptomyces* chromosomes. a. The distribution of biosynthetic gene clusters (BGCs) encoding terpene, non-ribosomal peptides (NRPS), polyketides (PKS), siderophore and ectoine in 100 linear *Streptomyces* chromosomes. Clusters encoding type 1 PKS, type 3 PKS and NRPS localise to the chromosome ends, whereas clusters encoding type 2 PKS, terpenes, siderophores and ectoines localise more towards the chromosome core. **b.** Proposed movement of BGCs throughout evolution. Upon niche separation, genomes in different environments take up different genetic material through horizontal gene transfer (HGT) and/or biosynthetic gene clusters (BGCs) are lost. In the long term, BGCs that are important for survival in that niche migrate towards the chromosome core, and have a greater chance of being maintained through vertical inheritance.

Control of antibiotic production

Just like ecological forces shapes the organization of BGCs across actinobacterial chromosomes, they also shape how the clusters are regulated. A range of external signals influence production either directly through activation of pathway-specific activators, or indirectly, via an interactive network of pleiotropic regulators and intracellular signalling molecules [34,83,120]. Additionally, cross-regulation can exist between different BGCs, as was recently shown for the two chemically unrelated specialised metabolites antimycin and candicidin produced by *S. albus* S4. Their BGCs are separated by 9 kb on the chromosome, but the pathway-specific regulator of candicidin, FscR1, can bind directly upstream of the genes encoding antimycin biosynthesis and is essential for the activation of this cluster [121]. This regulatory complexity is highlighted by the strong emphasis on regulation in the genome of the model organism *S. coelicolor,* which encodes close to 800 regulatory proteins, representing >10% of the total proteome [20]. Higher-level control is likely to be tied to the ecological conditions in which these adaptive responses have evolved. Environmental and physiological signals have been integrated into the regulation of specialised metabolism to ensure that these costly molecules are only produced when required [2,34]. The involvement of environmental signals in the control and complexity of specialised metabolism is illustrated by the control of the BGC for ferroverdins and bagremycins in *Streptomyces lunaelactis* [95]. The metabolite produced from the BGC depends on iron availability. In iron-limiting conditions, the antimicrobial bagremycins are produced. When there is an excess of iron, the amino group of bagremycin is replaced by a nitroso group to generate ferroverdin, a siderophore that is used as anticholesterol drug, and that in nature likely functions to limit iron-mediated oxidative damage [95].

In a laboratory setting, the absence of the triggers and cues that would activate antibiotic production in the original habitat offers a possible explanation as to why so many BGCs remain poorly expressed or silent under laboratory growth conditions (Box 1). However, the identity of environmental ligands and/or signals perceived by both pleiotropic and pathway-specific regulatory proteins is a major area of investigation, and if resolved could lead to the activation of silent BGCs, and thus drug discovery [83]. The key lies in understanding the biology of the producing bacteria and translating these insights into solutions to activate antibiotic production (see below).

Here, we provide some background on the complex transcriptional control of BGCs, and then look into the regulatory networks that control the well-established connection between the onset of development and antibiotic production. For detailed reviews we refer the reader elsewhere [82,83,122].

Principles of the control of antibiotic production

To coordinate the metabolic responses to specific ecological challenges, Actinobacteria have evolved complex multilevel regulatory networks. These networks are composed of multi-level transcriptional and translational control. This is required for the correct interpretation of the signals that reach the colony, and translate them into appropriate responses. Much of our knowledge of the control of antibiotic production has been obtained from the study of the BGCs for Act, Cda and Red in *S. coelicolor* and for streptomycin in *Streptomyces griseus* [82,83,122]. These clusters are controlled by the pathway-specific activators ActII-ORF4, CdaR, RedD, which belong to the SARP family of *Streptomyces* antibiotic regulatory proteins [123], and StrR (ParB-Spo0J family [124]), respectively. These cluster-situated regulators directly control the level of transcription of the BGC, which in turn dictates the production level of the cognate natural product [125,126]. Interestingly, *actII*-ORF4, *cdaR* and *redD* are all subject to translational control by the tRNA that recognises the rare UUA codon for leucine. This tRNA, encoded by the *bldA* gene, is also required for the proper translation of many developmental genes, and thus links morphological to chemical differentiation [127,128].

