Universiteit
Leiden
The Netherlands

Ecology and genomics of Actinobacteria and their specialised metabolism

Bergeijk, D.A. van

Citation
Bergeijk, D. A. van. (2022, October 19). *Ecology and genomics of Actinobacteria and their specialised metabolism*. Retrieved from https://hdl.handle.net/1887/3484350

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/3484350

Note: To cite this publication please use the final published version (if applicable).

Ecology and genomics of Actinobacteria and their specialised metabolism

Doris A. van Bergeijk

ISBN: 978-94-6458-503-2
Cover: *Hidden potential*, litho, 2022, by Doris A. van Bergeijk & Peeter Burgeik
Lay-out and design: Publiss | www.publiss.nl
Print: Ridderprint | www.ridderprint.nl

© Copyright 2022: Doris Antonia van Bergeijk. All rights reserved

Ecology and genomics of Actinobacteria and their specialised metabolism

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof.dr.ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op woensdag 19 oktober 2022
klokke 15:00 uur

door

Doris Antonia van Bergeijk
Geboren te Haarlemmermeer, Nederland
in 1992

Promotores:	Prof. dr. G.P. van Wezel
	Prof. dr. H.P. Spaink
Promotiecommissie:	Prof. dr. A.H. Meijer
	Prof. dr. D. Claessen
	Prof. dr. W. Bitter (Vrije Universiteit Amsterdam)
	Prof. dr. L. Gram (Technical University of Denmark)
`	Prof. dr. J.M. Raaijmakers

Voor mijn ouders Hanneke en Peter

&

voor mijn oma Cootje

Contents

Chapter 1. A brief history and thesis outline **9**

Chapter 2. Ecology and genomics of Actinobacteria: new concepts for natural product discovery **15**

Chapter 3. Taxonomic and metabolic diversity of Actinobacteria isolated from faeces of a 28,000-year-old mammoth **41**

Chapter 4. Animal stress hormones enhance siderophore production in *Streptomyces* **69**

Chapter 5. The ubiquitous catechol moiety elicits angucycline production in *Streptomyces* **87**

Chapter 6. Bioactive *Pseudonocardia* within the microbiome of zebrafish **125**

Chapter 7. Summarising general discussion **143**

Nederlandse samenvatting **153**

References **161**

Curriculum vitae **179**

List of publications **181**

A brief history and thesis outline

Filamentous Actinobacteria are among the most talented chemists on Earth [1]. They produce a plethora of chemically diverse natural products with a wide range of biological activities that help them survive in challenging and competitive environments [2]. These specialised metabolites serve as antioxidants [3], regulatory signals [4], metal scavengers [5], attractants [6], and biological weaponry [7,8]. Due to their chemical and biological diversity, numerous actinobacterial natural products have been the starting point for drug discovery and have found use in human and veterinary medicine, agriculture, and biotechnology [9-11].

The interest in microbial natural products for medical applications was set in motion in 1928, by the discovery of the antibiotic penicillin by Alexander Fleming [10,12]. However, it was the systematic screening of soil microorganisms by the Waksman group in the 1940s, resulting in the discovery of streptomycin, that really prompted the interest of pharmaceutical companies [13]. During the next 30 years, in the so-called “golden age of antibiotic discovery”, high-throughput screening efforts identified over 1000 natural products with antibacterial or antifungal activity, including many of the antibiotics that are still used in the clinic today [10]. In the next years, advances in screening methodologies led to the discovery of other clinically and industrially relevant molecules, such as chemotherapeutics, immunosuppressive drugs, herbicides, and hydrolytic enzymes [10,11]. In 2010, over 30,000 bioactive microbial natural products had been characterised, of which ~ 30% originating from Actinobacteria [11,14].

The discovery of antimicrobial natural products revolutionised the treatment of infectious diseases. Yet, human use of antibiotics has resulted in widespread emergence of multi-drug resistant pathogens [15]. Once again, bacterial infections pose a significant threat to human health, urging the need for novel antibiotics. However, the success of high-throughput screening has dramatically declined, primarily due to the rediscovery of known molecules (known as replication) [16]. Over the past 40 years, only three new classes of antibiotics have been discovered and many pharmaceutical companies have abandoned antibiotic research [14,17,18].

However, in the early 2000s, a renaissance in microbial natural product research was initiated by the development of genome sequencing techniques [10,16]. Bacterial genes responsible for the biosynthesis of specialised metabolites are typically clustered together on the genome, forming recognizable biosynthetic gene clusters (BGCs) [19]. Genome sequencing made it possible to identify these BGCs within the genome and bring the full biosynthetic potential of bacteria into the light. When the first two *Streptomyces* genomes were sequenced, it revealed numerous uncharacterised BGCs, even though these organisms had been studied extensively for decades [20,21]. For example, the extremely well-studied *Streptomyces coelicolor* A3(2) was known to produce no fewer than four antibiotics: the plasmid-determined methylenomycin A [22], the polyketide antibiotic actinorhodin [23],

the red-pigmented undecylprodigiosin [24], and the lipopeptide calcium-dependent antibiotic [25]. Yet, sequencing of its genome revealed the presence of a further 18 clusters [20] and renewed efforts led to the discovery of yet a fifth antibiotic in *S. coelicolor*, called coelimycin P1 [26].

Today, it has been estimated that we have characterised less than 3% of natural products potentially encoded in bacterial genomes [27]. Our apparent failure to uncover the full biosynthetic potential of bacteria is largely due to the fact that we lack the understanding that is required to activate BGC expression in the laboratory. It is now recognised that many biosynthetic pathways are silent: they have been identified in the genome but the cognate products are not synthesised under laboratory conditions [2,28,29]. Yet, heterologous expression and activation studies have shown that these BGCs encode functional pathways synthesizing novel molecules [30,31]. Silence of a BGC therefore may reflect that we do not yet understand the conditions that are required for their expression [29,32].

Laboratory conditions are in sharp contrast to the natural and rapid changing habitat of bacteria where specialised metabolism has evolved. Within their natural ecosystems, biotic and abiotic stresses, as well as interactions with other organisms, have shaped the regulation of natural product biosynthesis [32-34]. It is therefore highly likely that many bioactive metabolites are only produced under specific environmental conditions, in response to changes in nutrient availability, the presence of competitors, and the interactions with other organisms [8,35-37]. Identification of these ecological signals and the regulatory pathways involved is needed to unlock the full potential of Actinobacteria. Moreover, ecology plays a role in the search for gifted producers as ecological forces have shaped BGC diversity and distribution [38]. Lastly, understanding the ecological role of specialised metabolites within the microbial communities of plants, insects and animals, may guide us towards the use of antibiotic-producing Actinobacteria as probiotics in agriculture or human health. As part of the *metagenomics* PhD programme of the Netherlands Centre for One Health (NCOH) [39], this thesis connects to the NCOH research theme *tackling antimicrobial resistance* and aims to use ecological approaches to access the full biosynthetic potential of Actinobacteria.

In **Chapter 2**, we discuss how the specialised metabolism of Actinobacteria and the regulatory mechanisms governing this metabolism have evolved in the context of ecology and genomic structure. We provide background on the complex transcriptional control of BGCs and discuss chemical-ecological relationships as elicitors of antibiotic production. Lastly, we explore how ecological insights can be translated into approaches for computational and experimental genome mining strategies that yield novel bioactive molecules.

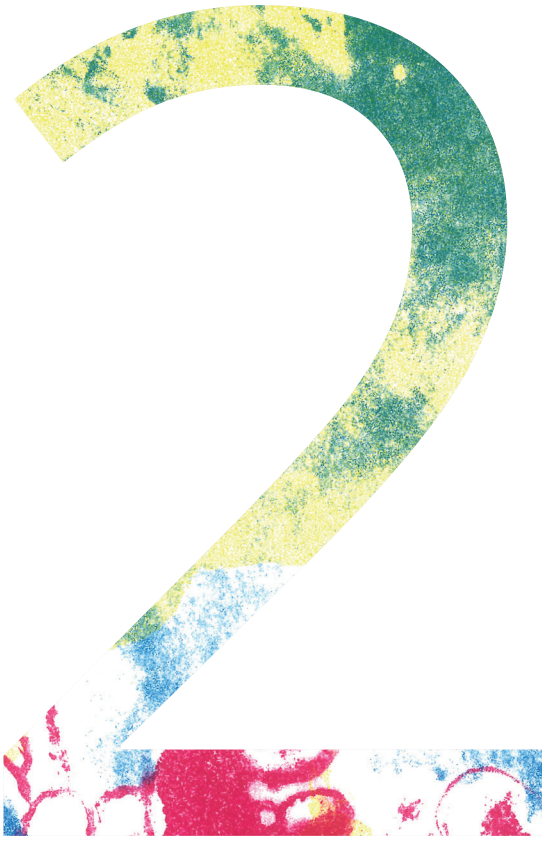
Screening efforts have focused on only a narrow band of habitats and a few taxonomic groups. It is likely that even more biosynthetic potential will come to light when more diverse environments and taxa are explored. Over the past years, Actinobacteria have been isolated from extreme environments and microbiomes of different organisms, resulting in the discovery of many novel molecules [40,41]. **Chapter 3** describes the isolation of Actinobacteria from a unique niche, namely a faecal sample of a 28,000-year-old mammoth. Subsequently, the biosynthetic potential and phylogeny of the isolated bacteria was analysed. Different ancient Actinobacteria were revived and their genomes were sequenced to compare them to currently known strains. This revealed significant phylogenetic distance to known modern strains. The isolates produced several known bioactive metabolites, but also harboured many uncharacterised biosynthetic gene clusters whose cognate natural products await discovery.

Although often described as free-living organisms, Actinobacteria live in and around a wide variety of other organisms, including higher eukaryotes such as plants, insects, marine organisms, and mammals [42,43]. As part of the microbiomes of these hosts, Actinobacteria are exposed to host-associated signaling molecules, many of which will likely influence their specialised metabolism [42]. Indeed, plant stress hormones, such as salicylic acid and jasmonic acid, can increase the antibiotic activity of endophytic streptomycetes [44]. **Chapter 4** focuses on the impact of human stress hormones on the specialised metabolism in *Streptomyces*. The data show that epinephrine (adrenaline), involved in the fight-or-flight response, enhances siderophore production. Catechol was established as the likely eliciting moiety, since similar responses were seen for catechol and for the catechol-containing molecules dopamine and catechin, but not for related molecules. Proteomic profiling demonstrated that the expression of proteins involved in iron uptake, siderophore production, and dithiolopyrrolone biosynthesis is increased in the presence of catechol compounds. Thus, we show that plant- and animal-associated molecules increase siderophore production in *Streptomyces*.

We were intrigued by the specificity of the response of *Streptomyces* to the catechol moiety. As a follow up, the eliciting potential of catechol itself was tested in **Chapter 5**. We explored the response of *Streptomyces* sp. MBT84 to catechol, which reproducibly produced strong antibacterial activity against *Bacillus subtilis* when exposed to catechol. A multi-omics approach showed that catechol elicits a BGC that produces different angucycline glycosides and a new member of this family, galtamycin D, was identified. Additionally, heterologous expression of catechol-degrading enzymes showed that catechol, and not one of its degradation products, is responsible for antibiotic elicitation. These results reveal that catechol, by itself and as part of animal stress hormones, can serve as elicitor of different classes of natural products.

The specialised metabolites of Actinobacteria mediate important ecological functions, allowing Actinobacteria to thrive in a wide range of environments. The beneficial effects provided by their metabolites are not limited to Actinobacteria themselves, but can also provide advantages to the many higher organisms that host Actinobacteria [45-47]. As part of the microbiomes of plants and insects, Actinobacteria can play an important role in the protection against pathogens [47,48]. Yet, the bioactive and functional potential of Actinobacteria within the microbiome of other organisms remains poorly characterised. Zebrafish are increasingly used as a model to study the role of the microbiome in health and disease. In **Chapter 6**, we investigated antibiotic-producing Actinobacteria associated with this model system. We analysed the gut microbiome of adult zebrafish and isolated Actinobacteria from zebrafish larvae, including a *Pseudonocardia* sp. with antibacterial activity. Genome sequencing revealed high similarity to ant-associated *Pseudonocardia*. Although further research is required to identify the bioactive metabolites produced by the *Pseudonocardia* isolate, this study provides a first step towards the use of zebrafish as a model to explore the role of bioactive Actinobacteria within the animal microbiome.

Finally, the main findings of the thesis are summarised and discussed in **Chapter 7**.



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.

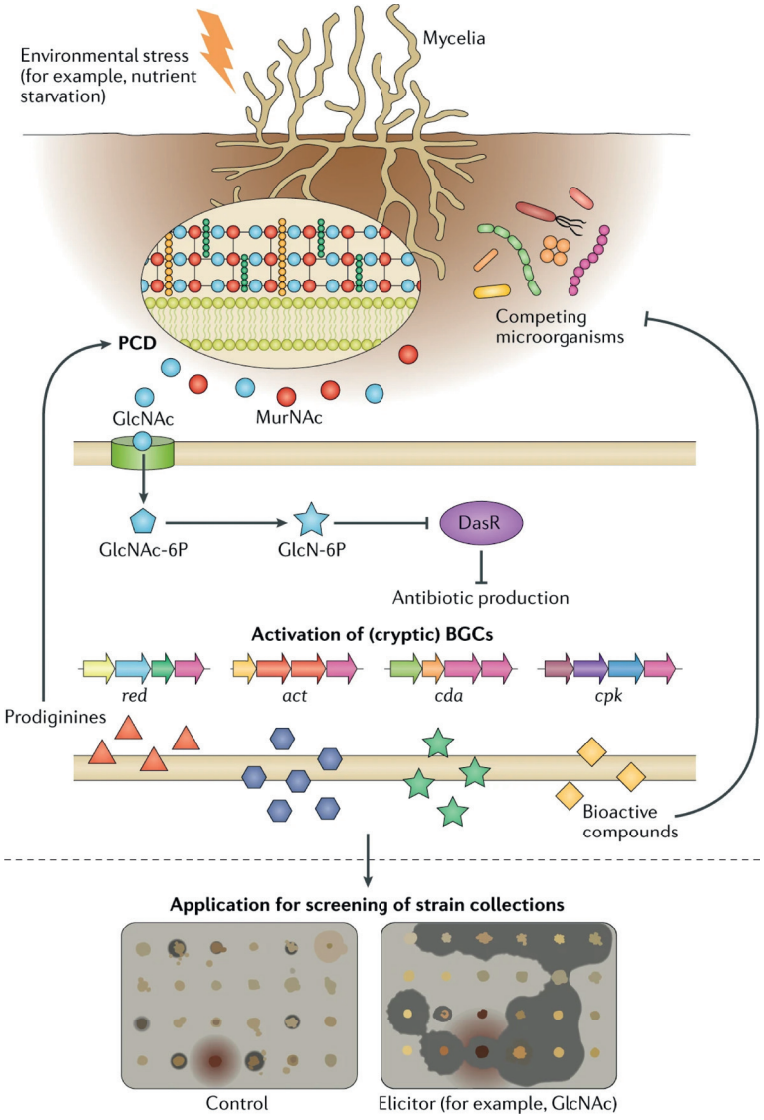


Figure 1. From biological understanding to elicitation of BGCs. Upon environmental stress, such as nutrient starvation, programmed cell death (PCD) leads to the autolytic degradation of the old mycelia, which liberates the necessary nutrients to fuel the onset of development, and for the new biomass. 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].

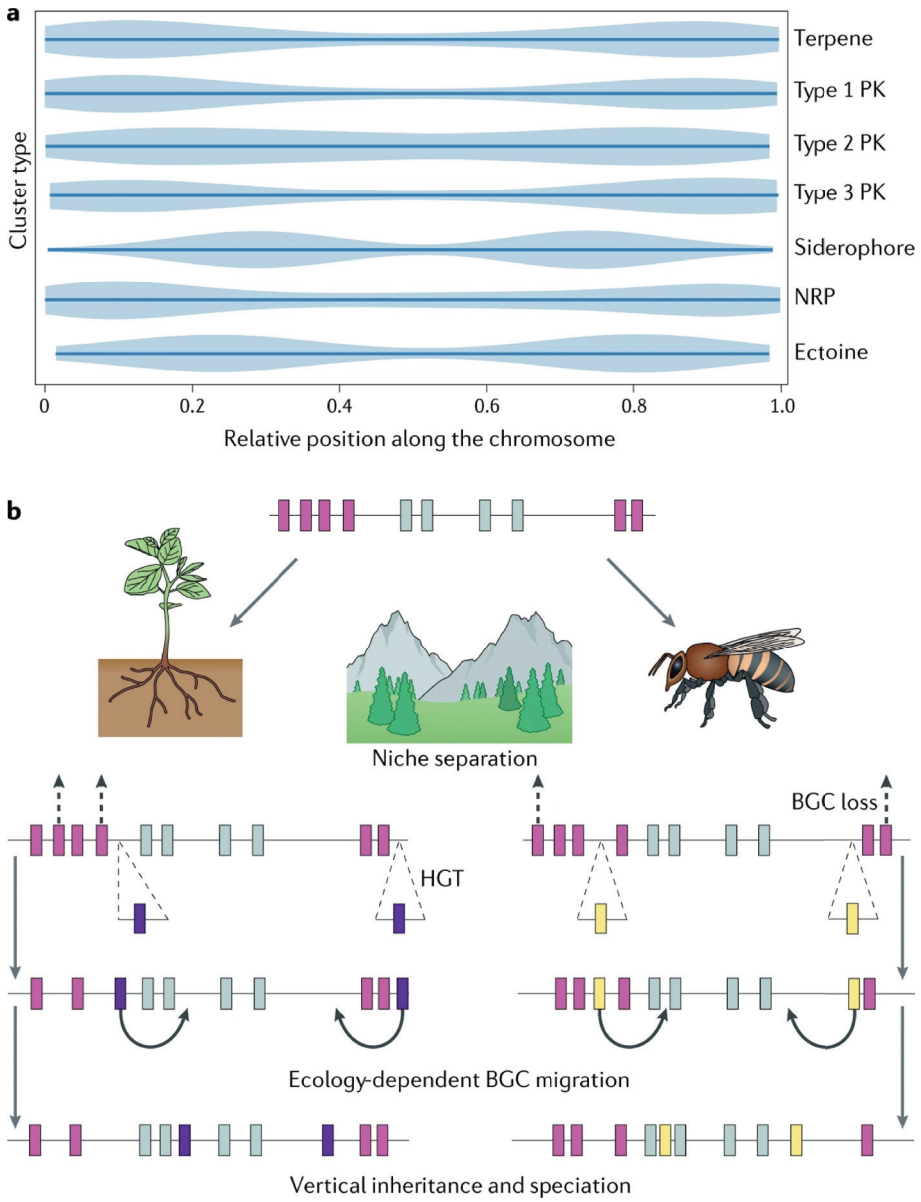


Figure 2. BGC distribution and evolution in *Streptomyces* chromosomes. a. The distribution of biosynthetic gene clusters (BGCs) encoding terpene, non-ribosomal peptides (NRPS), polyketides (PKS), siderophore and ectoine in 100 linear *Streptomyces* chromosomes. Clusters encoding type 1 PKS, type 3 PKS 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).

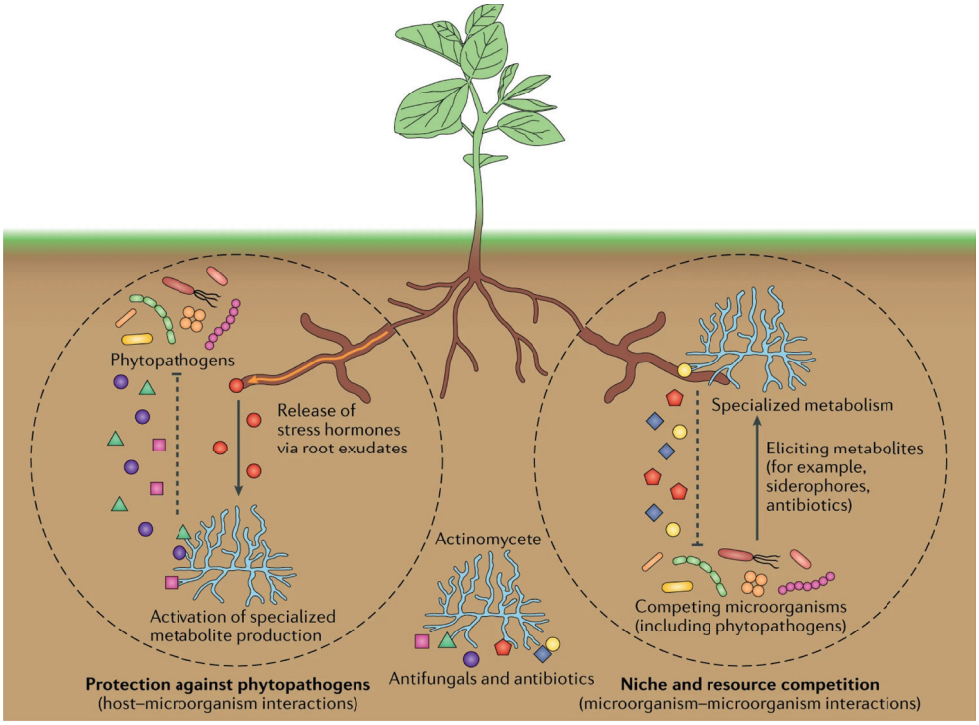


Figure 3. Natural products of Actinobacteria in host–microorganism and microorganism–microorganism interactions. Within the natural habitat of Actinobacteria, environmental signals are likely to have a key role in the activation of specialised metabolism. Chemical ecological interactions can for example 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).

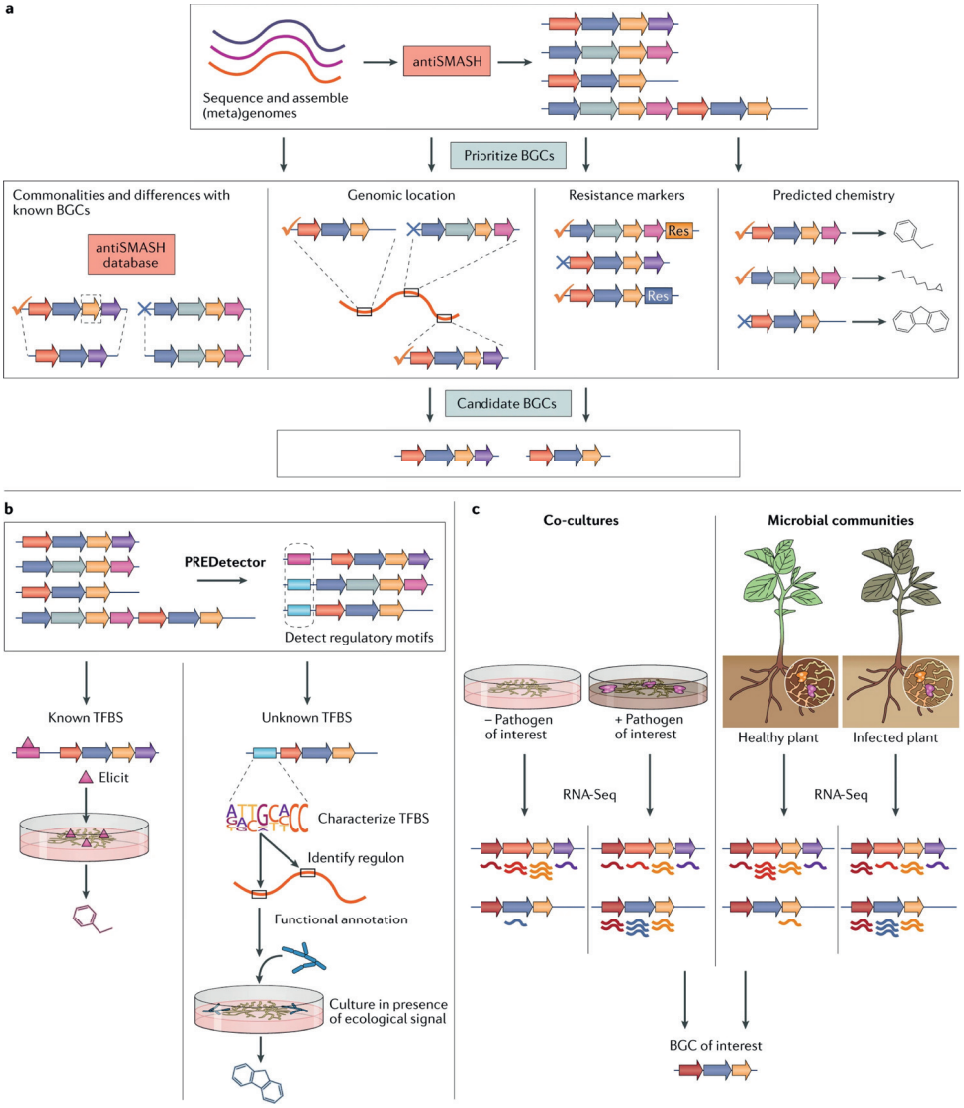


Figure 4. Omics strategies for BGC prioritization and elicitation. a. Genomics-based prioritization of BGCs. After identifying putative BGCs through antiSMASH, BGCs can be prioritised based on commonalities and differences with known BGCs, their location on the genome, resistance markers in the vicinity 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.

Taxonomic and metabolic diversity of Actinobacteria isolated from faeces of a 28,000-year-old mammoth

Doris A. van Bergeijk*, Hannah E. Augustijn*, Somayah S. Elsayed, Joost Willemse, Victor J. Carrión, Mia Urem, Lena V. Grigoreva, Maxim Y. Cheprasov, Bas Wintermans, Andries E. Budding, Herman P. Spaink, Marnix H. Medema, Gilles P. van Wezel

* Contributed equally.

Abstract

The yet unexplored microbial world is a vast resource that can provide us with new evolutionary insights, metabolic pathways, and chemistry. Ancient microbial communities of permafrost soils and frozen animal remains represent an archive that has barely been explored. Here, we show that Actinobacteria revived from a faecal sample from the intestinal tract of a 28,000-year-old Siberian mammoth are phylogenetically and metabolically distinct from their modern siblings. Ancient *Micromonospora*, *Oerskovia*, *Saccharopolyspora, Sanguibacter,* and *Streptomyces* species were recovered and subsequently grown up in the laboratory. Surprisingly, the genomes of these ancient ancestors showed a large phylogenetic distance to any known modern strains and harboured many novel biosynthetic gene clusters that may represent uncharacterised biosynthetic potential. The strains produced known molecules like antimycin, conglobatin and macrotetrolides, but the majority of the mass features could not be dereplicated. Thus, our work provides a snapshot into Actinobacteria isolated from a unique ancient animal faeces sample, yielding unexplored genomic information that is not yet present in current databases.

Introduction

Bacteria are key to life on Earth; they are major players in nitrogen cycling [196], decompose organic matter [197], and provide eukaryotic hosts with essential nutrients and protection against pathogens [198,199]. Additionally, bacterial natural products are a major source of the bioactive molecules that find applications in medicine, biotechnology and agriculture [9,18]. It is predicted that we have only uncovered a small percentage of the microbial world; the majority of microorganisms resist cultivation and many bacterial taxa have hardly been explored [200]. One reason is that scientists have so far primarily accessed environments that can be easily reached, while numerous bacteria exist in hard-to-reach places such as deep-sea sediments, caves, and permafrost soils [41]. This uncharacterised microbiology represents an important reservoir of biological information that may be harnessed for drug discovery.

Microbial communities that have been underexplored are those in ancient samples, such as permafrost soils, frozen animal remains, and deep soil layers. These samples not only represent unique environments, but may also allow a glimpse into the past, i.e., may teach us important lessons on the evolution of microbial and chemical diversification. Isolation and metagenome sequencing of ancient bacteria has provided important knowledge on bacterial evolution, changes in microbiome composition, ancient diseases and potential chemical novelty [201-204]. For example, large-scale *de novo* assembly of microbial genomes from human palaeofaeces samples revealed previously undescribed gut microorganisms and a markedly higher abundance of mobile genetic elements in our ancestral gut microbiome compared to modern industrial gut microbiomes [201]. A metagenomics survey of ancient Alaskan soil confirmed that homologues of different resistance genes existed in ancient bacteria, showing that antibiotic resistance predates the modern selective pressure of clinical antibiotic use [202]. Furthermore, ancient Actinobacteria isolated from thousands-year-old Arctic and Antarctic sediments showed promising bioactivity against drug-resistant pathogens, and genome mining showed low similarity to known antibiotics [203,204]. These studies illustrate how ancient sources can lead us to evolutionary insights, as well as novel micro-organisms and chemistry.

The New Siberian Islands, located between the Laptev Sea and the East Siberian Sea, contain permafrost deposits that have been preserved for more than 200,000 years [205]. These islands are considered a time capsule to ancient biology and have been an important source of ancient animal remains, including mammoths [206]. In August 2012, an adult female woolly Mammoth (*Mammuthus primigenius*) was recovered on Maly Lyakhovsky Island (74°07′ N, 140°40′ E) [207,208]. The mammoth carcass was submerged in permafrost, exposing skull, post-cranial elements, and partial trunk. The lower part of the body was surrounded by almost pure

ice and included lower parts of the head, distal portion of the trunk, chest, abdomen, front legs and distal half of the right hind leg [207-209]. The mammoth was determined to be 28,610 years old and the remains contained exceptionally well-preserved soft tissues. The skin retained its elasticity and mummification of the carcass was minimal [207-209]. In February 2014 the specimen was transported to Yakutsk for investigation.

The finding of this extraordinarily well-conserved ancient specimen provided a unique opportunity to explore an ancient microbiome. Its surprisingly well-preserved soft tissues included the intestinal tract, allowing to extract a faecal sample for isolation of bacteria. We were specifically interested in members of the highly diverse phylum Actinobacteria, well known for their capability to produce an unprecedented diversity of specialised metabolites. Their genomes are full of biosynthetic gene clusters (BGCs) that encode the cellular machinery required for the biosynthesis of natural products. Although many bioactive compounds produced by well-known Actinobacteria have been identified, genomic research shows that a large part of the metabolic diversity of this phylum is still unexploited [27]. In this study we isolated Actinobacteria from this extraordinary ancient sample, compared the genomes to those of their closest modern-day neighbours and analysed their bioactive potential. Sequencing of six isolated Actinobacteria revealed significant phylogenetic distance to currently known strains, with yet uncharacterised biosynthetic potential.

Results

Isolation of Actinobacteria from mammoth faeces

The remarkable diversity of Actinobacteria and their specialised metabolites is the result of millions of years of evolution [104,210]. However, the evolutionary drivers that have shaped this metabolic diversity remain largely unknown. Isolation of Actinobacteria from ancient samples may allow a glimpse into the past say 10,000-100,000 of years, and thus provide insights on how natural product biosynthesis evolved over this period of time. Additionally, the microbial communities of ancient samples have rarely been explored, representing unique environments that can yield novel microorganisms and chemistry. Yet, opportunities to study ancient samples do not occur often as most archaeological findings are not in a state that does justice to the biological conditions during this era.

The discovery of an exceptionally well-preserved mammoth on Maly Lyakhovsky Island [207,208] (Fig. S1) provided a unique opportunity to recover ancient Actinobacteria and compare their genomes and biosynthetic potential to their modern descendants. From 10-14 March 2014, we joined an international team of researchers during the dissection of this extraordinary

specimen. During the dissection, every day a deeper layer of the mammoth tissue became accessible, as the specimen gradually thawed. When the thawing had proceeded far enough to explore the abdominal cavity, we found a large part of the intestines fully intact, with the omentum still attached. To extract faecal samples, the intestines were exposed and a 60 cm intestinal specimen was gathered from the remains (Fig. S1). The intestinal lumen was thoroughly inspected for defects and was found to be intact. Two perpendicular incisions were made and the intersection was folded over, exposing the intestinal lumen. Faecal samples were carefully taken from the lumen under sterile conditions, thereby avoiding any cross-contamination (Fig. S1).

A fraction of the collected faeces was homogenised in sterile dH_2O and plated onto various media selective for Actinobacteria [37]. Plates were incubated aerobically at 4 °C and 30 °C, and anaerobically at room temperature. We were able to recover some Actinobacteria from the long-frozen sample. Strains were selected based on their filamentous morphology and grown on different media for phenotypic discrimination, resulting in the isolation of six morphologically distinct strains. The majority of the strains were isolated using selective humic acid agar and/or glucose agar plates incubated at 30 °C. No filamentous bacteria were observed after anaerobic incubation.

Taxonomic and phenotypic profiling of the isolated Actinobacteria

To determine the taxonomic origin of the isolates and gain insights into the relatedness to current known bacterial species, we obtained the full genome sequences of the isolates using a combination of Nanopore and Illumina sequencing. Comparing the 16S rRNA sequence to those within EzBioCloud database revealed that the isolates belonged to five genera of Actinobacteria, namely *Sanguibacter* (M9), *Micromonospora* (M12), *Oerskovia* (M15), *Saccharopolyspora* (M46), and *Streptomyces* (M10 and M19). To obtain a more detailed taxonomic classification, a maximum-likelihood tree was constructed based on the genome sequences of the six isolates and that of 578 Actinobacteria representing six bacterial families and over 40 different genera (Fig. 1). The relatedness of the isolates and neighbouring organisms was determined by calculating the average nucleotide identity (ANI) score in an all-to-all genome comparison. Interestingly, the ANI score revealed a large phylogenetic distance between the mammoth isolates and currently known strains, with similarity scores ranging from 79-90% (Table S1). Of the isolates, the largest phylogenetic distance was found for *Streptomyces* sp. M19 and its closest neighbours with an ANI score <80%. The species delineation threshold typically lies at approximately 95% gene identity [211], which suggests that the isolates may be novel species.

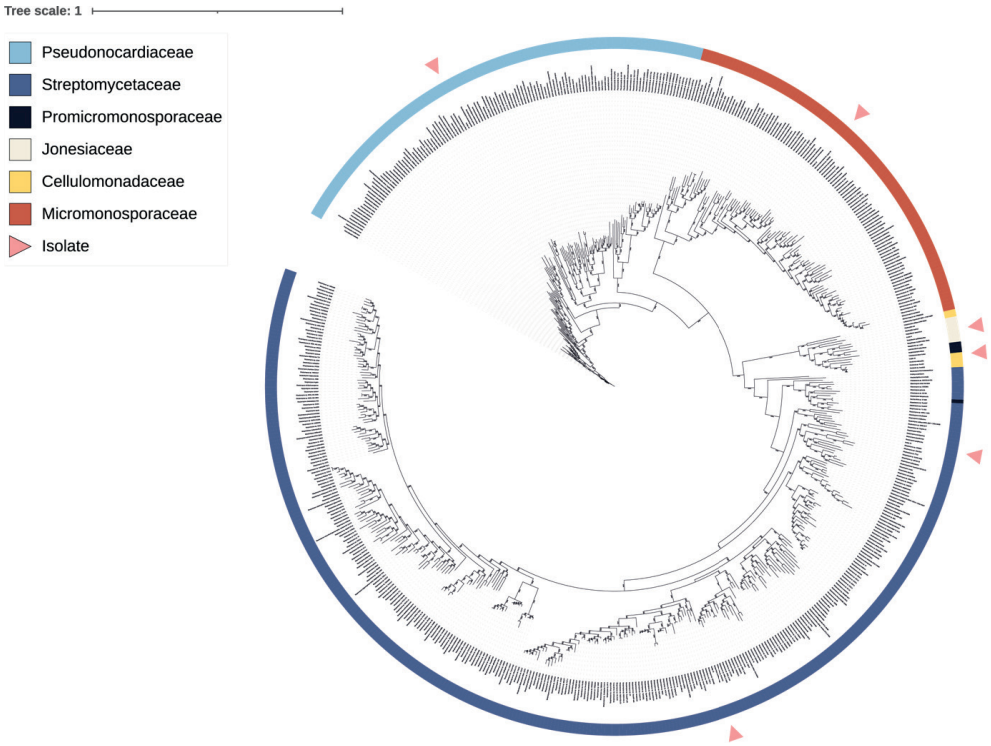


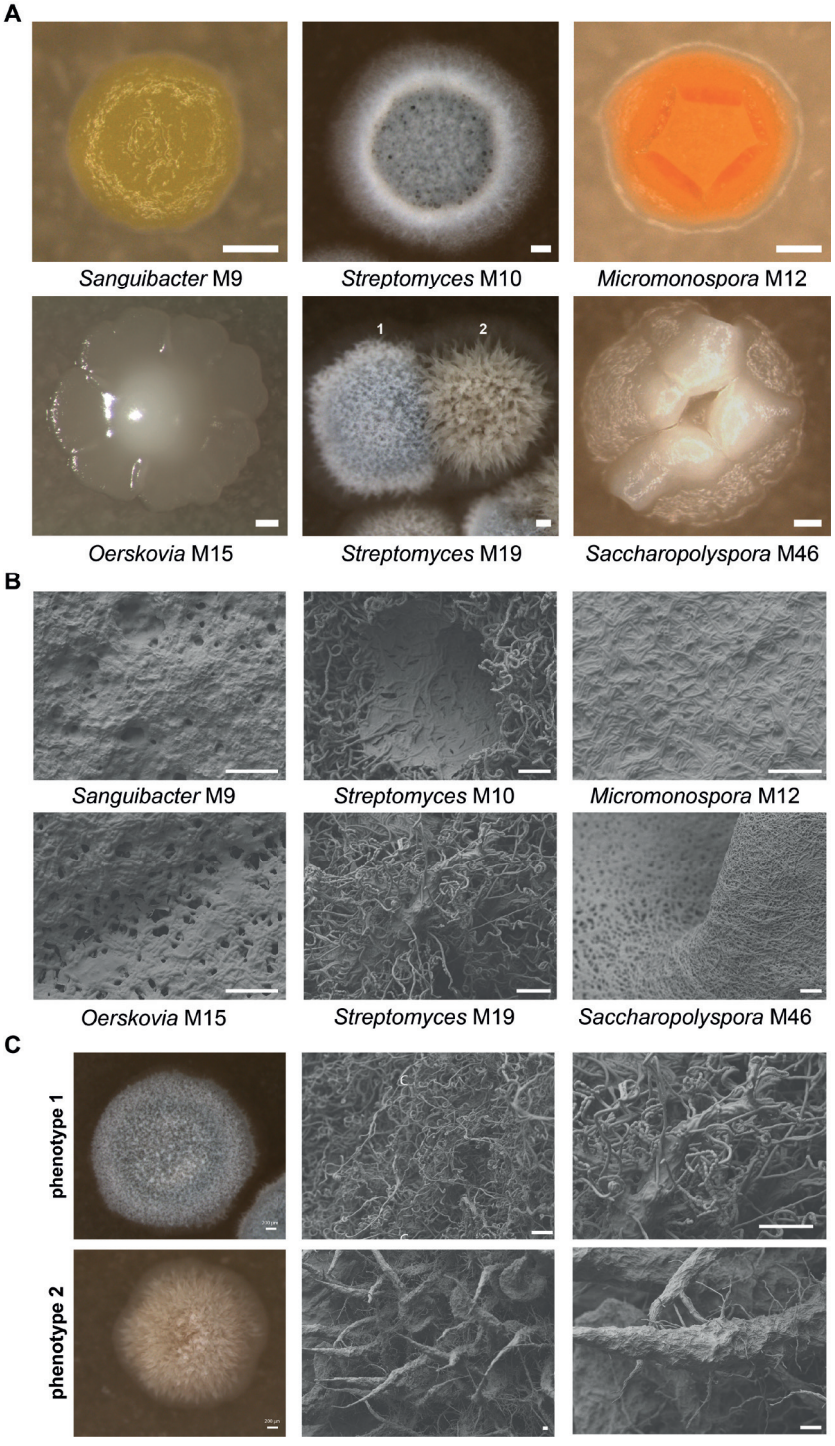
Figure 1. Phylogeny of mammoth isolates and their closest known neighbours. Maximum-likelihood tree of the six isolates compared to 578 Actinobacteria of six bacterial families (colour-coded in the outer ring), subdivided into 40 genera. The tree was rooted using *Amycolatopsis vastitatis* as outgroup and the numbers on the tree branches represent the bootstrap values in percentages of a total of 500 bootstraps. Triangles point at the location of the novel isolates. The phylogenetic analysis suggests the following likely classification based on their nearest neighbours: M9, *Sanguibacter*; M12, *Micromonospora*; M15, *Oerskovia*; M46, *Saccharopolyspora*; M10 & M19, *Streptomyces*.

The taxonomic analysis revealed a significant phylogenetic difference between the genomes of the isolates and those of their closest known neighbours. This was especially surprising for the *Streptomyces* isolates, which were compared to more than 300 genomes. We wondered whether the low similarity was a result of the ancientness of the strains or whether this was related to the underexplored environment the strains were sampled from. Therefore, we analysed the phylogenetic distance between other strains isolated from extreme environments and their closest neighbours. For this we selected the deep sea isolates *Streptomyces* sp. NTK 937 [212] and *Streptomyces* sp. SCSIO 3032 [213], the desert isolate *Streptomyces jeddahensis* [214], and *Streptomyces* sp. BF-3 and *Streptomyces* sp. 4F isolated from the Great Salt Plains in Oklahoma [215] (Table S2). In general, the results show higher relatedness to the closest neighbours of these strains (ANI scores > 93%, compared to 79-90% for our isolates), with

the exception of *Streptomyces* sp. SCSIO 3032 (ANI: 85.48%) and *Streptomyces jeddahensis* (ANI: 81.44%). It should be noted that the closest known neighbour of *Streptomyces jeddahensis* is an isolate from an insect, while this strain itself was isolated from a soil sample collected in the desert. Additionally, *Streptomyces* sp. SCSIO 3032 and its neighbour *Streptomyces* sp. MP131-18 both originate from different sampling environments although they both originate from the deep-sea. This suggests that the low level of similarity observed between the genomes of our strains and publicly available genomes may be due to the lack of bacteria isolated from similar environments, rather than the age of the strains themselves. However, we cannot ascertain this at this moment in time.

Next, we grew the isolates on SFM agar and assessed their morphology using stereomicroscopy, revealing a wide range of phenotypes (Fig. 2A). *Sanguibacter* sp. M9 produced bright yellow round colonies, *Streptomyces* sp. M10 produced cream-coloured substrate mycelia, a grey aerial spore mass, and a dark diffusible pigment, *Micromonospora* sp. M12 produced orange-coloured folded colonies, *Oerskovia* sp. M15 produced white colonies, and *Saccharopolyspora* sp. M46 produced cream-coloured folded colonies. *Streptomyces* sp. M19 displayed a heterogeneous phenotype. When this isolate was grown on SFM agar, two colony phenotypes were observed: a fully developed phenotype, and a variant with strong yellow pigmentation, sparse aerial mycelia, and lack of spores. Morphological heterogeneity was also observed within single colonies (Fig. S2), consistent with a high tendency to genetic heterogeneity [216]. Sequencing of the 16S rRNA strongly suggests that all morphological variants were indeed phenotypes of the same strain.

To obtain more insights into the morphology of the strains at high resolution, colonies were scrutinised using Scanning Electron Microscopy (SEM) (Fig. 2B). Colonies of *Streptomyces* sp. M10 produced hairy spores and the aerial mycelium consisted of both smooth and hairy hyphae, with dark-pigmented droplets on top of the colonies. *Saccharopolyspora* sp. M46 colonies consisted of a thick layer of interwoven mycelium made up of hyphae and extracellular matrix. SEM studies of the two distinct phenotypes of *Streptomyces* M19 revealed spiral spore chains in one variant (phenotype 1), while we failed to identify spores in the other (phenotype 2) (Fig. 2C). Instead, the non-sporulating colony produced large spikes made up of hyphae and extracellular matrix. These spikes could also be found in the fully developed colony, but in low abundance. Colonies of isolates *Sanguibacter* sp. M9, *Micromonospora* sp. M12, and *Oerskovia* sp. M15 were covered by an extracellular matrix and could therefore not be further characterised by SEM.


A
Sanguibacter M9
Streptomyces M10
Micromonospora M12
Oerskovia M15
1
2
Streptomyces M19
Saccharopolyspora M46
B
Sanguibacter M9
Streptomyces M10
Micromonospora M12
Oerskovia M15
Streptomyces M19
Saccharopolyspora M46
C
phenotype 1
phenotype 2

◂Figure 2. Phenotypic characterisation of Actinobacteria isolated from a faecal sample of a 28,000-year-old mammoth. Strains were grown on SFM agar plates for 9 days. **(A)** Stereomicrographs of isolates *Sanguibacter* sp. M9, *Streptomyces* sp. M10, *Micromonospora* sp. M12, *Oerskovia* sp. M15, *Streptomyces* sp. M19, and *Saccharopolyspora* sp. M46 (for phylogenetic analysis see Figure 1). *Streptomyces* sp. M19 showed two distinct phenotypes: fully developed colonies (1) and colonies with a bald appearance (2). Scale bar: 200 μm. **(B)** Scanning electron micrographs of isolates M9, M10, M12, M15, M19, and M46. Scale bar: 10 μm. **(C)** Scanning electron micrographs showing significant differences in morphology between the two colony phenotypes observed when *Streptomyces* sp. M19 is grown on SFM. Images of phenotype 1 show spirals of spore chains. Images of phenotype 2 show large spikes made up of hyphae and extracellular matrix. Also in the fully developed colony, such spikes can be found but in low frequency.16S rRNA sequencing strongly suggests that these morphological variants are phenotypes of *Streptomyces* sp. M19. Scale bar: 10 μm.

Overall, the taxonomic and phenotypic analyses show that we have isolated a diverse collection of Actinobacteria, of which *Streptomyces* sp. M19 stands out with a highly heterogeneous phenotype and low similarity to its closest neighbours. Moreover, the low ANI scores suggest that all isolates represent novel species yielding unexplored genomic information that is not yet present in current databases.

Potential of the strains in terms of natural product biosynthesis

We then studied the biosynthetic gene clusters (BGCs) of the isolates, to obtain an idea of how related they were to BGCs that are currently available in the databases. For this, the genome sequences were analysed using antiSMASH [217]. This identified a total of 179 putative clusters, namely four in *Oerskovia* sp. M15, six in *Sanguibacter* sp. M9, 22 in *Saccharopolyspora* sp. M46, 43 in *Micromonospora* sp. M12, 50 in *Streptomyces* sp. M10 and 54 in *Streptomyces* sp. M19 (Fig. 3A). Over 70% of the total BGCs shared less than 50% KnownClusterBlast similarity to BGCs within the MIBiG database, with 23% showing no significant similarity to any known BGC. To determine how unique the high level of uncharacterised biosynthetic potential is to ancient actinobacterial isolates, we compared the potential of our isolates to the potential of Actinobacteria isolated from different environments using the antiSMASH database (Table S3). This showed that similarly diverse uncharacterised biosynthetic potential can be found in Actinobacteria isolated from other environments, such as soil. Indeed, even in the well-studied model organism *Streptomyces coelicolor*, several BGCs still show less than 50% KnownClusterBlast similarity to other known BGCs. Therefore, the uniqueness of the biosynthetic potential may not be related to the age of the strains, but may instead be widespread even in contemporary and well-studied *Streptomyces* species. BiG-SCAPE analysis of the genomes showed that the genomes contained relatively high numbers of genes encoding enzymes associated with the biosynthesis of terpenes, in particular for *Micromonospora* sp. M12 (five), *Streptomyces* sp. M19 (four), *Streptomyces* sp. M10 (seven), and *Saccharopolyspora* sp. M46 (six).