Multiple cluster-situated regulators (CSRs) may control a single BGC, and in addition the BGCs are subject to global control. This enables the cell to coordinate specialised metabolism with growth and development, the balance in C-, N- and P-metabolism, and other major cellular pathways, thereby generating a complex hierarchy of regulatory networks. To enable efficient responses to external stimuli, the activity of many regulatory proteins is determined by small molecules. This includes the hormone-like γ-butyrolactones [4,129], feedback through biosynthetic intermediates [130-132] and sugar-based ligands [36,133]. The identification of such external signals, often referred to as elicitors, is of key importance for the rational activation (elicitation) of natural product biosynthesis, and thus for the revitalization of drug discovery.

PCD and the DasR regulatory network

As described above, PCD mediates the provision of nutrients at the onset of morphological and chemical differentiation. The signalling pathway from PCD to differentiation revolves around the global nutrient sensory GntR-family regulator DasR (Figure 1). Autolytic degradation of cell-wall peptidoglycan releases the aminosugars N-acetylglucosamine (GlcNAc) and N-acetylmuramic acid (MurNAc) around the colonies. Under nutrient-limiting conditions (famine), the accumulation of GlcNAc around colonies triggers development and antibiotic production, while under rich growth (feast) conditions, GlcNAc blocks both processes [36]. The rationale behind this is that under feast conditions, GlcNAc is seen as derived from chitin and signals nutrient abundance and promotes growth, whereas under famine conditions it signals hydrolysis of the bacterial cell wall and thus the need for development. GlcNAc-derived glucosamine-6-phosphate (GlcN-6P) and other

phosphosugars act as ligands for DasR and thereby inactivate the repressor, which results in the de-repression of BGCs [36]. DasR directly controls the pathway-specific activators of BGCs for all antibiotics and siderophores in *S. coelicolor* [134-136]. Addition of GlcNAc under nutrient-limiting conditions activates the transcription of antibiotic BGCs, including *cpk* which specifies the cryptic polyketide coelimycin; this principle is now also being applied in industrial screening regimes. Interestingly, there is direct competition between the DasR regulon and the regulatory networks governed by the transcription factors AtrA and Rok7B7: DasR represses *actII*-ORF4 and *nagE2*, encoding the GlcNAc transporter, while these genes are activated (and depend on) Rok7B7 and AtrA [137]. This system highlights the complex control of bioactive molecules in response to environmental changes. It also illustrates the value of discovering the ecological rationale behind antibiotic production and using such knowledge to activate antibiotic production (Figure 1).

Identifying the elicitors that activate cryptic BGCs

Various other approaches have been developed to identify the environmental signals involved in the regulation of actinobacterial specialised metabolism. The signals that trigger the expression of BGCs act through a transcriptional regulatory network, governed via *cis*-regulatory elements (CREs) targeted by transcription factors. Many transcription factors will respond to ligands, but how to uncover what these ligands are? Genomic context is one major pointer. For example, if a regulatory gene lies next to a metabolic operon (such as for sugar metabolism), this may be an important clue. Additionally, transcription factors are often autoregulatory, and the CRE is therefore typically found in the upstream region of the gene. With the CRE in hand, computational approaches can then be used to predict the regulatory network *in silico*. In the case of DasR, it was immediately obvious that the best hits in those predictions all related to GlcNAc metabolism or transport, and identifying glucosamine-6P as the ligand was then fairly straightforward [36]. Methods directed at single bacterial producer strains include varying the composition of growth media [58,138], inducing antibiotic resistance [139,140] and microbial cocultivation [141-143]. Screening for new chemical elicitors of antibiotic production is a promising approach, as this can enhance the chance of success in high-throughput screening of bacterial strain collections. Logical elicitors to include in such screens are those with a proven pleiotropic activity, such as GlcNAc [36], γ-butyrolactones [129,144] and histone deacetylase inhibitors [145]. Screening compound libraries for small molecules that perturb antibiotic production was shown to be an effective strategy to identify novel elicitors of antibiotic production [146]. Another example is bioactivity high-throughput elicitor screening technology (HiTES), in which a wild-type microorganism is subjected to a library of small molecules and the resulting induced metabolomes are screened for bioactivity against a chosen indicator strain [30,31]. Use of this method led to the identification of various cryptic antibiotics, including a novel lanthipeptide cebulantin [30] and a novel naphtoquinone epoxide hiroshidine [31]. It also identified atenolol, a β-blocker clinically used to treat hypertension, as a global elicitor [31].