The genome of *Streptomyces* sp. M19 showed the largest distance in terms of ANI scores to its nearest neighbours. We therefore subjected the strain to detailed analysis of its biosynthetic potential, to obtain insights into the chemical diversity of its specialised metabolites. The M19 genome had a similar number of predicted BGCs (34) as its nearest neighbours *Streptomyces* sp. CNZ306 (33 BGCs) and *Streptomyces* sp. CNQ-509 (28 BGCs) (Fig. 3B). Surprisingly, *Streptomyces* sp. M19 only shares two BGCs with these nearest neighbours, namely for desferrioxamines and for ectoine. The M19 genome encodes a high percentage of non-ribosomal peptide synthetases (NRPSs) and polyketide synthases (PKSs), and also has BGCs for butyrolactones, unusual polyketides / fatty acids (linked to a heterocyst glycolipid synthase-like PKS, and a ladderane-lipid-associated ketosynthase), a likely lasso peptide, an aryl polyene, and an aminoglycoside/aminocyclitol; none of these natural product classes could be found in modern nearest neighbours. Comparative genomic analysis of *Streptomyces* sp. M19 and its neighbours, CNZ306 and CNQ-509, displays the low similarity of approximately 79% and only few corresponding BGCs (Fig. 4). While a high degree of overlap was seen between the genomes of *Streptomyces* sp. CNZ306 and CNQ-509 in terms of overall genome synteny and in shared BGCs (Fig. S3), precious little homology was seen with the genome of *Streptomyces* sp. M19 (Fig. 4).

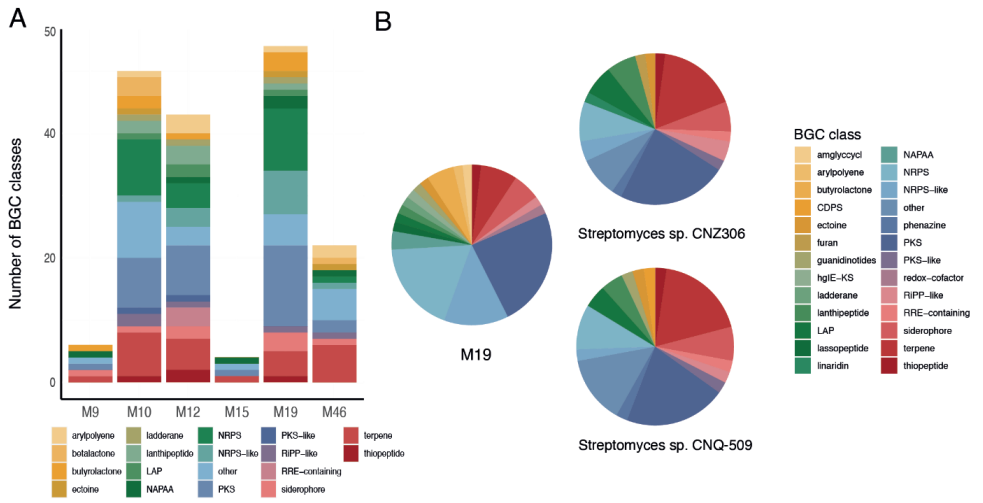


Figure 3. Predicted BGC classes for the six actinobacterial isolates. A) BGC classes for each isolate predicted with antiSMASH v6.0. Known types of BGCs representing <1% of all BGCs were grouped into the "other" category. The large number of BGCs in *Streptomyces* species M10 and M19, *Micromonospora* M12 and *Saccharopolyspora* M46, is particularly noteworthy. **B)** Overview of all predicted BGC classes for isolate M19 and its closest neighbours *Streptomyces* sp. CNZ306 and *Streptomyces* sp. CNQ-509. The comparison shows that only two BGCs are shared between all three isolates, namely BGCs for desferrioxamines and for ectoine. Moreover, the genome of *Streptomyces* sp. M19 contains a greater variety of BGC classes, with a surprisingly high proportion of NRPS BGCs.

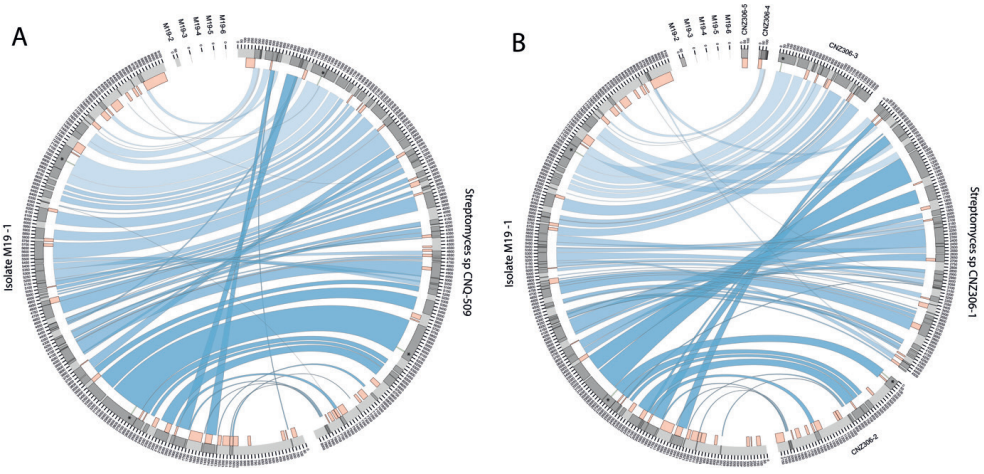


Figure 4. Comparative analysis between isolate M19 and its modern-day closest known neighbours *Streptomyces* sp. CNZ306 and *Streptomyces* sp. CNQ-509. A) synteny plot comparing *Streptomyces* sp. M19 to *Streptomyces sp. CNQ-509*. The blue lines are obtained by a pairwise BLASTP and collinearity calculation method. Resulting locations represent high similarity regions. In red, the location of predicted BGCs of antiSMASH v6.0 are indicated. Shared BGCs are indicated in green and with an asterisk. **B)** synteny plot of *Streptomyces* sp. M19 and *Streptomyces sp. CNZ306.*

Antibiotic activity and bioactive metabolites produced by the mammoth isolates

Next, the antibiotic-producing potential of the strains was assessed under six different culturing conditions. These were four agar-based media, namely the nutrient-rich Nutrient Agar (NA) and International *Streptomyces* Project 2 (ISP2) and the nutrient-poor Czapek Dox and Minimal Media supplemented with mannitol and glycerol (MM), and the liquid media versions of ISP2 and MM. We used the Gram-positive *Bacillus subtilis* 168 and the Gram-negative *Escherichia coli* ASD19 and *Pseudomonas aeruginosa* PA01 as indicator strains. Different types of growth inhibition were observed, namely complete inhibition, strong reduction in number of colony-forming units (cfu), and impaired growth (Fig. 5). Overall, the isolates displayed the strongest antibacterial activity when grown on NA. Therefore, further metabolic analysis was done on samples isolated from NA-grown cultures.

Besides soluble natural products, also volatile compounds (VCs) may have antibacterial activity [218,219]. To investigate whether (some of) the antimicrobial activity may have been due to the production of antibacterial VCs, the isolates were grown on plates where the Actinobacteria were separated from the indicator strains by an impermeable polystyrene divider. Growth of *E. coli* was completely inhibited by VCs produced by isolates *Streptomyces* sp. M10, *Streptomyces* sp. M19, and *Saccharopolyspora* sp. M46 (Fig. S4). While *Micromonospora* sp. M12 did not display antibacterial activity in the bioactivity assay described above, it partially inhibited growth

of *E. coli* in the volatile assay. None of the strains produced VCs with activity against *B. subtilis*. These data indicate that the observed antibacterial activity of strains M10, M19, and M46 against *E. coli* was at least in part caused by VCs, while the bioactivity of strains M10, M15, M19, M46 against *B. subtilis* was solely caused by the production of soluble antibiotics.

	B subtilis						*E. coli*						*P. aeruginosa*					
	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6
M9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0%
M10	0.7	1.1	0	0	0.6	1.7	0	0	0	0	0	2.2	0	0	0	0	0	IG
M12	0.6	0.6	0	0	0	0	0.6	0.5	0	0	0	0	0	0	0	0	0	0%
M15	0	0	0	0	0	0.6	0	0	0	0	0	0	0	0	0	0	0	0%
M19	1.9	0	1.8	1.7	0.6	1.1	0	0	0	0.9	0	3.2	0	0	0	0	0	0%
M46	0	0	0	0	0	3.6	0	0	0	0	0	3.3	0	0	0	0	0	IG

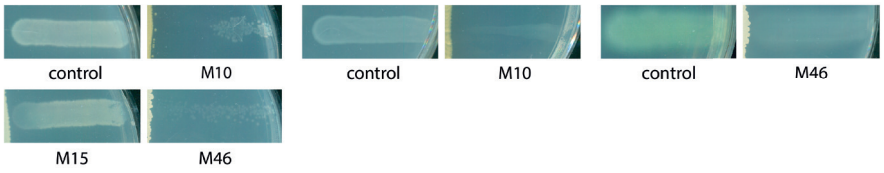


Figure 5. Antimicrobial activity of the actinobacterial isolates. After 7 days of growth, the bioactivity of the mammoth isolates was assessed against different indicator strains using different growth media and methods: 1 = MM, soft agar overlay; 2 = Czapek Dox, soft agar overlay; 3 = liquid culture ISP2; 4 = liquid culture MM; 5 = ISP2, cross streak, 6 = NA, cross streak. The zone of inhibition (cm) is indicated for each isolate and related to a colour scale ($n = 3$). Most activity was observed in the cross-streak assay on NA. The fields selected with a black border refer to the examples displayed below the table that illustrate the different types of growth inhibition observed. IG: impaired growth.

To gain more insights into the soluble antibiotics produced by isolates *Oerskovia* sp. M15, *Streptomyces* sp. M10, *Streptomyces* sp. M19, and *Saccharopolyspora* sp. M46, the strains were streaked on NA plates and grown for seven days. Metabolites were extracted using ethyl acetate (EtOAc) and tested for bioactivity against the different indicator strains. The crude extracts of *Streptomyces* sp. M10 and M19 showed activity against *B. subtilis*, while *E. coli* and *P. aeruginosa* were not inhibited (Fig. S5). MS/MS data were analysed using Global Natural Products Social molecular networking (GNPS) [220], resulting in a molecular network containing 2886 nodes clustered in 223 spectral families (Fig. 6). The highest number of unique nodes (491) was attributed to isolate M10, while the lowest number (280) was attributed to isolate M46. 44 nodes were unique to *Streptomyces* isolates M10 and M19. Dereplication based on matching MS/MS spectra against the GNPS spectral library annotated several *m/z* values as known natural products that were not present in the medium blank (Table S4), including *m/z* value 546.4879 $[M + H]^+$ as the hopanoid aminobacteriohopanetriol. Additionally, several *m/z* values were annotated as being known bioactive metabolites: antimycin A1 (1), antimycin A2 (2), and the macrotetrolide monactin (3) and

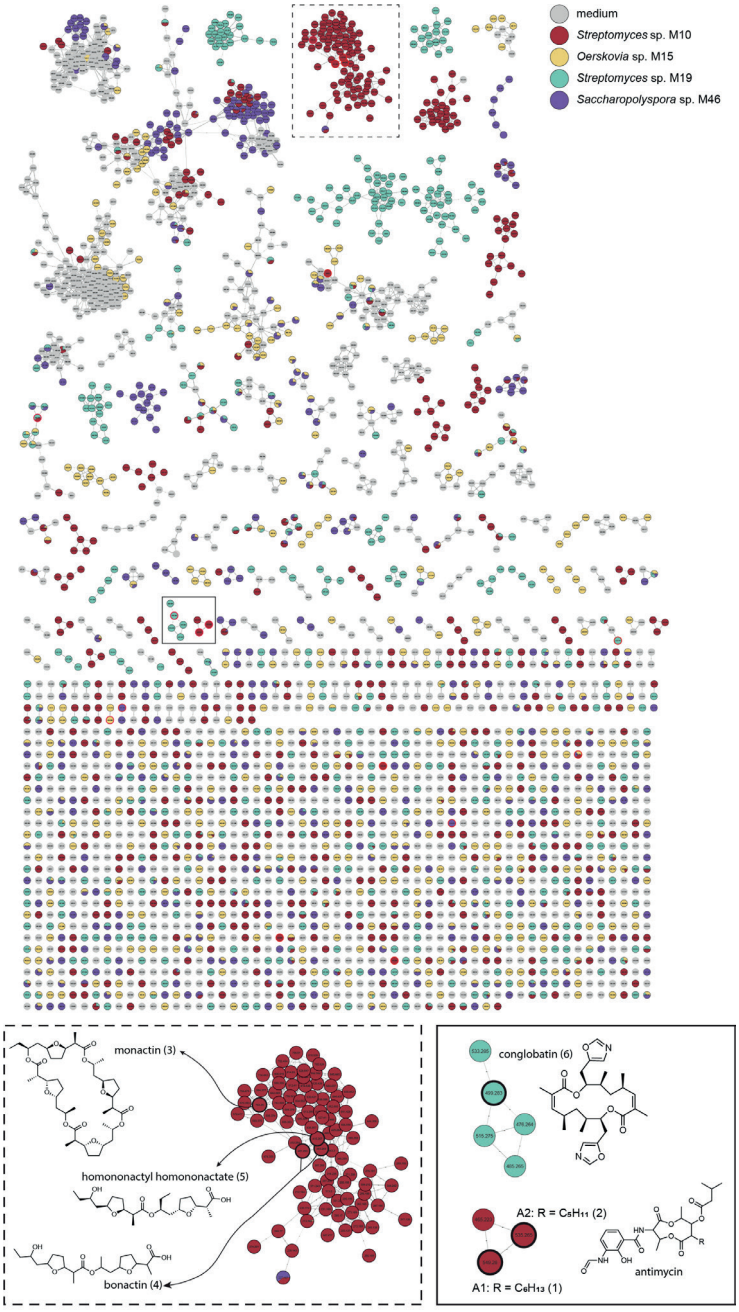


Figure 6. Molecular network of the ions detected in the crude extracts of the bioactive isolates revealing the presence of multiple known bioactive compounds. A pie chart was mapped to the nodes which represents the abundance of each *m/z* value in the crude extracts of the different bioactive isolates (M10, M15, M19 and M46) and the medium blank. Nodes highlighted represent dereplicated known bioactive metabolites. The isolates were grown for seven days on NA plates ($n = 3$).

related bonactin (4) and homononactyl homononactate (5) in the extracts of M10, and conglobatin (6) in the extracts of M19 (Fig. 6). To confirm these findings, we analysed the KnownClusterBlast output of antiSMASH for presence of the responsible BGCs. In the genome of M10, antiSMASH identified an antimycin BGC and a BGC with 75% similarity to a macrotetrolide BGC (MIBiG cluster BGC0000243). The macrotetrolide-associated BGC from strain M10 lacks three genes compared to the reference BGC, two encoding hypothetical proteins and one encoding an inositol monophosphatase-like enzyme. M19 harbours a BGC with low (36%) similarity to the conglobatin BGC (Fig. S6). Although the overall match for this cluster is lower, we could detect homologues of all known core conglobatin genes (*congA-E*) with the same domain architecture as in the reference BGC. M10 and M19 contain a hopene cluster with similarity scores of 84% and 76%, to MIBiG cluster BGC0000663, which is primarily due to the absence of two and three hypothetical genes, respectively. Taken together, the isolates produced some known bioactive molecules, but the majority of the mass features could not be dereplicated.

Discussion

A huge proportion of the bacterial world remains uncharacterised, representing an important reservoir of biological information and new chemistry as the possible basis for future drugs [41,200,221]. Metagenome sequencing, varying culturing techniques, and isolation of bacteria from rare environments is gradually revealing part of this microbial dark matter, and ancient samples are a promising and underexplored resource [201,202]. We had the unique opportunity to sample faeces from the intestinal tract of an exceptionally well-preserved 28,000-year-old mammoth, which we used to isolate ancient Actinobacteria [207]. These Actinobacteria belong to the well-studied genus *Streptomyces* and to the underexplored genera *Sanguibacter, Micromonospora, Oerskovia,* and *Saccharopolyspora.* Genomic analysis of the strains revealed a large phylogenetic distance to currently known strains and much uncharacterised biosynthetic potential. With a mammoth, multiple replicates are not readily obtained, to say the least. Therefore, minimizing the contamination risk during the excavation process is particularly important. The specimen was transported and stored in frozen condition. As thawing set in from the outside of the carcass and slowly proceeded inward, each day freshly thawed samples could be taken. As the intestines that we encountered were still intact, we could open them under sterile conditions and extract faeces with a low risk of external contamination. Hence, a lot of attention was paid to the careful removal of the faecal sample, and the specific sample we eventually used for isolation of the Actinobacteria was then extracted from the very core of the large faecal sample, so as to eliminate the risk of contamination to the absolute minimum. The intestinal lumen was thoroughly expected for defects and faeces was taken and stored using sterile materials.

A previous metagenome-based study revealed the presence of actinobacterial reads in different tissues of mammoth remains [222]. In this study, we have now been able to isolate Actinobacteria from such an environment, strengthening these findings and allowing us to further explore their functional role and potential. Studying metagenome data always has the risk for misassembly of (damaged) DNA, which is not the case when isolates are grown up and then sequenced. Taxonomic analysis revealed a significant phylogenetic difference between the isolates and their closest known neighbours, strongly suggesting that these isolates represent novel species. The low similarity to known strains can be attributed to the limited number of publicly available complete genomes for underexplored taxa such as *Oerskovia, Sanguibacter* and *Saccharopolyspora,* with three, four and eight available complete genome sequences, respectively, as compared to 323 complete *Streptomyces* genomes. Nevertheless, our *Streptomyces* isolates also showed low similarity scores, while many publicly available complete genomes were available for comparison. The mammoth isolates are therefore a valuable addition to the database and illustrate the power of harvesting microbes from unique environments, even for well-studied genera.

Ancient samples allow a glimpse into the past and may provide evolutionary insights. An interesting question that arises is what the significance is of the major differences that we observe between the isolates and their closest known relatives. The phylogenetic distance between the isolates and their known relatives makes it hard to evaluate evolutionary differences on a genomic level. For better comparison, genome sequences of more closely related strains are required; however, it is yet not known if such strains can be found in the available strain collections.

Genome mining of the isolates revealed a wide variety of BGCs, with 19% encoding PKSs, 14% NRPSs and 14% terpene biosynthetic enzymes. Most of the predicted BGCs showed low or no similarity to any known BGCs. Comparative genome analyses between M19 and its taxonomic neighbours found in publicly available sequence data, *Streptomyces* sp. CNZ306 and *Streptomyces* sp. CNQ-509, revealed a shared conserved internal region and a less conserved region near the ends of the chromosome, holding high numbers of unique and uncharacterised BGCs. This is consistent with results from previous studies showing that common BGCs are often located in the internal regions of the chromosome, while more unique genes are located towards the subtelomeric regions [21,42,115]. Surprisingly, only two BGCs, encoding the desferrioxamine and ectoine pathways, were shared between isolate *Streptomyces* sp. M19 and its neighbours. These specialised metabolites play and important role in survival and their BGCs are well conserved among Actinobacteria [42,223]. The limited amount of shared biosynthetic content between *Streptomyces* sp. M19 and its neighbours is

unexpected as phylogeny has been shown to be an important indicator of BGC distribution [100,224]. This underlines the opportunities offered in terms of the biosynthetic potential of bacteria isolated from underexplored environments.

The majority of the isolates displayed bioactivity against one of the tested indicator strains, both through production of volatile and non-volatile compounds. The antibacterial activity against Gram-positive bacteria by *Oerskovia* sp. M15 is surprising, because to the best of our knowledge, antibacterial activity has never been reported before for members of this genus. Four potential BGCs were identified in the genome sequence of M15, of which three had similarity scores below 12%. However, the crude extracts did not show any activity and chemical dereplication did not give any hits with known bioactive compounds. Future work in our laboratory will focus on identifying the BGC and cognate natural product linked to the bioactivity of *Oerskovia* sp. M15. The crude extracts of *Streptomyces* isolates M10 and M19 exhibited antibiosis against *B. subtilis*. Molecular networking and chemical dereplication using the GNPS platform resulted in the annotation of several known natural products, including antimycin and monactin in the extracts of M10, and conglobatin in the extracts of M19. This was supported by the detection of their BGCs by antiSMASH. Interestingly, antimycin- and conglobatin-related molecules were also detected in the crude extracts of polar Actinobacteria isolated from ancient sediment cores [203]. An evolutionary path of the antimycin BGC has been proposed in which the L-form BGC is appointed as the ancestor of other antimycin BGCs [225]. Therefore, it's not surprising that our detected ancient BGC is the ancestral L-form antimycin cluster. Additionally, analysis of the biogeographical and phylogenetic distribution of the antimycin BGC has revealed that this BGC is widespread among *Streptomyces spp.* and across the globe [225]. While these known metabolites might be responsible for the bioactivity of the crude extracts, other unknown metabolites with antibacterial activity might be produced as well. The GNPS network of the metabolome of these isolates revealed several spectral families unique to the isolates which could not be dereplicated. Proteomics or transcriptomics could be used to analyse whether other BGCs are expressed under these growth conditions.

Much interest has been directed towards the presence of antibiotic-producing Actinobacteria in the mammalian microbiome as this could point towards a protective role against infection by pathogens [99,181]. However, these bacteria have been isolated from a faeces sample and we therefore cannot distinguish whether these strains are part of the mammoth microbiome or whether they are simply passing through as part of digested plant material.

Overall, we have isolated a diverse collection of Actinobacteria from a unique ancient stool sample. The large phylogenetic distance between the isolates and their modern siblings and

the high percentage of uncharacterised biosynthetic potential, even in the well-studied genus *Streptomyces*, illustrate that we have by far captured all microbial diversity. Future studies on the evolutionary differences between the isolates and their modern siblings may allow for a unique glance into history.

Methods

Sample acquisition

In February 2014 the specimen was transported by truck to Yakutsk. Dissection of the specimen was done in the autopsy room of the medical faculty of the university of Yakutsk. The samples for microbiological analysis were extracted from 10-14 March 2014, the timespan in which the remains completely thawed. As thawing set on from the outside of the carcass and slowly proceeded inward, each day freshly thawed samples could be taken. Approached from the right side, the intestines where exposed on March 13 after carefully removing soft-tissue and ribs over the previous days. A 60 cm intestinal specimen was taken out of the remains and placed on a sterile surface for further examination. No defects of the intestinal lumen were found. After inspection, two perpendicular (5 cm) incisions where made, and the intersection was folded over, exposing the intestinal lumen. Faecal samples were carefully taken from the lumen using flocked swab collecting tubes (eSwab Copan) or deposited in sterile collection tubes with disposable tweezers with minimal possible cross-contamination. Sterile examination gloves and instruments were used during the whole procedure.

Isolation of Actinobacteria

Approximately 100 mg of faeces was aseptically placed into a sterile Eppendorf, dissolved in dH_2O, and serially diluted (10^{-1}–10^{-6}). Serial dilutions were plated onto different agar media. The media were glucose agar (GA) [138], humic acid agar (HA) [226], mannitol soya flour medium (SFM) [227], modified Starch-Casein agar (MSCA) [228], minimal medium without carbon sources (MM) [227], MM + 1% glycerol (w/v) (Y), MM + 1% mannitol (w/v) (A). All media contained nystatin (50 μg/mL) and nalidixic acid (10 μg/mL) for the inhibition of fungi and Gram-negative bacteria respectively. Plates were incubated at 30 °C, 4 °C, and anaerobically at room temperature. Single actinomycete colonies were streaked onto SFM agar plates until pure and cryopreserved with glycerol (20 %) and stored at −80 °C

Genome sequencing

Strains were cultured in TSBS at 30 °C with 200 rpm shaking speed. Genomic DNA was isolated by phenol-chloroform extraction as described previously [227] and sent to be commercially sequenced at Future Genomics Technologies, The Netherlands. Genomes were sequenced

using the MinION Nanopore sequencing platform and Illumina NovaSeq6000. Hybrid assembly (both Illumina and ONT reads) was performed for each isolate using Unicycler (v0.4.0.7) [229]. Briefly, Unicycler performs a SPAdes assembly of the Illumina reads and then scaffolds the assembly graph using long reads. Unicycler polishes its final assembly with Illumina reads and uses Pilon [230] to reduce the rate of small base-level errors.

Phylogenetic analysis

The initial taxonomy of the six isolates was determined using 16S rRNA sequences and BLASTN, with the highest scoring sequencing hits reported. To determine the phylogenetic class of the isolates on a whole genome scale, 578 high quality Actinobacteria genome sequences have been downloaded from the PhyloPhlAn v3.0 [231] database, using the phylophlan_setup_database function. Next, the phylophlan_write_config_file script is employed to create a configuration file with DIAMOND as mapping tool, MAFFT for the multiple sequence alignment, trimAl for alignment trimming and IQ-TREE for generating a phylogenetic tree. FastANI v1.32 [232] was used to calculate the relatedness of the isolates and neighbouring strains.

Bioactive potential and comparative genome analysis

AntiSMASH v6.0 [217] was used under default settings to predict BGCs from the six isolated bacteria and their modern-day closest known neighbours. The resulting predicted BGCs were then used as input for BiG-SCAPE [108] to analyse BGC clustering.

To annotate the genomes and identify coding regions, Prokka v1.14.6 [233] was applied on the assemblies of the isolates. The strain *Streptomyces sp.* CNQ-509 was downloaded from NCBI using accession number GCA_001011035.1. The EMBL-EBI database [234] was questioned to obtain the sequence information of strain *Streptomyces* sp. CNZ306 (PGEW01000000). Next, BLASTP was used to identify similarity between the isolates and their neighbours, whereafter MCScanX [235] was utilised to predict gene collinearity and synteny. Finally, the coordinates of the collinear regions and predicted BGCs were used as input for the visualizing tool Circos v0.69-8 [236].

Stereomicroscopy and Scanning Electron Microscopy (SEM)

Isolates were grown for nine days on SFM. Stereo microscopy was done using a Leica MZ16 FA microscope equipped with a Leica DFC420 C camera. SEM studies were performed using a JEOL JSM-7600F scanning electron microscope. Single colonies were excised from agar plates and the bottom layer of agarose was cut off to minimise the thickness of the sample. The sections were glued upon a Cryo-EM stub and the whole stub was submerged in non-boiling liquid nitrogen and frozen for 20 sec. The stub was then transferred to the gatan Cryo unit, heated to -90 °C for 2 minutes to remove any ice crystals formed during the transfer,

cooled to -120 °C, and coated with gold/palladium (80/20) using a sputter coater. Hereafter, the samples were transferred into the microscope and kept at -120 °C while imaging.

Antimicrobial activity assays

Bacillus subtilis 168, *Escherichia coli* ASD19 [237], and *Pseudomonas aeruginosa* PA01 were used as indicator strains for antimicrobial activity and were cultured in LB media at 37 °C. Antimicrobial activity assays were conducted in liquid and on plate, using different methods:

Liquid cultures: Strains were grown in ISP2 (DSMZ #987) and NMMP [227] medium for seven days. Wells were performed in soft LB agar (1.8% w/v agar) containing one of the indicator strains pre-grown in liquid LB to exponential phase (OD600 = 0.4 – 0.6) and filled with 100 μL culture supernatant. Plates were incubated overnight at 37 °C (± 18 hours) and the following day, the zone of inhibition was determined.

Cross streak method: each strain was independently inoculated on NA (Difco) and ISP2 agar plates as a single streak in the centre of the plate and incubated for seven days to allow the strains to grow and produce antibiotics. The plates were then seeded with the indicator strains pre-grown in liquid LB to exponential phase (OD600 = 0.4 – 0.6) by streaking perpendicular to the line of actinobacterial growth and incubated overnight at 37 °C (± 18 hours). The following day, the zone of inhibition was determined.

*The double-layer agar method***:** Strains were manually spotted (2 μL) on minimal medium agar plates (MM) supplemented with 0.5% mannitol and 1% glycerol (w/v) as non-repressing carbon sources, and Czapek Dox plates. After seven days of incubation, plates were overlaid with soft LB agar (1.8% w/v agar) containing one of the indicator strains pre-grown in liquid LB to exponential phase (OD600 = 0.4 – 0.6) and incubated overnight at 37 °C (± 18 hours). The following day, the zone of inhibition was determined.

Volatile assay: The antimicrobial activity of volatile production was assessed using a petri dish with two equally sized compartments, both filled with NA. Mammoth isolates were streaked on one site and plates were incubated for seven days after which, *E. coli* or *B. subtilis* were inoculated on the other side using a concentration of 10^4 and 10^3 CFU/mL respectively.

Metabolite profiling

Isolates M10, M15, M19 and M46 were grown confluently on NA plates for seven days. The agar plates were cut into small pieces, soaked overnight in ethyl acetate (EtOAc) to extract the metabolites, evaporated at room temperature, and dissolved in methanol (MeOH) to a concentration of 1 mg/mL. LC-MS/MS acquisition was performed using Shimadzu Nexera X2

ultra high-performance liquid chromatography (UPLC) system, with attached photodiode array detector (PDA), coupled to Shimadzu 9030 QTOF mass spectrometer, equipped with a standard electrospray ionisation (ESI) source unit, in which a calibrant delivery system (CDS) is installed. A total of 2 μL was injected into a Waters Acquity HSS C18 column (1.8 µm, 100 Å, 2.1 × 100 mm). The column was maintained at 30 °C and run at a flow rate of 0.5 mL/min, using 0.1% formic acid in H2O, and 0.1% formic acid in acetonitrile (ACN) as solvents A and B, respectively. The gradient used was 5% B for 1 min, 5–85% B for 9 min, 85–100% B for 1 min, and 100% B for 4 min. The column was re-equilibrated to 5% B for 3 min before the next run was started. The PDA acquisition was performed in the range of 200–600 nm, at 4.2 Hz, with 1.2 nm slit width. The flow cell was maintained at 40 °C. All the samples were analysed in positive polarity, using data dependent acquisition mode. In this regard, full scan MS spectra (*m/z* 100-1700, scan rate 10 Hz, ID enabled) were followed by two data dependent MS/MS spectra (*m/z* 100–1700, scan rate 10 Hz, ID disabled) for the two most intense ions per scan. The ions were fragmented using collision induced dissociation (CID) with fixed collision energy (CE 20 eV), and excluded for 1 s before being re-selected for fragmentation. The parameters used for the ESI source were: interface voltage 4 kV, interface temperature 300 °C, nebulizing gas flow 3 L/min, and drying gas flow 10 L/min.

MS/MS-based molecular networking and dereplication

MS/MS raw data (obtained from Shimadzu 9030 QTOF MS) were converted to a 32-bit mzML file using MSConvert (ProteoWizard) and a molecular network was assembled using the online workflow on the Global Natural Product Social Molecular Networking (GNPS) website [220]. Both the precursor ion and the MS/MS fragment ion mass tolerance were set to 0.02 Da. The minimum cosine score was set to 0.7 and the minimum matched peaks set to 6. The MSCluster algorithm was run with a minimum cluster size of 2 spectra. The spectra in the network were searched against the GNPS spectral libraries. For this, the precursor ion and the MS/MS fragment ion mass tolerance were set to 0.5 Da. Matches between network spectra and library spectra required a minimum score of 0.7 and at least 6 matched peaks. Cytoscape 3.8.2 was used for visualization of the generated molecular networks [238]. The edge thickness was set to represent the cosine score, with thicker lines indicating higher similarity between nodes. The molecular networking job in GNPS can be accessed at https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=e160b564fc7e48e6b82394991bfd79be.

Bioactivity crude extracts

The activity of the crude extracts was determined in triplicate. Indicator strains were pre-grown in liquid LB to exponential phase (OD600 = 0.4 – 0.6). Cultures were diluted to OD600 = 0.01 in LB and 100 μL diluted culture was loaded in wells of a 100-well honeycomb plate. 200

*μ*g crude extract was added to the bioactivity of the crude extract. Additionally, the following controls were added: LB, bacterial dilution (growth control), bacterial cells + 6 *μ*g ampicillin (positive control), and bacterial cells + MeOH (negative control). Subsequently, the optical density at 600 nm was measured every 30 min for 16 hours using a Bioscreen C Reader (Thermo Scientific, Breda, The Netherlands), with continuous shaking. The absorption at 600 nm was plotted against the time.

Supplementary information for Chapter 3

Table S1. Isolate taxonomic classification and identified closest neighbours.

Isolate ID	Taxonomic classification	Closest known neighbour(s)	Accession number	ANI score %
M9	*Sanguibacter*	*Sanguibacter antarcticus*	GCA_002564005.1	84.2
M10	*Streptomyces*	*Streptomyces* sp. CB02460	GCA_001905705.1	84.0
		Streptomyces sp. 2131.1	GCA_900105515.1	84.2
M12	*Micromonospora*	*Micromonospora chokoriensis*	GCA_900091505.1	89.5
M15	*Oerskovia*	*Oerskovia enterophila*	GCA_001624335.1	88.6
		Oerskovia sp. Root22	GCA_001429135.1	88.5
		Oerskovia sp. Root918	GCA_001428945.1	88.6
M19	*Streptomyces*	*Streptomyces* sp. CNZ306	PGEW01000000	79.8
		Streptomyces sp. CNQ-509	GCA_001011035.1	80.0
M46	*Saccharopolyspora*	*Saccharopolyspora flava*	GCA_900116135.1	88.1

Table S2. Isolates from extreme environments and their similarity scores with closest neighbours.

Isolate ID	Origin	Accession number	Closest known neighbour(s)	Origin neighbour(s)	Accession number neighbour(s)	ANI score (%)
Streptomyces sp. NTK 937	Deep-sea, Canary Basin, Canary Islands	GCA_000698495.1	*Streptomyces* sp. SM18	Deep-sea, Kilkieran Bay, Galway, Ireland	GCA_002910775.2	96.7
Streptomyces sp. SCSIO 3032	Deep-sea, Madeira archipelago, Portugal	GCA_002128305.1	*Streptomyces* sp. MP131-18	Deep-sea, Trondheim, Norway	GCA_001984575.1	85.5
Streptomyces jeddahensis	Desert, Jeddah, Saudi Arabia	GCA_001642995.1	*Streptomyces* sp. AmelKG-E11A	Insect	GCA_002705975.1	81.4
Streptomyces sp. BF-3	Great Salt Plains, Oklahoma	GCA_002104865.1	*Streptomyces* sp. Cmuel-A718b	Unknown	GCA_900092005.1	98.9
Streptomyces sp. 4F	Great Salt Plains, Oklahoma	GCA_002104855.1	*Streptomyces* sp. Alain-F2R5	Al Ain, Dubai, United Arab Emirates	GCA_002277855.1	93.1
			Streptomyces sp. CB02414	Beach soil, Dubai, United Arab Emirates	GCA_001905385.1	99.2

Table S3. Percentage of BGCs that share less than 50% KnownClusterBlast similarity to BGCs within the MIBiG database of the mammoth isolates and a selection of Actinobacteria

Strain	Isolation source	total # BGCs	% BGC < 50% similarity*
***Sanguibacter* sp. M9**	mammoth faeces	6	67%
Sanguibacter gelidistatuariae ISLP-3	ice sculpture, Antarctica	3	67%
Sanguibacter keddieii DSM 10542	bovine blood	5	80%
Sanguibacter massiliensis	human stool	2	50%
***Micromonospora* sp. M12**	mammoth faeces	28	79%
Micromonospora auratinigra DSM 44815	soil, Thailand	16	88%
Micromonospora chokoriensis DSM 45160	sandy soil	15	73%
Micromonospora inosilota DSM 43819	forest soil	9	89%
Micromonospora narathiwatensis DSM 45248	peat swamp forest soil	18	67%
Micromonospora terminaliae DSM 101760	plant	13	77%
Micromonospora zamorensis DSM 45600	rhizosphere	12	83%
***Oerskovia* sp. M15**	mammoth faeces	4	75%
Oerskovia sp. KBS0722	soil	6	83%
Oerskovia sp. Root22	roots of *Arabidopsis thaliana*	3	33%
***Saccharopolyspora* sp. M46**	mammoth faeces	20	80%
Saccharopolyspora coralli E2A	stony coral	10	70%
Saccharopolyspora erythraea NRRL 2338	soil	36	72%
Saccharopolyspora sp. ASAGF58	soil	29	69%
***Streptomyces* sp. M10**	mammoth faeces	31	65%
***Streptomyces* sp. M19**	mammoth faeces	34	65%
Streptomyces antibioticus DSM 40234	soil	26	50%
Streptomyces sp. MP113-05	sponge	28	79%
Streptomyces cacaoi H2S5	moss soil	31	45%
Streptomyces coelicolor CFB_NBC_0001	soil	27	41%
streptomyces ficellus NRRL 8067	soil	26	73%
streptomyces galilaeus ATCC 14969	soil	24	50%
streptomyces griseus ATCC 13273	soil, Japan	30	60%
Streptomyces leeuwenhoekii C34	desert soil	35	69%
Streptomyces lincolnensis NRRL 2936	soil	29	59%
Streptomyces lunaelactis MM109	moonmilk speleothem	28	71%
streptomyces olivochromogenes DSM 40451	soil	39	72%
streptomyces scabiei 87.22	plant pathogen	34	59%
streptomyces tsukubensis AT3	forest soil, China	36	64%

Table S4. *M/z* values annotated by GNPS present in the extracts of the isolated bacteria but not in the medium blank.

m/z value	Adduct	Annotated by GNPS as	Present in extracts of
203.179	$[M + H - H_2O]^+$	Alismol (terpenoid)	*Streptomyces* sp. M19
415.123	$[M+H]^+$	Futalosine	*Streptomyces* sp. M10
546.488	$[M+H]^+$	Aminobacteriohopanetriol	*Streptomyces* sp. M10, *Streptomyces* sp. M19
549.281	$[M+H]^+$	Antimycin A1	*Streptomyces* sp. M10
535.266	$[M+H]^+$	Antimycin A2	*Streptomyces* sp. M10
401.253	$[M+H]^+$	Bonactin	*Streptomyces* sp. M10
383.243	$[M + H - H_2O]^+$		
499.281	$[M+H]^+$	Conglobatin	*Streptomyces* sp. M19
415.269	$[M+H]^+$	Homononactyl homononactate	*Streptomyces* sp. M10
454.293	$[M+H]^+$	1-palmitoyl-2-hydroxy-sn-glycero-3-phosphoethanolamine	*Streptomyces* sp. M10
768.489	$[M+NH_4]^+$	Monactin	*Streptomyces* sp. M10
237.148	$[M + H - H_2O]^+$	Gliocladic acid (terpenoid)	*Saccharopolyspora* sp. M46
203.118	$[M+H]^+$	Nb-acetyltryptamine	*Streptomyces* sp. M10, *Saccharopolyspora* sp. M46, *Oerskovia* sp. M15
416.265	$[M+NH_4]^+$	NCGC00381146-01!2-[5-[2-[2-[5-(2-oxopropyl)oxolan-2-yl]propanoyloxy] butyl]oxolan-2-yl]propanoic acid	*Streptomyces* sp. M10
373.274	$[M + H - H_2O]^+$	12-Ketodeoxycholic acid	*Oerskovia* sp. M15
402.301	$[M+H]^+$	(Z)-N-hexadec-9-enoyl-L-phenylalanine	*Streptomyces* sp. M10

Figure S1. Origin and visualization of the mammoth sample. A) the mammoth carcass discovered in August 2012 on Maly Lyakhovsky Island, Russia. B) extracted intestinal specimen C) faecal sample.

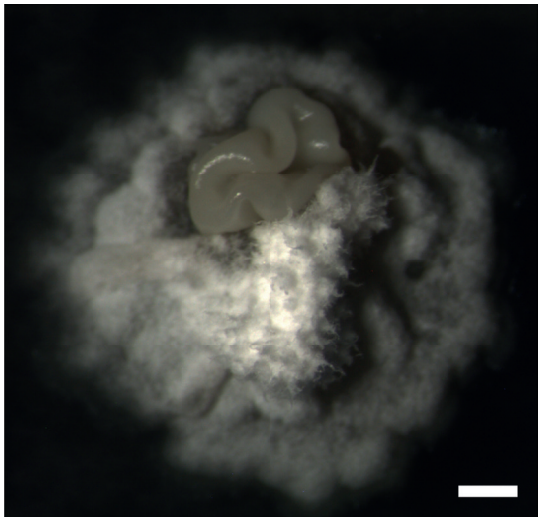


Figure S2. Different morphologies are observed within single colonies of isolate *Streptomyces* sp. M19.

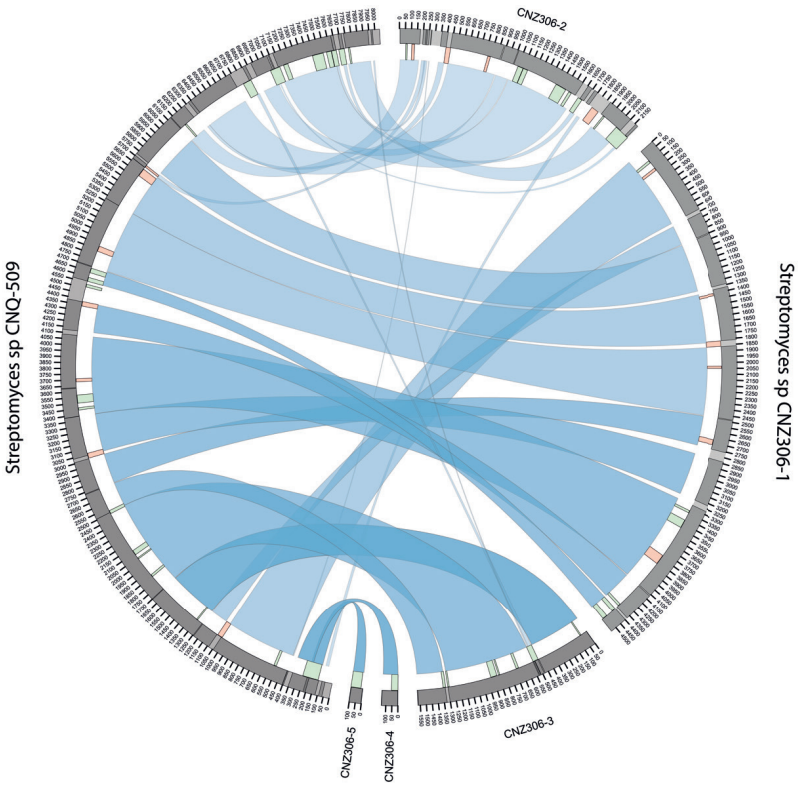


Figure S3. Synteny plot of *Streptomyces* sp. CNQ-509 and *Streptomyces* sp. CNZ306.

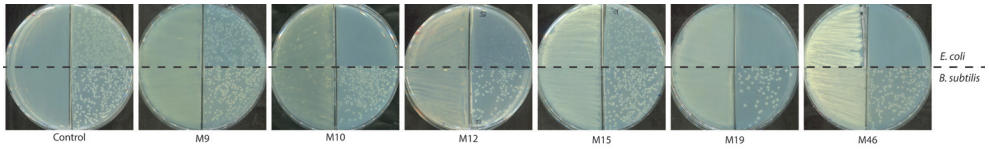


Figure S4. Bioactivity of volatiles released by the mammoth isolates against *E. coli* strain ASD19 and *B. subtilis* 168 after seven days of growth on NA.

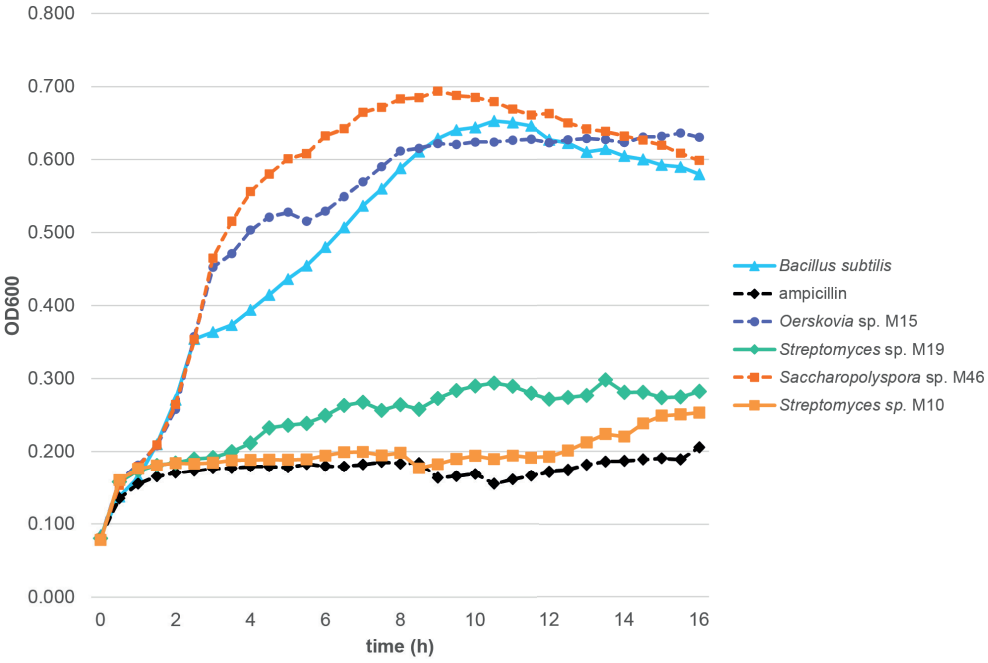


Figure S5. Antimicrobial activity of crude extracts against *B. subtilis*

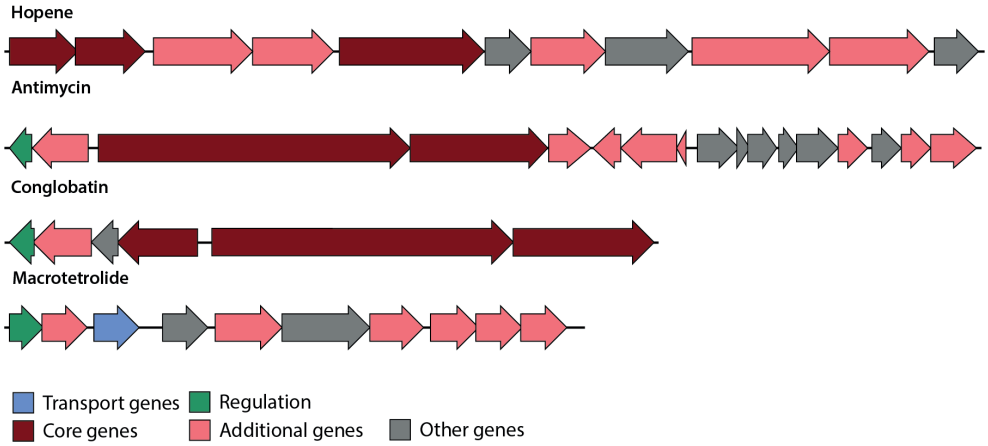
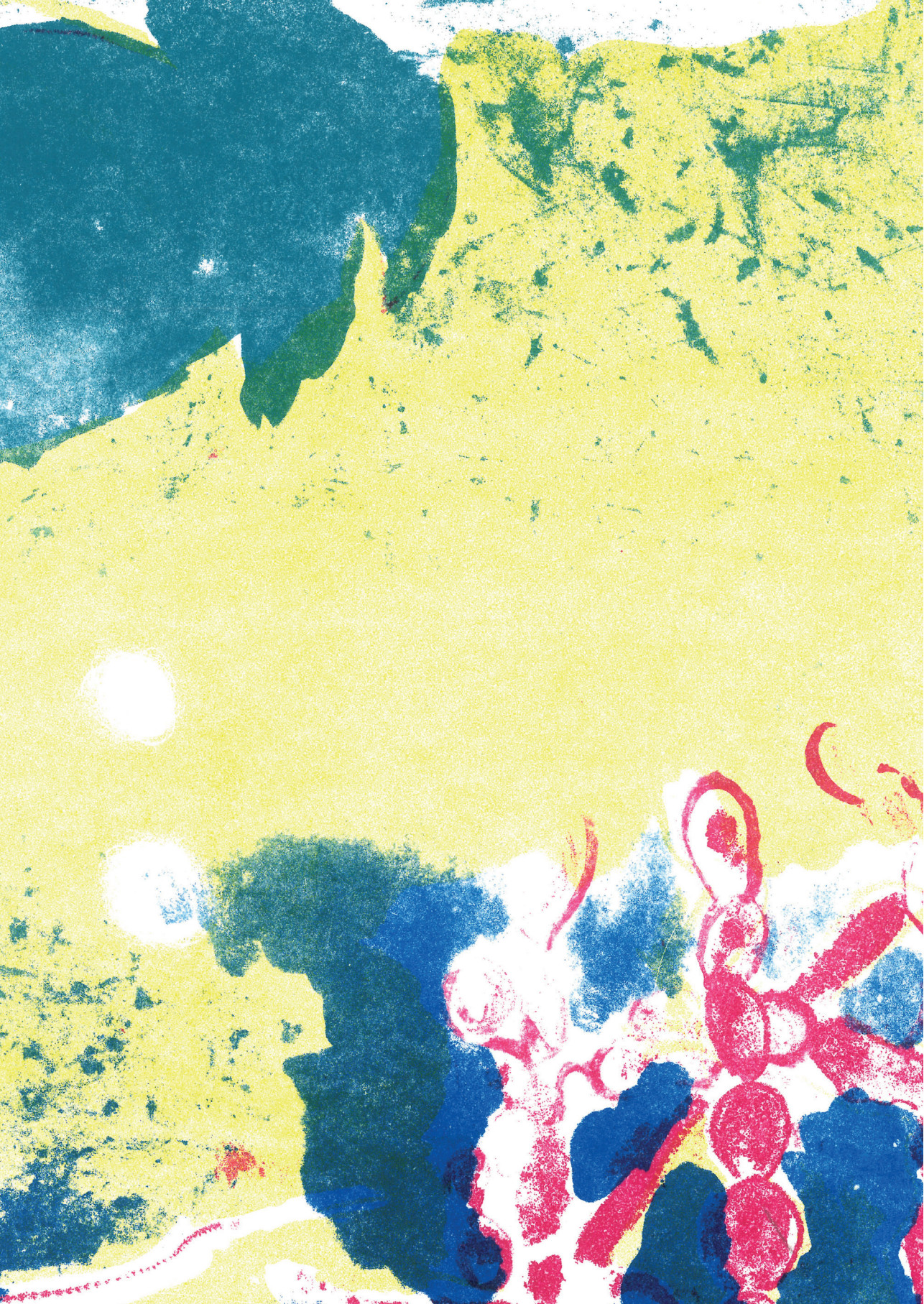


Figure S6. Predicted hopene cluster of M10 & M19, antimycin and macrotetrolide gene cluster of M10 and the conglobatin gene cluster of M19. The gene colours are based on the colouring system of antiSMASH v6.0.