Chemical ecological relationships as elicitors of antibiotic production

Within their natural environment, Actinobacteria are part of diverse microbial communities that include archaea, bacteria, fungi, protists and viruses. Within these communities, specific interactions have evolved and small molecules, like specialised metabolites, facilitate interactions between different microbial species (symbionts or competitors), including the activation of antibiotic production [5,147]. By mimicking these naturally occurring chemical-ecological relationships in so-called co-culture experiments, cryptic BGCs might be activated in the laboratory.

Indeed, co-cultivation of Actinobacteria with other bacteria or fungi changes their specialised metabolite production profile. Examples include the production of alchivemycin A by a *Streptomyces* strain following co-cultivation with the mycolic acid-producing *Tsukamurella pulmonis* [148], biosynthesis of a range of metabolites during co-culturing of *Aspergillus nidulans* and various streptomycetes [149], and the activation of a silent pathway of actinorhodin in *S. coelicolor* upon cocultivation with *A. niger* [89]. Co-culturing of marine-derived *Streptomyces spp.* with different human pathogens, including methicillin-resistant *Staphylococcus aureus*, resulted in increased production of different antibiotics and enhanced biological activity [150]. Co-culturing with multi-drug resistant bacteria might emerge as an effective, targeted approach to find novel bioactive compounds with activity against the pathogens for which new antibiotics are desperately needed.

The signals and cues that mediate the observed changes in specialised metabolite production are diverse and include physical cell–cell interactions [148,151], higher rate of nutrient depletion [152], enzymatic conversion of precursors to active metabolites [153], HGT [154] and microbial small molecules [64,155,156]. However, for many interactions the signals and specifically the molecular mechanisms are as yet unknown. The development of analytical techniques such as nanospray desorption electrospray ionization (Nano-Desi) and matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) imaging mass spectrometry enables the direct visualization of molecules exchanged during the chemical communication between microorganisms [157]. This method might help with the elucidation of the signals involved in above described interactions.

Actinobacteria are also found in close association with various eukaryotic hosts (Figure 3) and there are multiple examples of defensive symbioses between Actinobacteria and host, in particular for plants (suppressive soils) and insects (fungus-growing ants) [47,96,158]. An interesting example is provided by leaf-cutter ants, which live in symbiosis with the fungus *Leucoagaricus* and with *Pseudonocardia* bacteria [158]. The *Pseudonocardia* produce bioactive compounds to protect the fungal cultivar against infection by other fungi. Recent work showed that in return, the presence of *Pseudonocardia* elicits the production of antimicrobials by pathogenic *Escovopsis* fungi during infection of the cultivar [159]. This exemplifies an evolutionary arms race between the Actinobacterium and the fungus.

Interestingly, many plant growth-promoting bacteria produce phytohormones such as auxin and gibberelic acid [160]. This suggests that host and microorganism communicate (or hijack each other's communication channels) through the production of such metabolites. This may indeed work both ways, as plant hormones influence growth and specialised metabolism by endophytic Actinobacteria ([44], A. van der Meij, J.M. Raaijmakers and G. P. v. W. unpublished observations), which reveals that also host-specific signals can affect specialised metabolism (Box 2).