4

Animal stress hormones enhance siderophore production in *Streptomyces*

Doris A. van Bergeijk, Anna M. Rozeboom, Du Chao, Melanie Ofman, Hugo Rietdijk, Herman P. Spaink, Gilles P. van Wezel

Part of the work presented in this chapter is published in:
van Bergeijk, D.A., Elsayed, S.S., Du, C., Nuñez Santiago, I., Roseboom, A.M., Zhang, L., Carrión V.J., Spaink, H.P., van Wezel, G.P. The ubiquitous catechol moiety elicits siderophore and angucycline production in *Streptomyces*. *Commun Chem* **5**, 14 (2022).

Abstract

Actinobacteria are a rich source of bioactive molecules, and genome sequencing has shown that the vast majority of their biosynthetic potential has yet to be explored. However, many of their biosynthetic gene clusters (BGCs) are poorly expressed in the laboratory, which prevents discovery of their cognate natural products. To exploit their full biosynthetic potential, better understanding of the signals that promote the expression of BGCs is needed. Here, we show that the human stress hormone epinephrine (adrenaline) elicits siderophore production by Actinobacteria. Catechol was established as the likely eliciting moiety, since similar responses were seen for catechol and for the catechol-containing molecules dopamine and catechin, but not for molecules with related chemical structures. The response could be linked to the iron-chelating properties of the catechol moiety. Quantitative proteomics revealed that the expression of proteins involved in iron uptake, siderophore production and dithiolopyrrolone biosynthesis is increased in response to catechol. Thus, we show that plant- and animal-associated molecules increase siderophore production in *Streptomyces*.

Introduction

The phylum Actinobacteria represents a highly diverse group of bacteria with extraordinary metabolic potential. Their specialised metabolites include most of the clinically used antibiotics along with numerous cancer chemotherapeutics, immunosuppressants, and pesticides, and are therefore of great importance for application as clinical drugs or in agriculture [9]. This metabolic versatility makes Actinobacteria attractive sources for drug discovery, for which there is an urgent need due to the global rise of drug resistance [53,54]. However, traditional high-throughput screening suffers from low return on investments due to dereplication, in other words, the rediscovery of bioactive compounds that have been identified before [239,240].

Next-generation sequencing technologies revealed a huge repository of previously unseen biosynthetic gene clusters (BGCs) in Actinobacteria, which showed that their potential as producers of bioactive molecules had been grossly underestimated [20,50,241]. However, these BGCs are often not expressed under laboratory conditions, most likely because the environmental cues that activate their expression in their original habitat are missing [2,28]. Indeed, molecules of actinobacterial origin continue to be discovered that have important new structural and/or functional features, for example the antifungal cyphomycin [43], the glycopeptide antibiotic corbomycin [170], and the angucycline-derived polyketide lugdunomycin [60]. To exploit Actinobacteria more efficiently as resource of chemical diversity, we need to first understand the triggers and cues that promote their expression. This knowledge can then be translated to eliciting approaches to activate the expression of BGCs and produce their cognate bioactive compounds at high throughput [42].

One approach to identify the cues that activate BGC expression lies in understanding the ecological context of specialised metabolite production [42]. Actinobacteria inhabit a wide range of terrestrial and aquatic ecosystems [2]. Within these environments, specialised metabolites play an important role in survival through mediation of resource competition [7,8], protection against oxidative stresses [3] and uptake of essential nutrients [5]. This requires careful timing of production and it is therefore likely that environmental signals indicative of specific stresses (such as nutrient levels and the presence of competitors) have been incorporated in the regulation of BGC expression [32-34]. Indeed, nutrient availability and co-culture of Actinobacteria with other microorganisms significantly influence their specialised metabolite production [35,36].

Actinobacteria live in and around a wide variety of other organisms, including higher eukaryotes such as plants, insects, marine organisms, and mammals [42,43]. As part of the microbiomes of these hosts, Actinobacteria are exposed to host-associated signaling molecules, many of which will likely influence their specialised metabolism [42]. Indeed, plant stress hormones, such as salicylic

acid and jasmonic acid, increase the antibiotic activity of endophytic streptomycetes [44]. These hormones are excreted by plants under pathogenic stress and their release might represent a 'cry for help' through which the plant may activate the production of bioactive substances by members of their microbiome in order to counteract a pathogenic attack [44]. Besides plant hormones, also animal stress hormones influence bacteria. For example, the human opioid dynorphin stimulates production of pyocyanin in *Pseudomonas aeruginosa* [242]. Specifically, catecholamines, which include the well-known 'fight or flight' hormone adrenaline (also known as epinephrine), influence bacterial growth [243,244], biofilm formation [245], and horizontal gene transfer [246]. We hypothesise that animal stress hormones may play a role in the control of antibiotic production of Actinobacteria.

In this work, we demonstrate that the animal stress hormone epinephrine can influence specialised metabolism of streptomycetes, specifically siderophore production. The catechol moiety was key to this response and this finding could be translated to other host-specific compounds (human and plant) containing a catechol moiety. Proteomics revealed increased expression of proteins involved in iron uptake, siderophore production and dithiolopyrrolone biosynthesis. Taken together, these results illustrate that catechol, by itself and as part of plant- and animal-associated molecules, can serve as elicitor of specialised metabolism.

Results

Epinephrine alters antibiotic production of Actinobacteria

Actinobacteria live in close association with higher eukaryotes, such as plants and animals. Host stress molecules might play a role in the regulation of actinobacterial specialised metabolism, perhaps reflecting a 'cry for help' from host to bacterium [44]. We therefore investigated if animal stress hormones can affect growth and metabolism of Actinobacteria. For this, we analysed the effect of epinephrine, also known as adrenaline, on a selection of our in-house actinobacterial strain collection that was previously shown to require particular growth conditions for the production of antibiotics [37]. We used the Gram-positive *Bacillus subtilis* 168 and the Gram-negative *Escherichia coli* ASD19 as indicator strains. Minimal Medium agar (MM) and Nutrient Agar (NA) plates with or without epinephrine (bitartrate salt) or tartaric acid (as control for the added bitartrate) were inoculated with spots from spore stocks of different actinobacterial strains. As a concentration we used 100 μM [244]. To see if the addition of epinephrine to the growth media affected the susceptibility of the indicator strains against antibiotics, we added a diffusion disc with ampicillin (6 μg) to each plate and tested whether the presence of epinephrine in the growth medium affected the size of the inhibition zone of ampicillin against *E. coli* and *B. subtilis*. No differences were observed, confirming that addition of epinephrine to the growth medium did not affect the susceptibility of the indicator strains.

On MM, only a small number of strains showed a change in bioactivity in response to epinephrine (Fig. S1A). Both promotion and inhibition of bioactivity were observed. For some strains, this decrease in bioactivity could be linked to inhibition of growth in presence of epinephrine. On NA, no significant changes in bioactivity were observed except for the elicitation of a large semi-transparent halo surrounding *Streptomyces* sp. MBT42 by epinephrine, indicative of reduced growth of *B. subtilis* (Fig. S1B).

The catechol moiety is key to the eliciting effect of epinephrine on *Streptomyces* sp. MBT42

To analyse the eliciting effect of epinephrine in more detail, we selected *Streptomyces* sp. MBT42, as this strain reproducibly showed a strong response to the hormone when the strain was grown on NA plates (Fig. 1A). In particular, we were interested to see whether epinephrine itself or any of its chemical constituents would be responsible for the eliciting effect. To do so, the effects of different catecholamines and structurally related compounds were assessed (Fig. 1B). Besides epinephrine, we also included catechin, dopamine, levodopa, and norepinephrine. Interestingly, all of these compounds had an effect similar to or stronger than epinephrine and significantly elicited *Streptomyces* sp. MBT42 compared to control (Fig. 1C). Since all of the tested compounds contain a catechol moiety, we wondered if catechol (1,2-dihydroxybenzene) may be the moiety primarily responsible for the observed eliciting effect. We therefore tested the effect of catechol itself, and as controls we used several structural analogues, namely the *meta* isomer resorcinol (1,3-dihydroxybenzene) and the *para* isomer hydroquinone (1,4-dihydroxybenzene), as well as molecules containing a monohydroxy-substituted benzene ring, such as phenol, phenylephrine, and tyramine. Importantly, of these compounds, catechol itself significantly enhanced antibiotic production by *Streptomyces* sp. MBT42 compared to control, while most other compounds failed to elicit antibiotic production (Fig. 1C). Hydroquinone had a mild eliciting effect, similar to that of norepinephrine. However, this effect was significantly lower compared to the effect of the compounds with a catechol moiety such as epinephrine and catechol. These data strongly suggest that the catechol moiety is key to the response of *Streptomyces* sp. MBT42 to the stress hormone epinephrine.

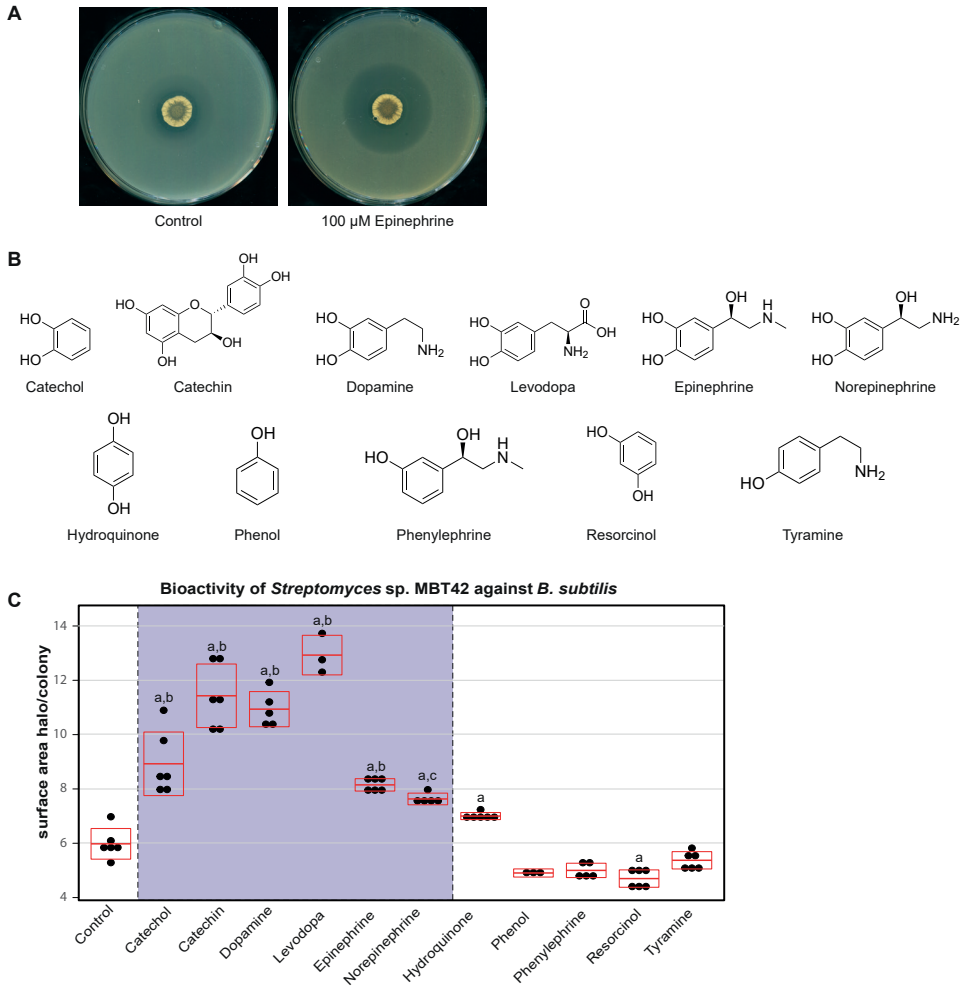


Figure 1. The catechol-moiety is key to the eliciting effect of epinephrine on *Streptomyces* sp. MBT42. A) *Streptomyces* sp. MBT42 was grown on NA supplemented with and without 100 μM epinephrine ($n = 3$). After 4 days of growth, plates were overlaid with *B. subtilis* to test for antimicrobial activity. Note the increased semi-transparent halo surrounding *Streptomyces* sp. MBT42 grown in presence of epinephrine. **B)** An overview of the structurally-related compounds tested to assess the specificity of the response **C)** Effect of different compounds on the bioactivity of *Streptomyces* sp. MBT42 against *B. subtilis*. Bioactivity was quantified by measuring the ratio between the surface area of the inhibition zone and the colony (data represent two independent experiments, the number of biologically independent replicates (= *n*) for each group is indicated in the figure, mean and standard deviation are indicated in red, the grey box highlights the response to compounds with a catechol moiety). One-way ANOVA, followed by a *post hoc* Tukey's honest significant difference (HSD) test, was performed to compare the difference in bioactivity between the growth conditions. The symbols indicate a significant ($p < 0.05$) increase in bioactivity compared to (**a**) the control, (**b**) to compounds lacking a catechol moiety, and (**c**) to compounds lacking a catechol moiety except hydroquinone.

Catechol enhances siderophore production in *Streptomyces* sp. MBT42

The catechol moiety has iron-chelating properties, and catechin is a strong siderophore [247]. We therefore hypothesised that the addition of catechol-containing molecules might result in lower iron availability, which may trigger siderophore production and further iron depletion, thus inhibiting growth of *B. subtilis*. To test this hypothesis, MBT42 spores were spotted onto NA with or without 100 μM catechol, and the plates overlaid with CAS agar solution after four days of growth. An orange halo was formed that is indicative of siderophore production, which was strongly increased in the presence of catechol (Fig. 2). The siderophore-induced halos matched the semi-transparent halos that were observed when *Streptomyces* sp. MBT42 was overlaid with *B. subtilis.*

Siderophore production is suppressed by iron. We therefore tested whether the addition of iron to the growth medium could compensate for the effect of catechol on *Streptomyces* sp. MBT42. Indeed, when 5 μM $FeCl_3$ was added to the medium, siderophore production was strongly reduced and this coincided with disappearance of the semi-transparent inhibition zones (Fig. 2). The same result was obtained when $FeSO_4$ was added (data not shown). Other metal ions did not have this effect (Fig. 2). Additionally, we tested whether the reduced growth of *B. subtilis* was related to iron depletion by adding a disc containing either 10 mM $FeCl_3$ or dH_2O on top of the *B. subtilis* overlay close to the spot of *Streptomyces* sp. MBT42. Within the semi-transparent halo, growth of *B. subtilis* was restored when $FeCl_3$ was present, but not in the control (Fig. S2). This strongly suggests that catechol reduces iron availability, which then enhances siderophore production by *Streptomyces* sp. MBT42 and leads to reduced growth of *B. subtilis.*

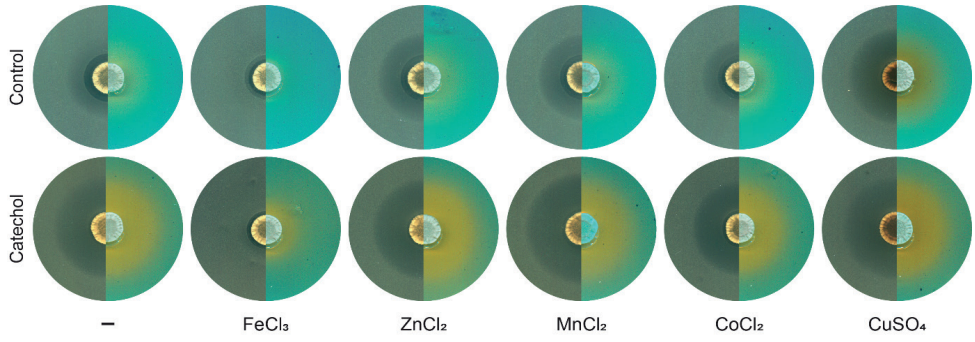


Figure 2. Catechol elicits siderophore production by *Streptomyces* sp. MBT42. MBT42 was grown on NA with and without 100 μM catechol, supplemented with various metal salts (5 μM) (n = 3). After 4 days of growth, plates were overlaid with *B. subtilis* (left) to test for antimicrobial activity or with CAS agar to detect the extracellular production of iron-chelating molecules (orange halos; right). Note that catechol inhibits growth of *B. subtilis* and induces siderophore production, and that these zones are highly comparable. When iron was added to the medium, siderophore production was almost completely inhibited and the semi-transparent halo was no longer visible.

Catechol enhances expression of proteins involved in dithiolopyrrolone and siderophore production

We next wondered which siderophore may be produced by *Streptomyces* sp. MBT42. For this, the full genome of *Streptomyces* sp. MBT42 was sequenced (GenBank accession number: JAJNOJ000000000) (Table S1) and the natural product BGCs predicted using antiSMASH 6.0 [217] (Table S2). This revealed four candidate BGCs for siderophores; one was identified as the desferrioxamine BGC, one with high similarity to the BGC for salinichelin and albachelin, and two predicted clusters that are not known to be associated with siderophore biosynthesis. To find out which siderophore BGC was induced by catechol compounds, MS-based quantitative proteomic analysis was applied, which is an efficient way to establish changes in expression of BGCs in response to eliciting signals [248,249]. The proteome samples of *Streptomyces* sp. MBT42 grown on NA plates with and without 100 μM dopamine, catechin, or phenylephrine were compared. Dopamine and catechin were chosen, because these compounds induced a similarly strong response despite their structural differences, while PE was added as a control, since this compound lacks a catechol moiety.

After five days of growth, biomass was harvested and snap-frozen in liquid nitrogen. Subsequent quantitative proteomic analysis was performed on four replicate samples per growth condition, yielding 1534 quantifiable proteins present in at least one treatment group. Of the 41 putative BGCs annotated by antiSMASH, proteins of 13 BGCs were expressed under these growth conditions (Table S2). To identify proteins involved in the response to catechin and dopamine, $\log_2$ fold changes in protein level were calculated compared to control cultures (no addition to growth medium). Proteins were considered significantly differentially expressed when a significant change was found in both dopamine- and catechin-cultures ($p \leq 0.05$) but not in phenylephrine-cultures ($p > 0.1$). According to these criteria, 13 proteins were significantly differentially expressed, three of which were predicted to be involved in NP biosynthesis; MBT42_4625 belongs to BGC2.5 that bears significant similarity to the salinichelin BGC, and MBT42_9695 and MBT42_9705 belong to BGC2.14A, which shows similarity to the BGCs for the dithiolopyrrolones holomycin and thiolutin (Fig. 3A). Expression of these proteins was significantly increased in dopamine- and catechin-grown cultures. Comparison of the total $\log_2$ fold changes of the proteins within both BGCs showed increased expression of siderophore BGC2.5 in both catechin and dopamine-grown cultures compared to phenylephrine-grown cultures ($p \leq 0.05$) (Fig. 3B+C, BGC border predictions are based on literature [109,250]). For the dithiolopyrrolone BGC2.14A, only part of the cluster showed increased expression in response to catechol-containing compounds.

Besides BGC-associated proteins, also the expression of an EfeM/EfeO family lipoprotein (MBT42_26995) involved in iron uptake was significantly increased in response to catechin and dopamine (Table 1). Taken together, this shows that catechol compounds act as an elicitor of siderophores, which is most likely explained by the sequestering of iron.