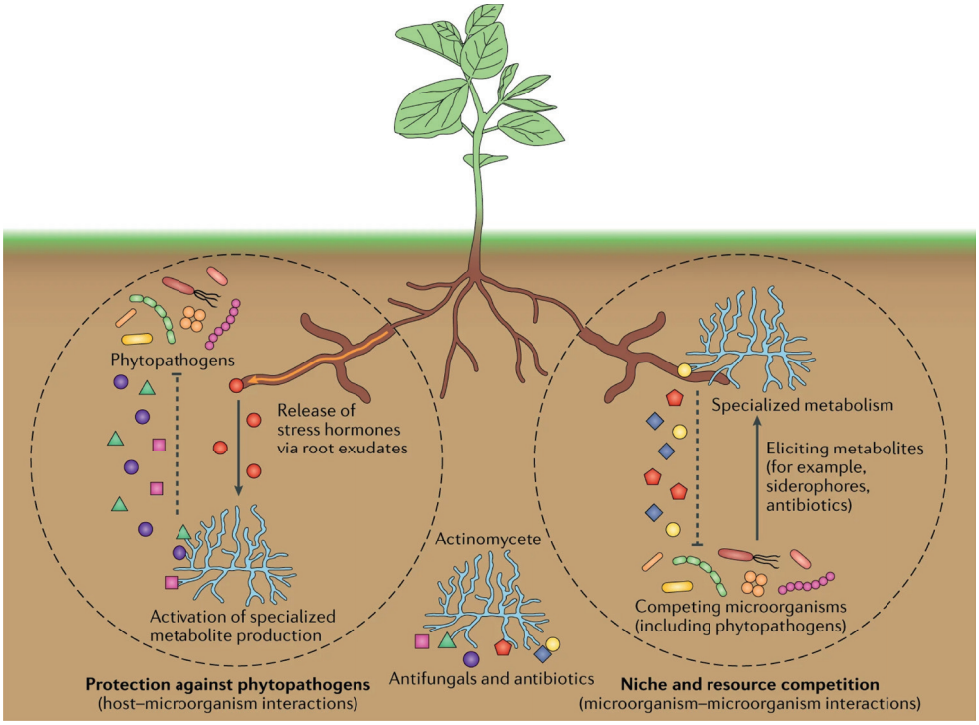


Figure 3. Natural products of Actinobacteria in host–microorganism and microorganism–microorganism interactions. Within the natural habitat of Actinobacteria, environmental signals are likely to have a key role in the activation of specialised metabolism. Chemical ecological interactions can for example be found in the rhizosphere of plants. Upon stress, plants release stress hormones (for example, jasmonic acid) via their root exudates. These hormones can activate antibiotic production by Actinobacteria, which can protect the plant against phytopathogens ('*cry for help*'). Additionally, competitive interactions occur between different members of the microbial soil community (both symbionts and pathogens). Metabolites, such as siderophores and antibiotics, from the competing microbial community can elicit changes in the metabolic profile of Actinobacteria, including antibiotic production. These molecules have an important role in shaping the rhizosphere microbiome and inhibiting (opportunistic) pathogens, thus also protecting the host.

Box 2. Host-derived signals as elicitors

The high abundance of Actinobacteria in plant microbiomes, specifically members of the genus *Streptomyces*, suggests that these bacteria are actively recruited by plants, possibly via root exudates [161-163]. Plants exudate up to 30%-60% of their photosynthate into the rhizosphere, which contain large quantities of sugars and amino acids that can influence the growth of beneficial bacteria as well as their antibiotic production (for example, through catabolite repression) [2,44,161]. These exudates contain plant metabolites, such as the phytohormone salicylic acid, which has been positively correlated with the abundance of operational taxonomic units attributed to *Streptomycetaceae* in the soil, as well as endophytic Actinobacteria [162]. As such hormones are released under pathogenic stress, an interesting hypothesis is that exudation of such compounds may be a means to 'cry for help': release of stress hormones by the host results in the recruitment of antimicrobial-producing Actinobacteria to reduce the severity of pathogenic infection [2,164] (Figure 3). Interestingly, when Actinobacteria are grown in the presence of plant stress hormones, like jasmonic acid and salicylic acid, changes in the production of specialised metabolites are observed, often leading to increased antimicrobial activity ([44]; A. van der Meij, J. M. Raaijmakers and G. P. v. W, unpublished observations). This could indicate that evolved regulatory networks between the host and members of its microbiome exist, opening up the intriguing possibility to design microorganisms that produce bioactive molecules specifically in response to stress induced by pathogen-derived signals. We hypothesise that similar use of Actinobacteria as 'medicine producers' may happen in our own microbiome, with mammalian stress hormones as activators specialised metabolism of Actinobacteria during infection.

Genome mining strategies

The biosynthetic diversity found in Actinobacteria is enormous. Genomic datasets have become increasingly larger, with massive strain collections, pan-genomes and metagenomes sometimes containing genetic information of thousands of Actinobacteria at once. A range of computational tools (for example, antiSMASH [165] or PRISM [166]) have been developed in the past decade that automate the identification of BGCs in these genomes and, to a certain extent, facilitate prediction of the structures of their products (Box 3).

Given the increasing size of genomic or metagenomic datasets, studying BGCs on a case-by-case basis is often no longer feasible. To address this, sequence similarity networking approaches have been developed to automatically relate predicted BGCs to gene clusters of known function (from, for example, the MIBiG database [167]) and to group them into gene cluster families (GCFs) [71,100,168]. The members of a GCF are then predicted to produce the

same or highly similar molecules. Using their streamlined computational framework for BGC similarity networking, researchers recently studied 3,080 actinobacterial genomes, and found that they contained around 18,000 distinct GCFs, the vast majority of which has no known products [108]. This constitutes an enormous potential for discovery, and to some extent enable us to differentiate between BGCs that are likely to produce compounds we have seen before and BGCs that may encode novel chemistry. However, the number of GCFs to which no known functions or chemistries can be linked is so great that it is difficult to know which of the BGCs belonging to them encode the production of the pharmaceutically most interesting molecules.

Some advances have been made to predict functions of BGCs. For example, the Antibiotic Resistance Target Seeker (ARTS [169]) prioritises BGCs based on co-localisation with resistance genes. The rationale behind this is that bacteria need a mechanism to protect themselves against the antibiotic molecules they are producing. Therefore, resistance markers may function as beacons to prioritise specific BGCs for antibiotic discovery. Unfortunately, only a small percentage of BGCs have distinguishable self-resistance markers. Ecological insights are needed to provide complementary strategies to know which microorganisms are most likely to encode biosynthetic pathways of interest, which BGCs among these are functionally most desirable, and how they can likely be activated.

Where to find chemical novelty?

To effectively mine Actinobacteria for drug discovery, there is a need for guidance towards chemical novelty. First of all, this requires a better understanding of the environmental and taxonomic distributions of BGCs. Such knowledge can help estimate where to search for novel producers and whether this search should be based on phylogeny, geography, or on specific environmental niches [70,101,168]. Taxonomic groups that are particularly gifted in terms of their natural product diversity include *Streptomycetales, Streptosporangiales, Frankiales, Micromonosporales* and *Pseudonocardiales* [168]. To decrease the risk of rediscovery of known molecules, known as replication, focus is directed towards rare Actinobacteria, of which many taxa have been greatly underexplored. Indeed, genera like *Micromonospora, Amycolatopsis, Salinispora, Nocardia* and *Verrucosispora* are a source of chemically unique metabolites with potent antibacterial activities, such as abyssomycins and proximicins [177]. Besides phylogeny, geographic location has also been proposed as an indicator of BGC distribution [178]. However, comparative genomics of *Amycolatopsis* and *Salinispora* strains both show that taxonomy is a more important indicator of BGC distribution than geographic location [97,100,101]. Intriguingly, regardless of their geographic origin, strains of *Salinispora arenicola* all produced rifamycin, staurosporine and saliniketal [97], suggesting that independent of the niche, strains may produce the same molecules or analogues. To a certain

degree, 'everything is everywhere and the environment selects', as Dutch microbiologist Baas-Becking proposed almost a century ago. Indeed, a study of biosynthetic diversity in soils from Central Park in New York City, United States, suggests that the degree of novelty found in common areas may be similar to that in more exotic locations [179].

Box 3. Computational tools for genome mining of biosynthetic diversity

The most commonly used tool for predicting BGCs in bacterial genomes is antiSMASH. Based on core genes, antiSMASH predicts gene cluster type (non-ribosomal peptide (NRP), polyketide (PK), terpene, siderophore, etc.). It also annotates and groups accessory genes, and minimally predicts the specialised metabolite core scaffold, enabling researchers to quickly identify BGCs of interest in an uploaded genome [165]. For instance, antiSMASH output was used as the starting point for a phylogenetic prioritisation method leading to the discovery of the novel compound corbomycin [170]. Other BGC identification tools also exist, such as BAGEL and PRISM [166,171]. The database MIBiG, which charts known BGCs and their products, provides a means of easily comparing predicted clusters to experimentally characterised ones that have known chemical products [167].

For BGC mining on a larger scale, the networking tool BiGSCAPE can be used to cluster both full and fragmented BGCs into gene cluster families, to obtain a comprehensive overview of BGC diversity in large genome collections or metagenomes. In such large datasets, the tool CORASON can help with prioritising BGCs of interest by providing insights into the evolutionary context of BGCs through phylogenetic analysis [108]. Genome-mining approaches can also be combined with proteomics for more efficient prioritisation [172]. Additionally, metabolomics data can be coupled to structure predictions yielded by genomics data or to absence-presence patterns of BGCs, in order to link molecules to BGCs [173,174]. The power of these methods will only increase now that tools are able to predict the function of natural products from their (predicted) structures [175]. Since structure prediction is easier on a substructure level, there are ongoing efforts to attempt linking substructure predictions to mass shifts, which can be a great aid in elicitation studies and dereplication [176].