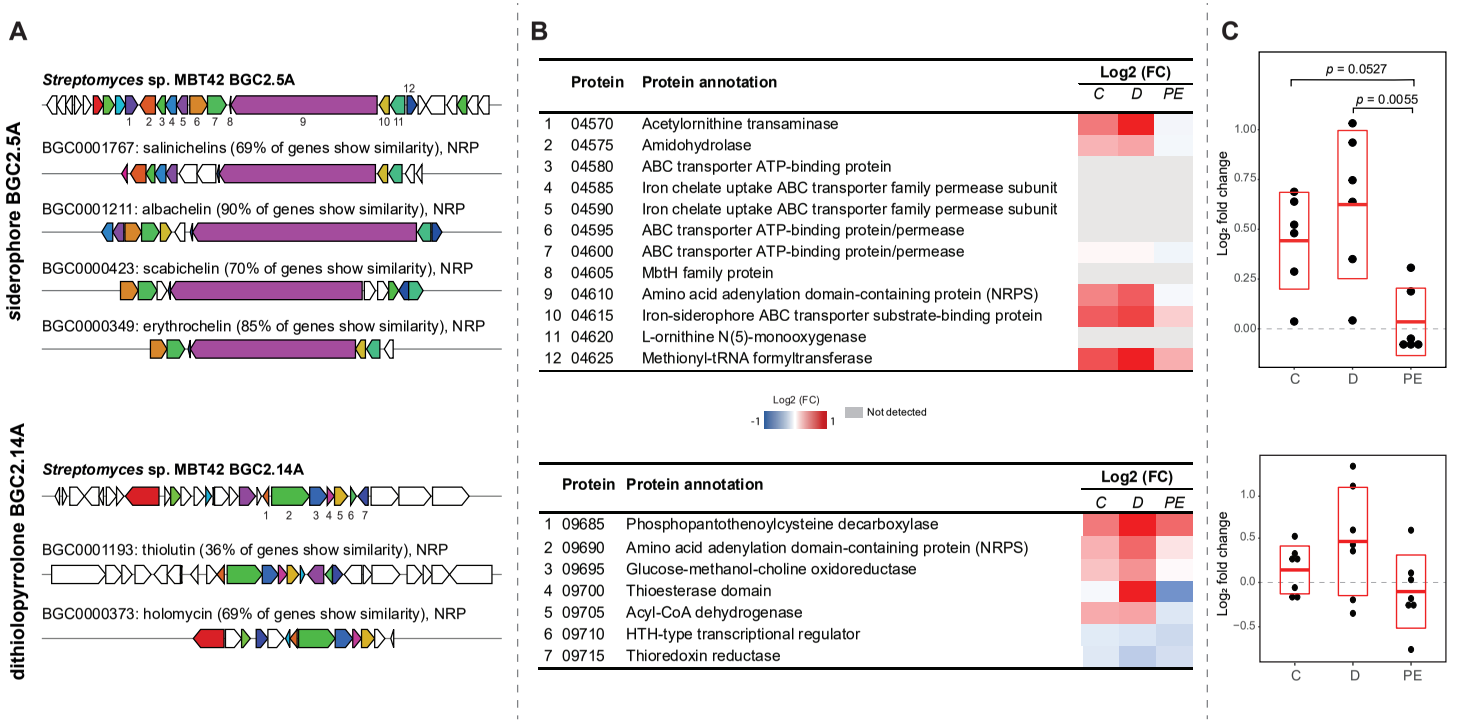


Figure 3. Eliciting effect of catechol compounds on BGCs related to siderophore (top) and dithiolopyrrolones (bottom) production. A) KnownClusterBlast output from antiSMASH which displays the similarity of BGC2.5A and BGC2.14A to clusters from the MIBiG database. Genes marked with the same colour are functionally related; white genes have no relationship. **B)** Changes in protein expression in response to catechin (C), dopamine (D), and phenylephrine (PE). Heatmaps showing log_2 fold changes in protein level compared to control ($n = 4$). Note the similarity between catechin and dopamine. C) Scatter bar plots showing the average log_2 fold change of each protein. The mean log_2 fold change in protein expression of the entire BGC and standard deviation are indicated in red ($n = 4$). The average log_2 fold changes (Log2 (FC)) in protein level were compared between the different groups by one-way ANOVA, followed by a *post hoc* Tukey's HSD test. Top figures represent siderophore BGC2.5A, figures below represent dithiolopyrrolone BGC2.14A.

Table 1. Proteins showing significantly altered expression in response to dopamine and catechin.

Protein ID	Protein annotation	Log2 (FC) *			BGC
		C	*D*	*PE*	
MBT42_04625	methionyl-tRNA formyltransferase	0.69	1.03	0.31	2.5
MBT42_06775	ATP-binding protein	-3.43	-3.43	0.03	-
MBT42_06860	ATP synthase subunit b	-0.53	-0.53	-0.35	-
MBT42_07290	sulfite oxidase	-0.35	-1.00	-0.25	-
MBT42_09695	GMC family oxidoreductase N-terminal	0.23	0.43	0.03	2.14A
MBT42_09705	Acyl-CoA dehydrogenase	0.33	0.36	-0.18	2.14A
MBT42_10280	Vitamin B12-dependent ribonucleotide reductase	-0.32	-0.32	-0.24	-
MBT42_16950	Phosphoserine transaminase	0.31	0.46	0.28	-
MBT42_21175	small secreted protein	-0.29	-0.35	-0.07	-
MBT42_22810	ABC-family ATP-binding cassette	-0.27	-0.36	-0.16	-
MBT42_26995	EfeM/EfeO family lipoprotein	0.25	0.36	-0.01	-
MBT42_28135	Anthranilate synthase component 1	-0.39	-4.27	-0.64	-
MBT42_33795	acyl-CoA/acyl-ACP dehydrogenase	-0.09	-0.12	-0.09	-

* For statistical analysis, Log_2 fold changes compared to control were calculated ($n = 4$). Average protein levels between control and each growth condition (catechin = C, dopamine = D, phenylephrine = PE) were compared using a two-sample t-test. Proteins that showed a significant change in protein level in catechin- and dopamine-cultures ($p < 0.05$) but not in PE-cultures ($p > 0.1$) are shown.

Discussion

Actinobacteria live in close association with a wide variety of eukaryotic hosts, including plants and animals. Plant stress hormones affect the bioactivity of *Streptomyces*, indicating that host-associated molecules can play a role in the regulation of BGC expression [2,44]. Here we show that the animal stress hormones adrenaline, dopamine, and noradrenaline and the plant-associated metabolite catechin increase siderophore production in *Streptomyces*, resulting in reduced growth of *B. subtilis*. We could link this response to the iron-chelating properties of the catechol moiety of these molecules.

Iron is an essential nutrient for bacterial growth and the cause of fierce battle within bacterial communities [5]. In cocultivation experiments, many of the observed changes in bioactivity and metabolome production have been attributed to changes in iron availability and siderophore production [35,83,95,251,252]. Here we present evidence of plant- and animal-associated hormones as important players in the battle for iron. Quantitative proteomics experiments revealed that the expression of proteins associated with iron uptake, siderophore production and dithiolopyrrolone biosynthesis is significantly increased in response to dopamine and catechin. To the best of our knowledge, elicitation of siderophore production by animal- and plant-associated molecules has not been reported in *Streptomyces* before.

Catecholamines increase the growth and virulence of bacteria, mainly by liberating iron from human transferrin, making it more accessible to bacteria [253-255]. They increase the expression of iron uptake systems in *Pseudomonas aeruginosa* and repress the expression of siderophore biosynthetic genes [256]. However, the data presented in our work show that catecholamines can also decrease iron availability, thereby resulting in increased siderophore production, with concomitant inhibition of the growth of *B. subtilis*. The same response was found in the presence of the plant metabolite catechin. Indeed, plant growth-promotion by bacteria has previously been attributed to their ability to produce siderophores thereby depriving other bacteria of iron [257]. This provides a novel perspective for catechol compounds as mediators of host-microbe interactions, and this interesting concept awaits further investigation in an *in vivo* situation.

To coordinate production of specialised metabolites, Actinobacteria have evolved a vast array of complex, multi-level regulatory pathways of which many remain to be elucidated [32]. We propose that the presence of catechol-containing molecules (catechols) results in lower iron availability, which triggers siderophore production. Our data indicate that catechols additionally increase the expression of proteins involved in dithiolopyrrolone biosynthesis, visualizing a possible regulatory connection between siderophore and dithiolopyrrolone biosynthesis. The concurrent increased expression of proteins involved in siderophore and dithiolopyrrolone biosynthesis in response to catechols may also point towards a shared ecological function of these natural products. While a lot of attention is directed towards the discovery of new bioactive molecules for therapeutic use, still relatively little is known about the ecological roles of specialised metabolites within the natural environment of Actinobacteria. Although the ecological purpose of siderophores is well known, little is known about the role of dithiolopyrrolones. Dithiolopyrrolones have antibacterial and anticancer activity, which has been correlated to their ability to inhibit RNA polymerase and to disrupt intracellular metal homeostasis [258,259]. In *E. coli*, mutants defective in iron and zinc were more sensitive to dithiolopyrrolones [258]. Hence, siderophore and dithiolopyrrolones production may be used by *Streptomyces* sp. MBT42 as a strategy to increase the sensitivity of competitors to dithiolopyrrolones. Further studies will be needed to address this hypothesis and to further elucidate the regulatory mechanisms involved.

Methods

Bacterial strains, growth conditions and antimicrobial activity assay

All media and routine *Streptomyces* techniques have been described previously [227]. The actinobacterial strain collection used in this study was obtained from the Leiden MBT strain collection [138]. *Bacillus subtilis* 168 and *Escherichia coli* ASD19 [237] were used as indicator strains for antimicrobial activity and were cultured in LB media at 37 °C.

Antimicrobial activity assays were conducted using the double-layer agar method. Strains were spotted on minimal medium agar plates (MM) supplemented with 0.5% mannitol and 1% glycerol (w/v) as non-repressing carbon sources, and nutrient agar (NA) (Difco) plates, using a pin replicator. For individual testing of strains, 2 μL spore stock was manually spotted. Growth media were supplemented with 25 mM TES buffer and 100 μM of either (+)-catechin hydrate (Sigma-Aldrich, CAS# 225937-10-0) , catechol (Sigma-Aldrich, CAS# 120-80-9), dopamine hydrochloride (Sigma-Aldrich, CAS# 62-31-7), (−)-epinephrine (+)-bitartrate salt (Sigma-Aldrich, CAS# 51-42-3), hydroquinone (Sigma-Aldrich, CAS# 123-31-9), levodopa (Sigma-Aldrich, CAS# 59-92-7), norepinephrine bitartrate monohydrate (MCE, CAS# 108341-18-0), phenol (VWR, CAS# 108-95-2), (R)-(−)-phenylephrine hydrochloride (Sigma-Aldrich, CAS# 61-76-7), tartaric acid (Sigma-Aldrich, CAS# 87-69-4), or tyramine hydrochloride (Sigma-Aldrich, CAS# 60-19-5).

After four days of incubation at 30 °C, plates were overlaid with soft LB agar (1.8% w/v agar) containing one of the indicator strains (2% v/v) pre-grown in liquid LB to exponential phase (OD600 = 0.4 – 0.6) and incubated overnight at 37 °C (± 18 hours). The following day, antibacterial activity was quantified by measuring the ratio between the surface area of the inhibition zone and the spot. All statistical analyses were performed in RStudio v1.4.1717.

Chrome azurol S (CAS) assay for siderophore detection

The medium for 1 L CAS was prepared according to the method of Schwyn and Neilands (1987) [260] without the addition of nutrients: 60.5 mg CAS, 72.9 mg hexadecyltrimethyl ammonium bromide (HDTMA-Br), 30.24 g piperazine-1,4-bis(2-ethanesulfonic acid) (PIPES) and 10 mL of 1 mM $FeCl_3 \cdot 6H_2O$ in 10 mM HCl. Agarose (0.9% w/v) was used as gelling agent. The CAS agar solution was overlaid onto the spots of *Streptomyces* sp. MBT42. Following incubation overnight at 30 °C (± 18 hours*)* the plates were examined visually for halos.

Genome sequencing, assembly, and annotation

Streptomyces sp. MBT42 was grown in 25 mL of YEME supplemented with 0.5% glycine and 5 mM $MgCl_2$ and cultivated at 30 °C with 200 rpm shaking speed. Genomic DNA was isolated by phenol-chloroform extraction as described previously 227. PacBio sequencing and assembly of Streptomyces sp. MBT42 was performed by Novogene (Novogene Europe, Cambridge, UK) using PacBio Sequel platform in continuous long reads mode. Raw sequences were demultiplexed with Lima v1.10.0, assembled using Flye v2.8.1 261, and polished using Arrow v2.3.3. The genome has been deposited at GenBank under accession numbers JAJNOJ000000000*.* Biosynthetic gene clusters were identified using the genome mining tool antiSMASH 6.0 [217].

Proteomics sample preparation

Streptomyces sp. MBT42 spores were spotted on NA plates supplemented with and without 100 μM catechin, dopamine, or phenylephrine (n = 4) and incubated for five days at 30 °C. Biomass was scraped off and snap-frozen in liquid nitrogen, lysed in a precooled TissueLyser adaptor (Qiagen, The Netherlands) and proteins extracted using lysis buffer [4% SDS, 100 mM tris-HCl (pH 7.6), 50 mM EDTA]. Total protein was precipitated using the chloroform-methanol method[262] and dissolved in 0.1% RapiGest SF surfactant (Waters, USA) at 95°C. The protein concentration was measured using the BCA method. Protein samples were reduced by adding 5 mM DTT and incubated at 60°C for 30 min, followed by thiol group protection with 21.6 mM iodoacetamide incubation at room temperature in the dark for 30 min. Then 0.1 µg trypsin (recombinant, proteomics grade, Roche) per 10 µg protein was added, and samples were digested at 37°C overnight. After digestion, trifluoroacetic acid was added to 0.5% and samples were incubated at 37°C for 30 min followed by centrifugation to degrade and remove RapiGest SF. Peptide solution containing 6 µg peptide was then cleaned and desalted using STAGE-Tips [263]. Briefly, 6 µg of peptide was loaded on a conditioned StageTip with 2 pieces of 1 mm diameter C18 disk (Empore, product number 2215), washed twice with 0.5% formic acid solution, and eluted with elution solution (80% acetonitrile, 0.5% formic acid). Acetonitrile was then evaporated in a SpeedVac. Final peptide concentration was adjusted to 40 ng·µL^{-1} using sample solution (3% acetonitrile, 0.5% formic acid) for analysis.

Proteomics measurement and data analysis

The desalted peptides solution was separated on an UltiMate 3000 RSLCnano system (Thermo Scientific) set in a trap-elute configuration, coupled to QExactive HF (Thermo Scientific) mass spectrometer. The LC system used a Waters nanoEase M/Z Symmetry C_{18} trap column (5 μm, 100 Å, 180 μm × 20 mm) for peptide loading/retention, and Waters nanoEase M/Z HSS T3 C_{18} analytical column (1.8 μm, 100 Å, 75 μm × 250 mm) for peptide separation. Mobile phase A was 0.1% formic acid (FA) in ULC-MS grade H_2O (Biosolve), while mobile phase B was 0.1% FA, 10% H_2O in ULC-MS grade ACN (Biosolve). The flow rate used was 0.3 μL/min. The gradient was programmed with linear increment from 1% to 5% B from 0 to 2 min, 5% to 13% from 2 to 63 min, 13% to 22% from 63 to 85 min, 22% to 40% from 85 to 104 min, 90% at 105 min and kept at 90% to 113 min.

The MS was operated in positive mode with data dependent acquisition. The survey scan for the tuning was set to scan range of 350–1,400 *m/z* at 60,000 resolution, 1 microscan, automatic gain control (AGC) of 1 × 10^6, max injection time (IT) of 50 ms, no sheath, aux or sweep gas, spray voltage ranging from 1.7 to 3.0 kV, capillary temp of 250 °C and an S-lens value of 80. The MS method settings were: the survey scan was taken with a scan range of 350–1,400 *m/z* at 120,000 resolution, AGC target of 3 × 10^6, and maximum IT time of 100 ms. For individual

peaks, a threshold of 1.00×10^3 for the minimum AGC target yielding an intensity of 2.0×10^4 was set to trigger an MS/MS event. No apex trigger was used, unassigned, 1 and charges >8 were excluded with peptide match mode preferred, and dynamic isotope exclusion was set to 20 s. For MS/MS events, the loop count was set to 10 and the settings were: resolution at 15,000, AGC target 1×10^5, max IT time 50 ms, isolation window of 1.6 *m/z*, fixed first mass of 120 *m/z* and normalised collision energy (NCE) at 28.

Raw LC-MS/MS files were analysed using MaxQuant software (v1.6.17.0) [264] with label free quantification (LFQ) method applied. Proteins that were not detected in three out of four replicates in at least one treatment group were removed prior to analysis, resulting in the removal of 132 protein quantification results. Log2 fold changes were calculated compared to control growth conditions (growth without supplementation). Proteins were considered significantly differentially expressed when a significant change was found in both dopamine- and catechin-cultures ($p \leq 0.05$) but not in PE-cultures ($p > 0.1$). The mass spectrometry proteomics data has been deposited to the ProteomeXchange Consortium via the PRIDE [265] partner repository with the dataset identifier PXD029669.

Supplementary information for Chapter 4

Table S1. Features of the genome of *Streptomyces* sp. MBT42

	***Streptomyces* sp. MBT42**
Number of contigs	2
Largest contig	8,925,615
Total length	8,967,886
N50	8,925,615
CDS	7,748
rRNAs	21
tRNAs	85

Table S2. Biosynthetic gene clusters identified by antiSMASH 6.0 [217]

Proto-cluster	Type	Most similar known cluster	Core biosynthetic protein(s)
2.1	lanthipeptide class IV	venezuelin (100%)	01555
2.2A	NRPS		02925, 02930, 02940
2.2B	NAPAA	stenothricin (13%)	03080, 03085
2.3A	T2PKS	spore pigment (83%)	03725, 03730
2.3B	NRPS		03765 - 03780
2.3C *	T1PKS	lobosamide A-C (17%)	03850 - 03865, 03875 - **03890**
2.4A	thiopeptide/LAP		04320 – 04340, 04360
2.4B *	terpene	2-methylisoborneol (100%)	**04395**
2.5A *	NRPS	salinichelin (69%)	**04610**
2.5B	NRPS	paenibactin (83%)	04685, 04690
2.6	melanin	melanin (28%)	05445
2.7 *	lantipeptide class III	SapB (100%)	**06250**
2.8 *	RiPP-like		**06385**
2.9 *	terpene	hopene (69%)	**06695**, 06710, 06715
2.10	RRE-containing/T2PKS	alnumycin A/B/C/P (75%)	07460, 07560, 07565
2.11 *	T1PKS/NRPS	formicamycins A-M	**07975**, 7980
2.12	butyrolactone	coelimycin P1 (16%)	08580
2.13	RiPP-like		09110
2.14A *	NRPS	thiolutin (36%)	**09690**
2.14B *	NRPS		**09770**
2.15	NRPS	friulimicin A-D (24%)	10020, 10025
2.16*	siderophore	ficellomycin (3%)	**10300**, 10310
2.17	siderophore	murayaquinone (6%)	10590, 10595
2.18	thiopeptide, LAP	BD-12 (14%)	11945-11965, 11990
2.19A	other	-	12130
2.19B	butyrolactone	A-factor(100%)	12175
2.2	melanin	istamycin (8%)	14320
2.21	linaridin	cypemycin (88%)	14495, 14505
2.22	NRPS-like	lankamycin (16%)	17300
2.23	siderophore	desferrioxamin B (100%)	24525
2.24	CDPS	malacidin A/B (5%)	27340
2.25 *	NAPAA	-	**33510**, 33525
2.26A	terpene	-	33630
2.26B	lantipeptide class II	-	33670
2.27 *	terpene	geosmin (100%)	**35570**
2.28	ectoine	ectoine (100%)	35825
2.29	NRPS/T1PKS	Herboxidiene (2%)	36730
2.30	CDPS	-	37735
2.31	terpene	-	38015
2.32 *	T1PKS/NRPS	elloramycin (8%)	**38455**, 38480
2.33	terpene	isorenieratene (100%)	38700, 38715

* BGCs were considered expressed if 1 ≥ core biosynthetic proteins were detected (expressed proteins in **bold**)

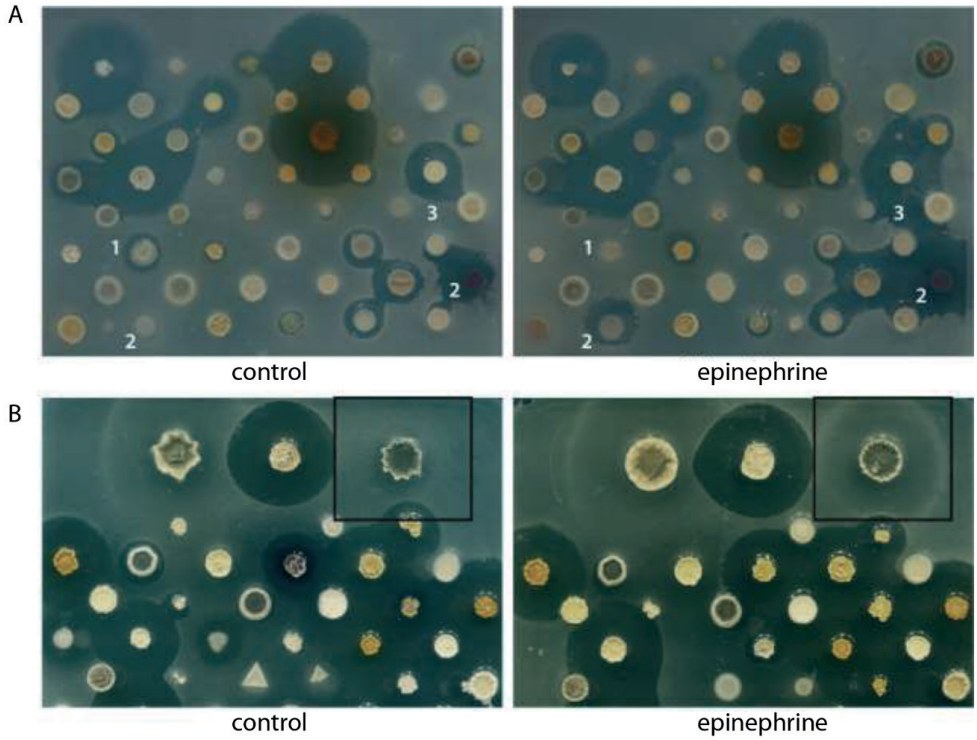


Figure S1. Epinephrine alters the bioactivity of different *Streptomyces*. A) On MM, both inhibition (1) and promotion (2) of antibiotic production in the presence of epinephrine is observed. Additionally, epinephrine can influence interactions between different strains illustrated by the change in halo shape (3) ($n = 3$). B) On NA, epinephrine elicits a semi-transparent halo surrounding *Streptomyces* sp. MBT42 (black square) ($n = 3$).