Still, exploration of extreme or unusual environments like hyper-arid deserts, permafrost soils, mangrove trees, caves and deep-sea sediments that are characterised by challenging conditions (aridity, high salinity, low nutrient sources and extreme temperatures) showed high diversity of BGCs [41,180]. Between 2010 and 2018 alone, taxonomically diverse microorganisms originating from extreme environments have been the source of nearly 200 new specialised metabolites, of which many were produced by Actinobacteria [41]. Microorganisms from the

permafrost soil synthesise a broad range of chemical compounds [177]. Other interesting sources of gifted Actinobacteria are the microbiomes of diverse eukaryotic hosts, including insects, sponges and humans [43,99,181]. Within microbiomes, pathogen pressure selects for Actinobacteria that produce efficacious and relevant antimicrobials. Furthermore, this host association could potentially enrich for compounds with low toxicity to animals. This makes host microbiomes a promising source of novel molecules with possibly a higher potential to be successfully used in the clinic. Metabolomic analysis of *Streptomyces spp*. from insect microbiomes displayed immense potential for novel chemistry in these strains [43]. The same study demonstrated how PCA analysis can be leveraged for strain prioritisation; a strain characterised as metabolic outlier produced cyphomycin, a novel antifungal agent active against multidrug-resistant fungi [43].

However, we should remind ourselves that even well-studied organisms like *S. coelicolor* still harbour undiscovered biosynthetic pathways. For instance, it was shown that inactivation of the biosynthetic genes for the common antibiotics streptothricin and streptomycin resulted in the production of hidden antibiotics, such as the rare amicetin [182]. It is intriguing that despite the extensive research into such organisms, metabolic products of several putative BGCs so far eluded discovery. Also, there are still many exciting questions about the regulation of these specialised metabolite pathways, the signals that can activate production, and the ecological role of many of these molecules. Even now, many lessons remain to be learned from this and other well-studied model streptomycetes. The challenge is to find ways to leverage this potential, and ecology can play a key role in this.

Ecology to identify BGCs of interest

Even when selecting Actinobacteria from under-mined taxa and from high-potential environments, one will still end up with thousands of distinct BGCs to study, many more than can realistically be targeted for experimentation with currently available tools. Even the most high-tech synthetic biology approaches — while certainly being game changers that allow by-passing regulation for specified BGCs — will not enable synthetic refactoring of sufficiently large numbers of BGCs to facilitate a global screening of all actinobacterial biosynthetic diversity for years to come. One way to somewhat narrow down that number is to focus our attention on BGCs within non-core regions of actinobacterial genomes, which are more likely to encode compounds of chemical and functional novelty since HGT (and therefore the uptake of new, less conserved BGCs) mostly happens in these regions, as seen in *Amycolatopsis* [101]. However, even then there are still far too many BGCs to explore. Hence, further prioritization is required. For this, genomic and meta-omic data are potentially very useful (Figure 4).

First of all, predicting how BGCs are regulated can shed light on both their ecological functions and which triggers or cues can be used to activate their expression in the laboratory (Figure 4B). Computational tools that predict regulons can help uncover the regulatory networks responsible for BGC control. One such tool is PREDetector [183], which uses position weight matrices of transcription factor binding sites to predict which regulators are likely to bind DNA sequences within BGCs and thus likely regulate their expression. Often, a BGC encodes a pathway-specific regulator that is in turn regulated by a more pleiotropic (global) regulator. The identity of the global regulator can potentially be very informative about the ecological function of the BGC, and therefore the function of its product. For example, in a plant microbiome setting, BGCs regulated by DasR are likely to respond to N-acetylglucosamine, which is also a breakdown product of fungal cell walls; hence, this could point to a possible antifungal role of a compound produced by such a BGC. Computational searches have the potency to identify the entire regulons associated with them [183,184], and the gene content of these regulons may provide valuable data on the ecological functions of this regulon, and by proxy of the BGC in question [185]. The specific molecules eliciting the activation of these regulons would then still need to be identified; potentially, paired metabolomics and metatranscriptomics of native communities where the Actinobacterium resides may provide means to identify which molecules are specifically present when expression of the BGC is triggered. Such predicted regulatory cues can in turn feed back into tools like PREDetector to find novel BGCs in other species that may be similarly elicited.