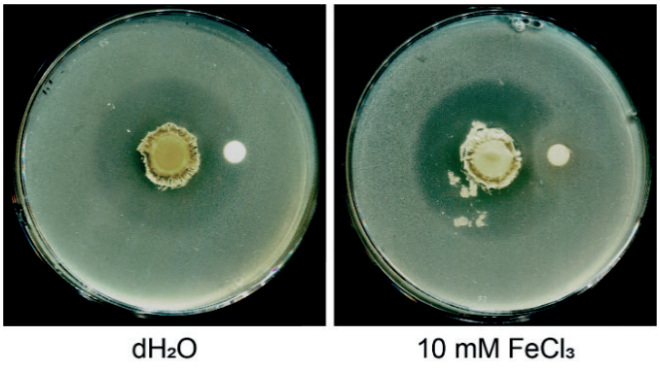
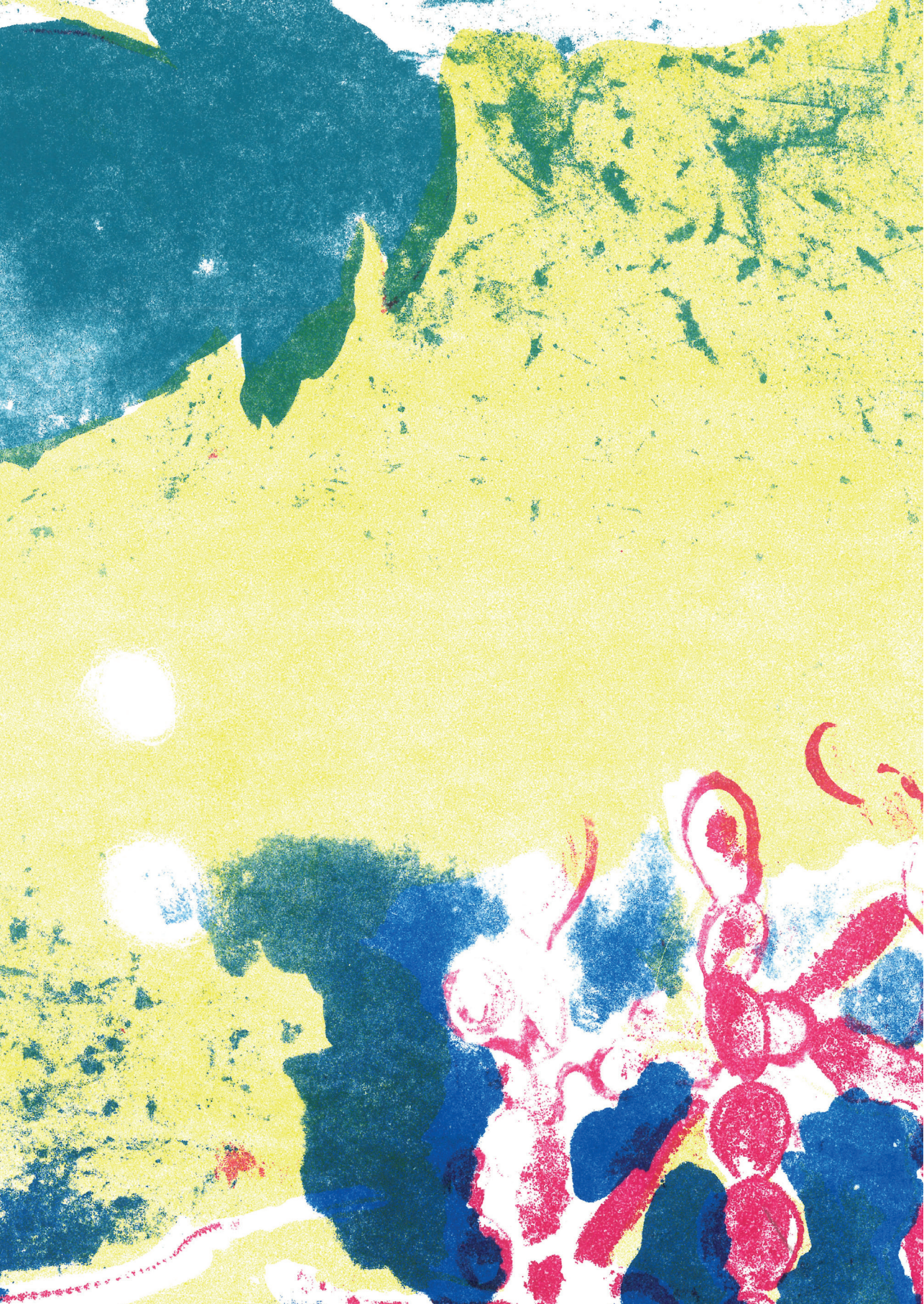
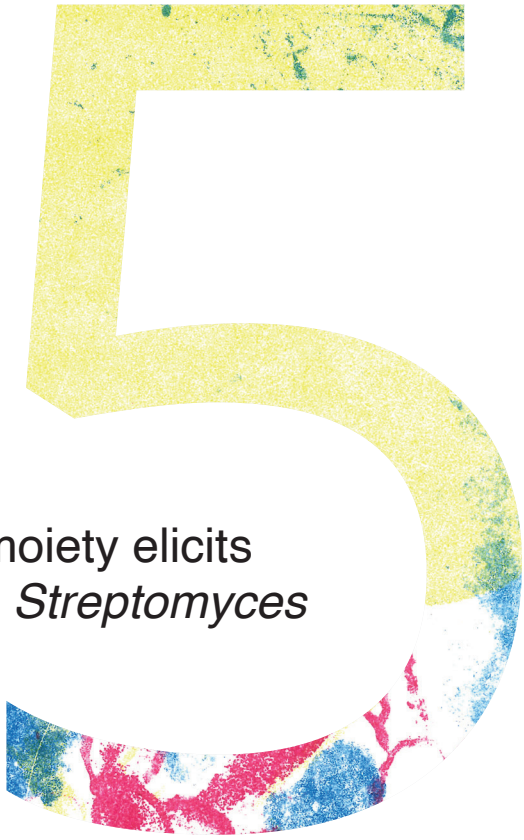


Figure S2. Iron restores growth of *B. subtilis*. *Streptomyces* sp. MBT42 was grown on NA for 4 days, followed by a soft agar overlay containing *B. subtilis*. On top of the overlay, a diffusion disc was added containing either 10 mM $FeCl_3$ or dH_2O followed by overnight incubation. Within the semi-transparent halo, growth of *B. subtilis* was restored when $FeCl_3$ was present, but not in the control.





The ubiquitous catechol moiety elicits angucycline production in *Streptomyces*

Doris A. van Bergeijk, Somayah S. Elsayed, Chao Du, Isabel Nuñez Santiago, Le Zhang, Victor J. Carrión, Herman P. Spaink and Gilles P. van Wezel

The work described in this chapter is part of the publication:
van Bergeijk, D.A., Elsayed, S.S., Du, C., Nuñez Santiago, I., Roseboom, A.M., Zhang, L., Carrión V.J., Spaink, H.P., van Wezel, G.P. The ubiquitous catechol moiety elicits siderophore and angucycline production in *Streptomyces*. *Commun Chem* **5**, 14 (2022).

Abstract

Actinobacteria are prolific producers of bioactive specialised metabolites. To unlock their full biosynthetic potential, better understanding of the regulatory networks and cognate signals that control antibiotic production is needed. Animal- and plant-associated molecules containing a catechol moiety, such as the human stress hormone adrenaline, enhance siderophore production in *Streptomyces* (Chapter 4). In this chapter, we analysed the response of Actinobacteria to catechol. Exploration of the catechol-responsive strain *Streptomyces* sp. MBT84 using mass spectral networking and proteomics revealed elicitation of a BGC that produces the angucycline glycosides aquayamycin, urdamycinone B, and galtamycin C. Heterologous expression of the catechol-cleaving enzymes catechol 1,2-dioxygenase or catechol 2,3 dioxygenase counteracted the eliciting effect of catechol. Taken together, these results illustrate that catechol, by itself and as part of plant- and animal-associated molecules, can act as elicitor of specialised metabolism.

Introduction

The global rise in drug resistance has urged the need for novel bioactive molecules [15,54]. Actinobacteria have been a prolific source of bioactive natural products with applications across medicine, agriculture, and biotechnology [10,11]. Over the past 20 years it has become clear that a huge amount of yet unexplored biosynthetic potential lies hidden in the genomes of Actinobacteria, which contain a wealth of uncharacterised biosynthetic gene clusters (BGCs) [20,27,50,241,266]. Many of these BGCs are silent, in other words, they are not (or hardly) expressed under routine laboratory growth conditions; this prevents discovery of their cognate natural products [29,93]. Heterologous expression and activation of these silent BGCs have shown that they encode functional biosynthetic pathways that produce novel molecules [29,30,267]. Hence, why are these BGCs silent and what activates their expression?

The metabolic versatility of Actinobacteria has been linked to their ecological diversity: Actinobacteria are widely distributed across Earth's ecosystems [2]. Within these environments, ecological forces have shaped how BGCs are regulated to ensure that the biosynthesis of their metabolically expensive products is only activated when needed [32-34]. Consequently, environmental signals have a major impact on the level and timing of specialised metabolism [37,44,138,252,268]. Identification of these eliciting signals and the regulatory pathways involved is needed to unlock the full biosynthetic potential of Actinobacteria.

In the previous chapter, we showed that the human stress hormone adrenaline enhances siderophore production by *Streptomyces* and this response was linked to the catechol moiety (Chapter 4). Interestingly, catechol is widely distributed across nature. It is common in plant metabolites [247,269] and is part of bacterial catecholate-type siderophores [5]. Additionally, catechol is present in the environment as degradation product of aromatic compounds such as polycyclic aromatic hydrocarbons originating from plant material or industrial pollution [270].

The specificity of the response of *Streptomyces* to the catechol moiety and its wide distribution in nature suggested catechol may be applied as a general elicitor of natural product biosynthesis in Actinobacteria. Here we show that catechol promotes the expression of a BGC that produces different (novel) angucycline glycosides. Expression of catechol-degrading enzymes counteracted the eliciting effect of catechol, highlighting the importance of the catechol moiety in the response. Taken together, our results provide further evidence for the potential of catechol as elicitor of specialised metabolism in *Streptomyces*.

Results

Catechol elicits the production of bioactive specialised metabolites

We have shown that plant- and animal-associated catechol compounds can impact the specialised metabolism of *Streptomyces* (Chapter 4). The specificity of this response to the catechol moiety suggested that catechol may be applied as a general elicitor of natural product biosynthesis in Actinobacteria. We therefore tested the effect of catechol on the bioactivity of some of the well-characterised Actinobacteria in our collection [138].

Similar to epinephrine, we observed both promotion and inhibition of bioactivity in the presence of catechol (Fig. S1). Catechol had a particularly strong effect on *Streptomyces* sp. MBT84, reproducibly enhancing its antibacterial activity against *B. subtilis* (Fig. S1, Fig. 1A). The bioactivity coincided with enhanced production of a yellow/brown pigment (Fig. 1A). When we investigated the specificity of this response by testing a library of structurally-related compounds including epinephrine, none of the other compounds increased antibiotic production (Fig. 1B). In Chapter 4, the effect of catechol on specialised metabolism was linked to iron availability. However, supplementation of the growth medium with iron did not affect the response of *Streptomyces* sp. MBT84 to catechol. Additionally, no siderophore production was observed in a CAS agar overlay. This strongly suggests that the change in bioactivity and pigment production was not due to altered siderophore production.

Catechol enhances angucycline production in *Streptomyces* sp. MBT84

To identify the nature of the metabolites produced by *Streptomyces* sp. MBT84 in response to catechol, the strain was streaked on MM agar plates and the metabolites were then extracted from the spent agar after five days of growth using ethyl acetate (EtOAc). The crude extracts of the catechol-grown cultures showed increased bioactivity as compared to those grown under control conditions, while media blanks showed no activity (data not shown). We wanted to rule out that catechol potentiated the bioactivity via an effect on the indicator bacteria rather than as elicitor. For this, we tested whether the presence of catechol in the growth medium affected the size of the inhibition zone of different antibiotics and the crude extracts of MBT84 against *B. subtilis*. No differences were observed (Fig. S2), which shows that addition of catechol to the growth medium does not affect the susceptibility of the indicator strains against antibiotics, nor does it potentiate the activity of the crude extracts.

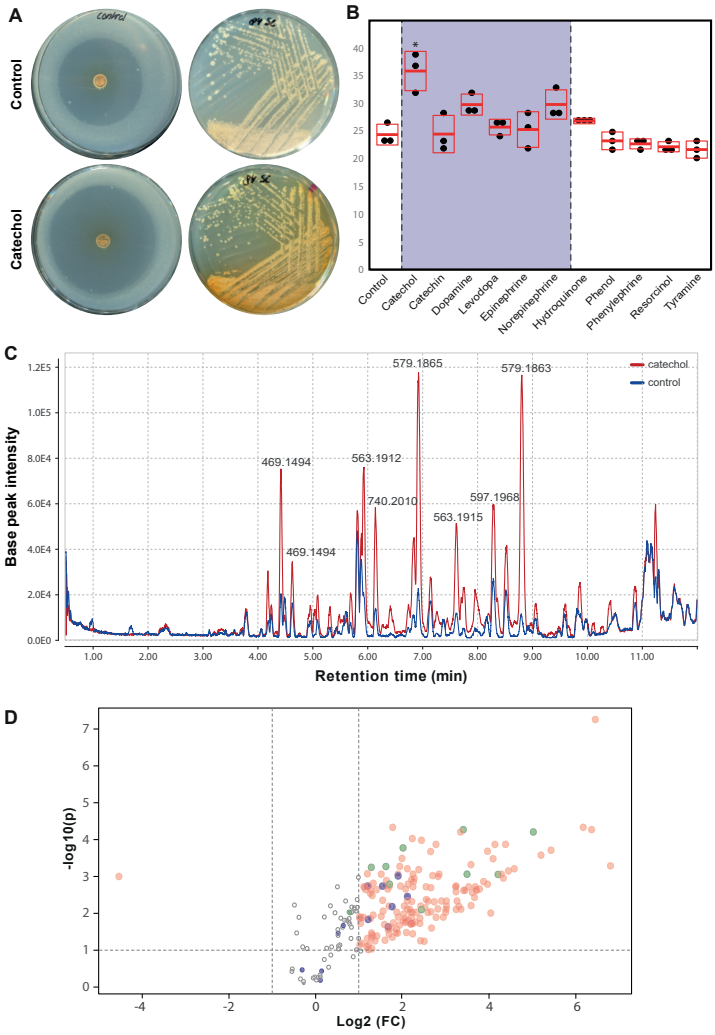


Figure 1. Catechol elicits the bioactivity and metabolite production of *Streptomyces* sp. MBT84. **A)** The antibacterial activity against *B. subtilis* 168 and yellow pigment production of *Streptomyces* sp. MBT84 are increased in the presence of catechol. **B)** None of the structurally related compounds, including epinephrine, increased the antibiotic production by *Streptomyces* sp. MBT84. One-way ANOVA, followed by a *post hoc* Tukey's HSD test, was performed to compare the difference in bioactivity between the growth conditions (* = $p < 0.001$, $n = 3$, mean and standard deviation are indicated in red, the grey box highlights the response to compounds with a catechol moiety). **C)** LC-MS chromatogram overlay of the crude extract of MBT84 grown with and without catechol. Multiple peaks were increased in intensity in the presence of catechol. **D)** Volcano plot highlighting the increased metabolite production by *Streptomyces* sp. MBT84 in response to catechol. The x and y axes represent the $\log_2$ fold changes and the corresponding $-\log_{10}$ FDR-adjusted *p*-value of all the mass features, respectively. Red circles represent the mass features in catechol-grown cultures with a significant intensity difference of more than 2-fold compared to control cultures (FDR-adjusted $p \leq 0.1$). Circles situated in the top left and top right quadrants represent the mass features which are most induced or repressed, respectively, by catechol with high statistical significance. The *m/z* values 469.149 and 579.186 are shown in purple and green, respectively ($n = 6$).

Liquid chromatography-mass spectrometry (LC-MS) analysis revealed multiple peaks in the chromatogram that were substantially overrepresented in the samples obtained from the catechol-grown cultures (Fig. 1C). The metabolomic profile of *Streptomyces* sp. MBT84 grown with and without catechol was compared through a volcano plot generated using MetaboAnalyst [271] (Fig. 1D). In the volcano plot, the majority of the mass features were significantly upregulated in catechol-grown cultures. The most notable ones were those with *m/z* values 469.1494 and 579.1864, which represented several highly intense mass features sharing the same exact masses but with different retention times. Additionally, multiple in-source fragments were among the mass features which were significantly upregulated with catechol. The mass features related to *m/z* 469.1494 and 579.1864 are likely due to yellowish-coloured metabolites, since they showed in the ultraviolet (UV) chromatogram corresponding peaks having UV absorption maxima of 420-440 nm. This corresponds well to the observed increase in yellow pigmentation when *Streptomyces* sp. MBT84 was grown in the presence of catechol.

To identify which *m/z* values correlated to the increased bioactivity of *Streptomyces* sp. MBT84, crude extracts were analysed using at-line nanofractionation coupled to LC-MS, using a method described previously [272]. In brief, the LC effluent is split post-column, after which one tenth of the flow goes to the MS, while the rest was fractionated into a 384-well plate (6 seconds/ well). Following LC solvent evaporation, the antibacterial activity of the collected nanofractions is assessed in a resazurin reduction assay and directly correlated to the mass spectral peaks. Reconstruction of the bioactivity chromatograms revealed multiple negative maxima, reflecting bioactivity. After alignment to the LC-MS chromatogram, the *m/z* values of the bioactive peaks were determined, which included those which correspond to the yellowish metabolites (*m/z* 469.149 and 579.186) (Fig. 2). In the extracted ion chromatogram of *m/z* 469.15, the multiple peaks observed were either due to quasi-molecular ions (like peaks number 7 and 9), or due to fragment ions of *m/z* 487.16 (like peaks number 1 and 2). The same pattern was observed for the peaks due to *m/z* 579.19, which were mostly fragment ions of *m/z* 597.20.

To gain more insight into which metabolites were induced by catechol and whether these were structurally related, molecular networking was employed using the Global Natural Products Social (GNPS) molecular networking web tool. This generates a network wherein molecules with related scaffolds cluster together [220]. Due to the presence of multiple isomeric metabolites, we used the classical molecular networking workflow while enabling the MSCluster algorithm [220,273]. This allows merging all mass features with similar MS and MS/MS spectra into one node, regardless of their different retention times. A network representing the ions detected in the crude extract of *Streptomyces* sp. MBT84 grown with and without catechol was constructed, revealing 258 nodes clustered in 21 spectral families (Fig. 3). Within the biggest spectral family,

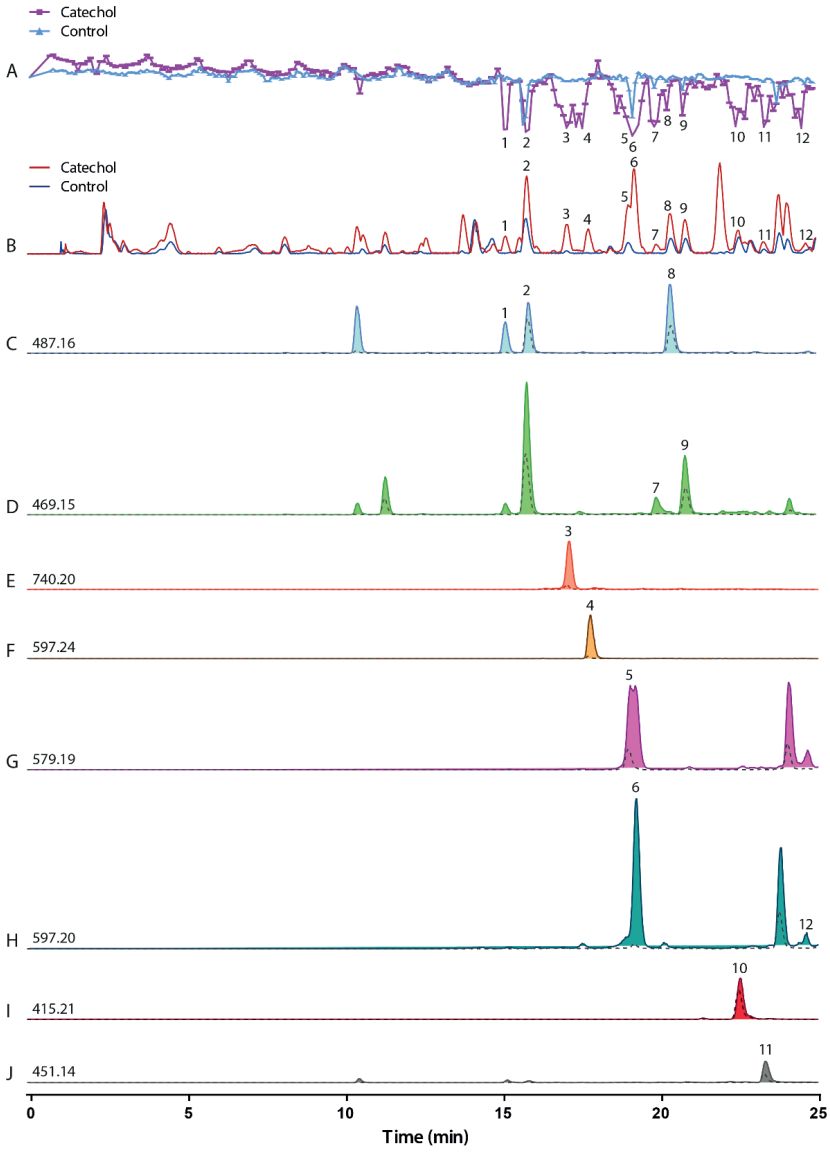


Figure 2. Nanofractionation reveals induction of several bioactive metabolites in *Streptomyces* sp. MBT84 by catechol. A) Bioactivity chromatogram overlay of the nanofractionated control and catechol-grown extracts. The negative peaks represent the bioactive compounds. The results of the resazurin reduction assay have been plotted against their corresponding LC-MS chromatograms after adjusting for the delay. **B)** Total ion current LC-MS chromatogram overlay of control and catechol-grown extracts. **C-J)** Extracted ion chromatogram overlay of m/z values 487.16 (C), 469.15 (D), 740.20 (E), 597.24 (F), 579.19 (G), 597.20 (H), 415.21 (I), and 451.14 (J) in both control (dashed line) and catechol-grown (solid line) extracts, which could be aligned to the negative maxima of the bioactivity chromatogram.

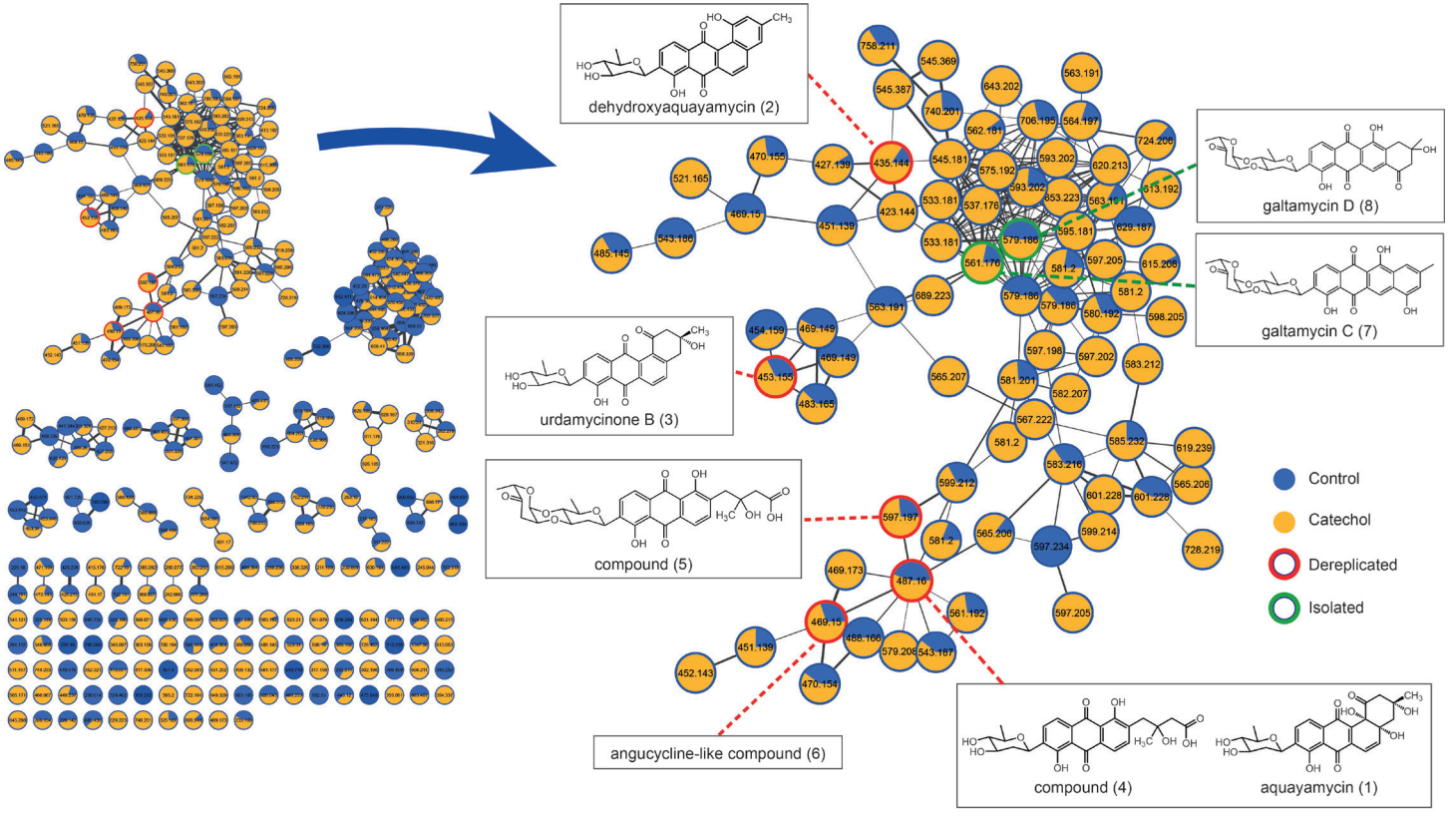


Figure 3. Molecular network of the ions detected in the crude extracts of *Streptomyces* sp. MBT84 revealing a large spectral family of angucycline compounds elicited by catechol. A pie chart was mapped to the nodes which represents the abundance of each m/z value in the control (blue) and catechol-grown crude extracts (yellow). Nodes highlighted in red represent dereplicated metabolites, while those highlighted in green represent the compounds isolated and identified in this study. *Streptomyces* sp. MBT84 was grown for five days on MM agar plates with or without 100 μM catechol (n = 6).

dereplication based on matching MS/MS spectra against the GNPS spectral library annotated several *m/z* values as being known metabolites belonging to the class of angucyclines, namely aquayamycin (**1**), dehydroxyaquayamycin (**2**), and urdamycinone B (**3**). Additionally, two structurally related compounds (**4,5**) as well as an angucycline-like compound (**6**), were annotated. This angucycline spectral family was more abundant in the presence of catechol and included the majority of the mass features that were correlated to bioactivity by nanofractionation (Fig. 2).