Furthermore, using metatranscriptomic data (or transcriptomic data of co-cultures) is likely to be a powerful technology to predict the roles of BGCs in interaction with other organisms (Figure 4C). Knowing under which conditions members of certain GCFs are expressed can illuminate their likely functions and hint toward how they may be regulated. For example, determining which bacterial BGCs in a plant endosphere microbiome were upregulated upon fungal infection recently led to the identification of a gene cluster essential for disease suppression [186]. The attractive concept that chitooligoaccharides produced from hydrolysis of the fungal cell wall elicit the production of the antifungal needs to be tested. Additionally, expression of BGCs can be correlated to the absence/presence/abundance of specific other organisms in the community, to identify whether they might either be triggered by their presence, or whether they might effectuate their loss from the community by e.g. antibiosis. Adding metabolomics to the equation may provide further means of prioritisation, as candidate products for a BGC can be identified that specifically appear when it is expressed, and MS/MS analysis algorithms [176,187-190] can be used to dereplicate them and predict (parts of) their structures to assess their novelty (Box 3).

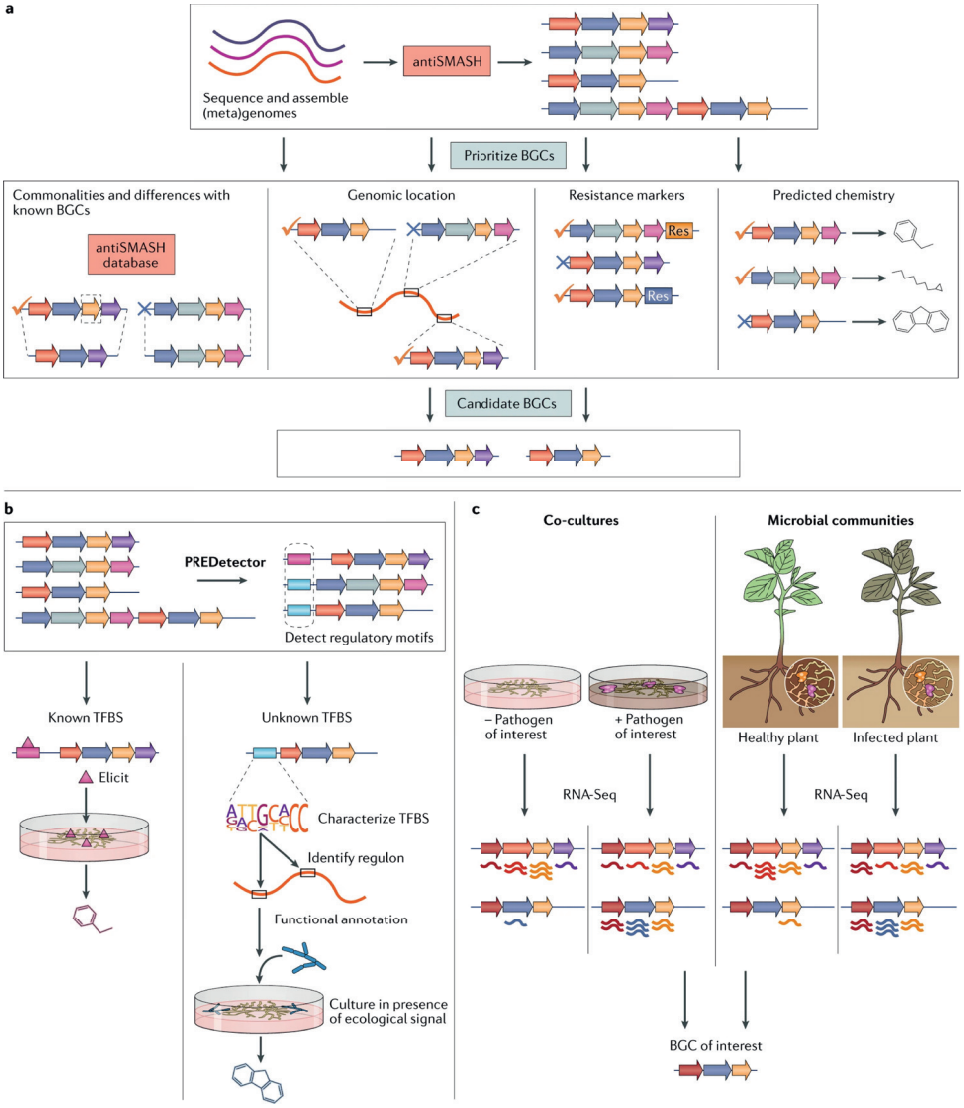


Figure 4. Omics strategies for BGC prioritization and elicitation. a. Genomics-based prioritization of BGCs. After identifying putative BGCs through antiSMASH, BGCs can be prioritised based on commonalities and differences with known BGCs, their location on the genome, resistance markers in the vicinity of the BGC and predicted chemistry of the compound the BGC produces. **b.** Genomics-based prediction of BGC regulatory sequences and elicitors. After predicting both known and unknown transcription factor binding sites (TFBS) with PREDetector, species with known TFBSs can be directly elicited. Unknown TFBSs can be characterised by generating a sequence profile of similar TFBSs, and searching the genome for the entire regulon regulated by the corresponding transcription factor. From this, the genes in the regulon can be functionally annotated, and the ecological signal that triggers BGC expression can be inferred. **c.** Transcriptomics-based prioritization of BGCs. Differential gene expression of BGCs can be quantified through RNASeq of co-cultures or microbiomes with or without a pathogen of interest. BGCs that are expressed when the pathogen of interest is present are more likely to have a role in targeting this pathogen and host protection.

Synthetic biology approaches to express gene clusters

Although computational strategies can be used to prioritise the most novel BGCs for experimental characterization, we then face the challenge of identifying the cognate chemical products of these BGCs. This currently remains a bottleneck, as thousands of potentially interesting BGCs can be found across publicly available genome sequences. By direct sequencing of environmental DNA, metagenomics even enables us to predict the chemical space of microbial 'dark matter': thus far uncultivated bacteria, which represent a promising source of novel natural products [191]. Synthetic biology is a powerful strategy to facilitate expression of BGCs observed in genome or metagenome sequences [192]. This is illustrated by recent work on the identification of bioactive molecules from the human microbiome [192,193]. The synthesis of a metagenomic BGC and subsequent heterologous expression in *S. albus* enabled the isolation and identification of new polyketides, designated metamycin A, metamycin B, metamycin C and metamycin D. Advances in synthetic biology and genome engineering (reviewed in detail in Ref. [192]) can become very useful in the expression of cryptic BGCs and the identification of their chemical products. However, designing a DNA sequence for a large pathway that will be functional in a model production host has remained more challenging than previously anticipated, as achieving the required stoichiometry between transcriptional units in a BGC for it to produce fully elaborated products is non-trivial [194]: required precursors or co-factors may be lacking in the heterologous host [195] and low production titres from synthetic BGCs may hamper chemical characterisation. Although all these bottlenecks are being addressed, understanding the regulatory mechanisms behind BGC expression for now remains of key importance for the identification of their products. And once high-throughput refactoring of complex BGCs becomes a reality, we anticipate that ecological and regulatory information will be crucial to predict BGC functions and thus prioritise them for synthesis and expression.

Conclusions and future perspectives

Despite our advances in niche exploration revealing great potential for drug discovery, the current state of knowledge regarding BGC diversity and distribution in terms of ecology and phylogeny limits our ability to guide drug discovery. It is therefore necessary to further characterise the extant microbial diversity from different ecological niches and create a global survey of niche-correlated natural product diversity. Moreover, characterising their functions in their native microbial communities as well as their modes of actions will be crucial to advance our understanding of the regulation of specialised metabolism, and hence for our effective ability to prioritise BGCs and elicit their expression. New technologies will be required for this, and in the 'omics' area, we specifically envision a

larger role for transcriptomic and metatranscriptomic studies of specialised metabolism, as well as regulatory network reconstruction, targeted to the most relevant microbiomes and ecological niches. Once we better understand which cellular and ecological conditions induce the expression of BGCs, this will greatly facilitate prioritising gene clusters that are likely to have functions of interest and to predict which molecular stimuli are likely to activate them.