To ascertain that this network indeed consisted of angucycline glycoside-like compounds, we isolated the molecules with *m/z* 561.1760 (**7**) and 579.1866 (**8**). For this, *Streptomyces* sp. MBT84 was grown confluently on 12 x 12 mm MM agar plates supplemented with catechol and the metabolites were extracted with EtOAc. Following chromatographic isolation, **7** (red amorphous powder) and **8** (yellow amorphous powder) were obtained. The final structures of **7** and **8** were elucidated using nuclear magnetic resonance spectroscopy (NMR) (Supplemental data). Based on that, **7** was identified as galtamycin C, which has previously been isolated from an intertidal sediments-derived *Streptomyces* sp.[274]. As for **8**, it was found to be 1-oxo-3-hydroxy-3,4-dihydro-2*H*-galtamycin C based on the different 2D NMR correlations observed in comparison with **7**, together with the molecular formula and degrees of unsaturation obtained from its accurate mass (Supplemental data). To our knowledge **8** has not been previously described, and was thus designated as galtamycin D. Taken together, our data show that catechol enhances the biosynthesis of angucycline glycosides in *Streptomyces* sp. MBT84.

Identification of the angucycline glycoside BGC in *Streptomyces* sp. MBT84

To identify the gene cluster responsible for the biosynthesis of the angucycline glycosides in *Streptomyces* sp. MBT84, we obtained its full genome sequence using PacBio sequencing (GenBank accession number: JAHTGP000000000). A draft genome was assembled resulting in 3 contigs (Table S3). Analysis using antiSMASH 6.0 [217] revealed 21 putative BGCs (Table S4) of which BGC4 shows high similarity to the saquayamycin A BGC, and was therefore most likely responsible for the production of the angucycline glycosides (Fig. 4A, Table S5).

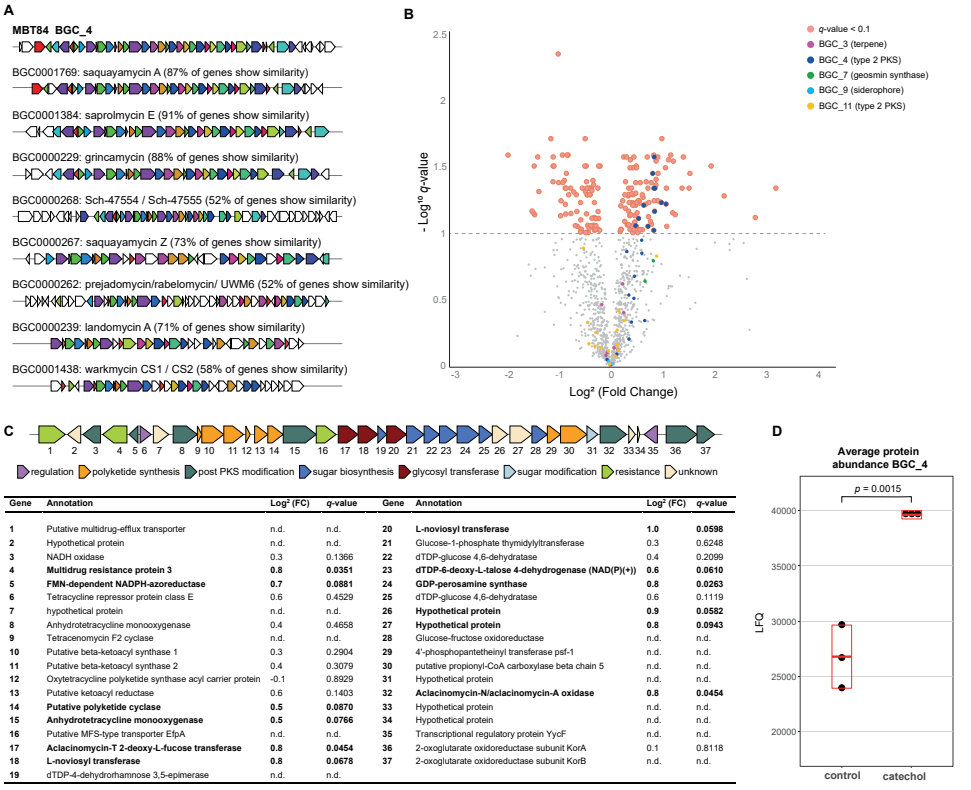
We have previously shown the applicability of quantitative proteomics combined with metabolomics to connect BGCs to the natural product(s) they specify [249]. Therefore, MS-based quantitative proteomic analysis was performed on total protein samples obtained from *Streptomyces* sp. MBT84 grown on MM agar plates with or without 100 μM catechol covered with a cellophane disc. After five days of growth, the biomass was harvested and snap-frozen in liquid nitrogen. Subsequent quantitative proteomics analysis was performed on three replicate samples per growth condition, yielding 1322 quantifiable proteins, of which 187 were significantly differentially expressed between catechol-grown and control cultures (FDR-

adjusted *p*-value < 0.1; Fig. 4B). Proteins belonging to five of the 21 putative BGCs annotated by the antiSMASH algorithm were expressed under the chosen conditions. However, only the biosynthetic proteins of BGC4 were mostly significantly upregulated in catechol-grown cultures, which provided validation of the involvement of this BGC in the observed change in bioactivity induced by catechol (Fig. 4B and C). The minimal polyketide synthase (PKS) enzymes responsible for the generation of the initial angucycline or angucyclinone backbone [275], namely two ketosynthase units (KSα and KSβ) and an acyl carrier protein, were not significantly differentially expressed (Fig. 4C).

To confirm that BGC4 was indeed responsible for the production of the angucycline glycosides, we knocked down transcription of the gene for the KSβ of the minimal PKS using CRISPR-dCas9 interference [276,277]. The knock-down was enforced from the CRISPRi construct pGWS1517 that expresses a spacer targeting the non-template strand of the gene for KSβ. Pigment production was strongly inhibited in the strain with reduced expression of KSβ, while no inhibition was seen in the control strain harbouring a construct (pGWS1516) that targets the template strand (Fig. S3). Metabolomics analysis revealed that the production of the angucycline glycosides was significantly reduced when the expression of KSβ was inhibited (Fig. S3). All the evidence together proves that indeed BGC4 specifies the angucyclines.

Expression of catechol-degrading enzymes counteracts antibiotic elicitation

Various microorganisms (e.g., *Pseudomonades*) enzymatically degrade aromatic compounds [270]. To establish whether catechol itself is essential for the eliciting effect, we expressed two enzymes in *Streptomyces* sp. MBT84 that degrade catechol. Catechol can be cleaved into *cis, cis*-muconic acid by catechol 1,2-dioxygenase (C12O) (*ortho*-pathway) or into 2-hydroxymuconic semialdehyde by catechol 2,3-dioxygenase (C23O) (*meta*-pathway) (Fig. S4) [270]. We wondered whether these degradation products would have a similar eliciting effect on angucycline production or whether the effect of catechol would disappear once degraded. To this end, we separately cloned the *catA1* gene encoding C12O from *Pseudomonas putida* KT2440 and the *xylE* gene encoding C23O from the promoter-probe vector pIJ4083 [278] onto the multi-copy vector pWHM3. The strong constitutive promoter SF14 [279] was used to drive expression. The resulting constructs pGWS1519 and pGWS1520 were introduced in *Streptomyces* sp. MBT84 via protoplast transformation, whereby the empty vector was used as a control. Of each recombinant strain (named MBT84-pWHM3, MBT84-C12O, MBT84-C23O) three independent transformants were tested for their ability to respond to catechol.



Gene	Annotation	Log^2 (FC)	*q*-value	Gene	Annotation	Log^2 (FC)	*q*-value
1	Putative multidrug-efflux transporter	n.d.	n.d.	**20**	**L-noviosyl transferase**	**1.0**	**0.0598**
2	Hypothetical protein	n.d.	n.d.	**21**	Glucose-1-phosphate thymidylyltransferase	0.3	0.6248
3	NADH oxidase	0.3	0.1366	**22**	dTDP-glucose 4,6-dehydratase	0.4	0.2099
4	**Multidrug resistance protein 3**	**0.8**	**0.0351**	**23**	**dTDP-6-deoxy-L-talose 4-dehydrogenase (NAD(P)(+))**	**0.6**	**0.0610**
5	**FMN-dependent NADPH-azoreductase**	**0.7**	**0.0881**	**24**	**GDP-perosamine synthase**	**0.8**	**0.0263**
6	Tetracycline repressor protein class E	0.6	0.4529	**25**	dTDP-glucose 4,6-dehydratase	0.6	0.1119
7	hypothetical protein	n.d.	n.d.	**26**	**Hypothetical protein**	**0.9**	**0.0582**
8	Anhydrotetracycline monooxygenase	0.4	0.4658	**27**	**Hypothetical protein**	**0.8**	**0.0943**
9	Tetracenomycin F2 cyclase	n.d.	n.d.	**28**	Glucose-fructose oxidoreductase	n.d.	n.d.
10	Putative beta-ketoacyl synthase 1	0.3	0.2904	**29**	4'-phosphopantetheinyl transferase psf-1	n.d.	n.d.
11	Putative beta-ketoacyl synthase 2	0.4	0.3079	**30**	putative propionyl-CoA carboxylase beta chain 5	n.d.	n.d.
12	Oxytetracycline polyketide synthase acyl carrier protein	-0.1	0.8929	**31**	Hypothetical protein	n.d.	n.d.
13	Putative ketoacyl reductase	0.6	0.1403	**32**	**Aclacinomycin-N/aclacinomycin-A oxidase**	**0.8**	**0.0454**
14	**Putative polyketide cyclase**	**0.5**	**0.0870**	**33**	Hypothetical protein	n.d.	n.d.
15	**Anhydrotetracycline monooxygenase**	**0.5**	**0.0766**	**34**	Hypothetical protein	n.d.	n.d.
16	Putative MFS-type transporter EfpA	n.d.	n.d.	**35**	Transcriptional regulatory protein YycF	n.d.	n.d.
17	**Aclacinomycin-T 2-deoxy-L-fucose transferase**	**0.8**	**0.0454**	**36**	2-oxoglutarate oxidoreductase subunit KorA	0.1	0.8118
18	**L-noviosyl transferase**	**0.8**	**0.0678**	**37**	2-oxoglutarate oxidoreductase subunit KorB	n.d.	n.d.
19	dTDP-4-dehydrorhamnose 3,5-epimerase	n.d.	n.d.				

Figure 4. Identification of the BGC responsible for the production of angucycline glycosides in *Streptomyces* sp. MBT84 using antiSMASH and MS-based quantitative proteomics. A) KnownClusterBlast output from antiSMASH which shows similar clusters from the MIBiG database. Genes marked with the same colour are interrelated; white genes have no relationship. **B)** Volcano plot of MS-based quantitative proteomics for cultures grown with and without catechol ($n = 3$). Proteins with an FDR-adjusted p-value ≥ 0.1 are greyed out. Proteins with a positive $\log_2$ fold change are higher expressed in catechol-grown cultures. **C)** BGC4 coding for the biosynthesis of angucycline glycosides. Annotations are based on BLAST homology searches and genes are colour-coded based on putative function. Significantly differentially expressed proteins are depicted in bold (FDR-adjusted $p < 0.1$, two-sample t-test, $n = 3$) **D)** Average protein level of BGC4 of cultures grown with and without catechol ($n = 3$). Mean and standard deviation are indicated in red. Two sample t-test was done showing BGC4 is significantly differentially expressed in presence of catechol.

The transformants were grown for five days on MM agar plates with and without 100 μM catechol, and the metabolites were extracted as described above. While catechol induced a clear increase in yellow pigmentation in MBT84 harbouring the control plasmid, such induction was no longer visible when MBT84-C12O was grown in the presence of catechol (Fig. 5A). In strain MBT84-C23O, the effect of catechol on the yellow pigmentation was strongly reduced, but in some transformants still some induction was observed. The visual

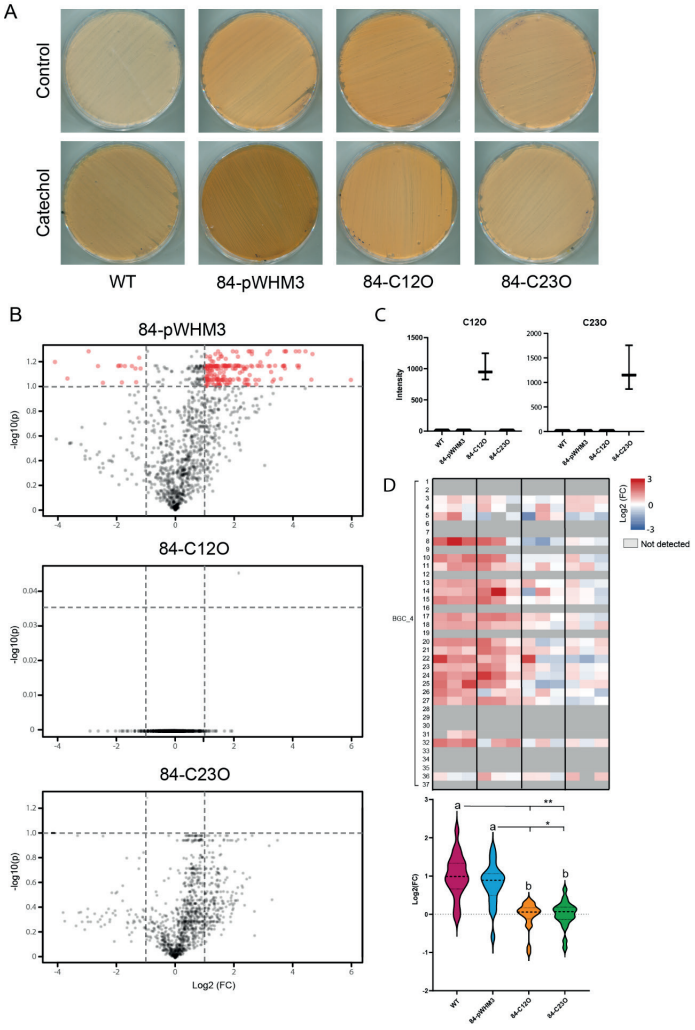


Figure 5. Heterologous expression of the catechol cleaving enzymes C12O and C23O resulted in a strong reduction of the eliciting effect of catechol. A) After five days of growth on MM with and without 100 μM catechol a clear increase in yellow pigmentation is visible in the WT and MBT84-pWHM3 (empty plasmid control) strain. This increase is not visible when C12O and C23O are expressed. **B)** Volcano plot highlighting the changes in metabolite production of the different recombinant strains in response to catechol. While no difference in metabolite profile was found when C12O is expressed, the metabolite profile of MBT84-C23O follows a similar distribution as the control but none of the mass features are significantly upregulated in the catechol-grown cultures (fold change >2 and FDR-adjusted *p*-value < 0.1) (n = 3). **C)** Protein abundance of the catechol dioxygenases in the WT and recombinant strains (n = 3) **D)** Expression profile of BGC4 coding for the biosynthesis of angucycline glycosides in the WT and recombinant strains. The heatmap shows the $\log_2$ fold change of protein level of each transformant (or WT replicate) comparing samples from catechol-grown cultures and control cultures (n = 3). The violin plots show the distribution of the $\log_2$ fold change of BGC4 expression in response to catechol (median = dashed line, quartiles = dotted line). Note that the empty vector control follows the same pattern as the WT strain, while this pattern is not visible when C12O and C23O are expressed. The average $\log_2$ fold changes in protein level of BGC4 in response to catechol were compared by one-way ANOVA, followed by a *post hoc* Tukey's HSD test. Similar letters indicate no significant difference (* = $p < 0.05$, ** $p < 0.01$).

assessment of antibiotic production was confirmed by detailed metabolomic analysis (Fig. 5B). Indeed, while catechol again elicited the production of metabolites in the control strain, catechol did not alter the metabolic profile of transformants expressing the catechol-degrading enzyme C12O or C23O. However, for the latter, the volcano plot showed a similar skewing to the upregulated side as the control, which suggests a reduced response.

To analyse the expression of the catechol dioxygenases and of BGC1.4 in the recombinant strains, we used quantitative proteomic analysis as described previously [216]. C12O and C23O were expressed individually in the respective recombinant strains (Fig. 5C). To study the protein levels of the gene products of BGC4 of the transformants or their parent with control plasmid, a heatmap of the $\log_2$ fold changes in expression between catechol and control was generated. Additionally, violin plots were generated to show the distribution of the expression differences of BGC1.4 proteins (Fig. 5D). The small sample size and variability between the different transformants, did not allow for a thorough statistical analysis. However, while catechol induced a similar upregulation of the expression of BGC1.4 in MBT84-pWHM3 compared to WT, this pattern was no longer visible in MBT84-C12O and MBT84-C23O. These results indicate that catechol is the elicitor and not one of its degradation products.

Discussion

Actinobacteria harbour many BGCs that are not or poorly expressed under laboratory growth conditions. A better understanding of the environmental signals that influence BGC expression as well as the regulatory pathways involved, will allow us to further explore actinobacterial specialised metabolism for therapeutic uses. Our work reveals catechol as elicitor of angucycline production in *Streptomyces* sp. MBT84. Angucyclines constitute the largest group of aromatic polyketides and are well known for their structural diversity and their therapeutic potential as anticancer and antibiotic compounds [275]. Our recent discovery of lugdunomycin and a range of novel angucyclines revealed that this extensively explored family of polyketides still has potential for drug discovery [60]. Indeed, the new compound described in this study, designated as galtamycin D, is another example that many angucyclines and derivatives are still to be discovered. Environmental cues may not only elicit the expression of (cryptic) biosynthetic pathways, but also of previously unseen branches of known pathways. Indeed, co-cultivation of the model streptomycete *Streptomyces coelicolor* A3(2) with *Aspergillus niger* elicited the production of GTRI-02, which is produced by the actinorhodin biosynthetic pathway [89]. The type II polyketide actinorhodin has been studied extensively by many laboratories for over 60 years, and the surprising discovery of an entirely new branch of its biosynthetic pathway underlines the importance of microbial interactions for drug discovery. Similarly, supplementation of

catechol to growth conditions might offer a new strategy to further explore the immense chemical space of angucyclines and other polyketides.

To coordinate appropriate production of specialised metabolites, Actinobacteria have evolved a vast array of complex, multi-level regulatory pathways of which many remain to be elucidated. The identification of a signal that influences specialised metabolism can provide a first step towards the elucidation of a pathway, as was illustrated by the discovery that GlcNAc had a stimulating effect on antibiotic production of many *Streptomyces* sp.[36]. Elucidating the signal transduction pathway identified DasR as pleiotropic transcriptional repressor of several pathway-specific activators, including ActII-ORF4 and RedD, whereby GlcN-6P acts as ligand for DasR [36]. However, the mechanism via which catechol impacts angucycline production in *Streptomyces*, remains yet unexplained. Our experiments show that degradation of catechol will annihilate its effect. Heterologous expression of the catechol-cleaving enzymes C12O or C23O in *Streptomyces* sp. MBT84 strongly reduced its response to catechol. This strongly suggests that catechol is the elicitor and not one of its degradation products. Of the two enzymes, C23O had slightly less impact than C12O, whereby the latter fully counteracted the effect of catechol. Our proteomics experiments show that both enzymes are expressed well, and we anticipate that the difference may lie in the nature of their degradation products. C23O is responsible for the extradiol ring cleavage of catechol, producing 2-hydroxymuconic semialdehyde in which the 1,2 di-hydroxy architecture of catechol is preserved, while C12O does not maintain this scaffold. This could indicate that specific protein interactions are involved in the response to catechol and might also explain why *Streptomyces* sp. MBT84 only responds to catechol. Catecholamines are protonated at neutral pH, which could interfere with protein binding. The next step will be to elucidate the signal transduction pathway for catechol and identify the key regulatory network it targets inside the cell.

Catechol impacts different types of specialised metabolites, as exemplified by its effect on the biosynthesis of angucyclines and siderophores (Chapter 4). Preliminary experiments in our laboratory revealed many elicited mass features in the metabolome of other *Streptomyces* species grown in presence of catechol (not shown). Our data indicate that these mass features belong to other classes of specialised metabolites, such as non-ribosomal peptides, which we are currently investigating. Interestingly, high throughput elicitor screening (HiTES) technology identified the plant-derived piceatannol as elicitor of the expression of a BGC for a cryptic NRPS[280]. Considering the data presented in our work, we hypothesise that this response might also be mediated through the catechol moiety that is present in piceatannol.

Our work again provides a new example of the power of using chemical elicitors to activate cryptic BGCs and identify the corresponding metabolites. Crude extracts are complex mixtures of metabolites and identifying the bioactive metabolites is often time consuming. Additionally, many 'cryptic' BGCs have not yet been linked to a natural product. We propose catechol as a new addition to the catalogue of elicitors that could be used in screens for novel bioactive molecules, which may allow for prioritization of bioactive metabolites for isolation from complex crude extracts and assist in 'decrypting' cryptic BGCs. A major issue to solve in the future is to identify the signal transduction pathway by which streptomycetes sense catechol, from internalization through metabolism to the activation of BGCs. Understanding of the mechanisms involved is of great interest as they may offer new tools for natural product-based drug discovery.

Methods

Bacterial strains, growth conditions and antimicrobial activity assay

All media and routine *Streptomyces* techniques have been described previously[227]. The bacterial strain collection used in this study were obtained from the Leiden MBT strain collection. *Streptomyces sp.* MBT84 has been isolated from soil samples collected from the QinLing Mountains (Shanxi province, China) [138]. *Bacillus subtilis* 168 and *Escherichia coli* ASD19 [237] were used as indicator strains for antimicrobial activity and were cultured in LB media at 37 °C.

Antimicrobial activity assays were conducted using the double-layer agar method. Strains were spotted on minimal medium agar plates (MM) supplemented with 0.5% mannitol and 1% glycerol (w/v) as non-repressing carbon sources, and nutrient agar (NA) (Difco) plates, using a pin replicator. For individual testing of strains, 2 μL spore stock was manually spotted. Growth media were supplemented with 25 mM TES buffer and 100 μM of either (+)-catechin hydrate (Sigma-Aldrich, CAS# 225937-10-0) , catechol (Sigma-Aldrich, CAS# 120-80-9), dopamine hydrochloride (Sigma-Aldrich, CAS# 62-31-7), (−)-epinephrine (+)-bitartrate salt (Sigma-Aldrich, CAS# 51-42-3), hydroquinone (Sigma-Aldrich, CAS# 123-31-9), levodopa (Sigma-Aldrich, CAS# 59-92-7), norepinephrine bitartrate monohydrate (MCE, CAS# 108341-18-0), phenol (VWR, CAS# 108-95-2), (R)-(−)-phenylephrine hydrochloride (Sigma-Aldrich, CAS# 61-76-7), tartaric acid (Sigma-Aldrich, CAS# 87-69-4), or tyramine hydrochloride (Sigma-Aldrich, CAS# 60-19-5).

After four days of incubation at 30 °C, plates were overlaid with soft LB agar (1.8% w/v agar) containing one of the indicator strains (2% v/v) pre-grown in liquid LB to exponential phase (OD600 = 0.4 – 0.6) and incubated overnight at 37 °C (± 18 hours). The following day, antibacterial activity was determined as the ratio between the inhibition zone diameter and the spot diameter.

Bioactivity of crude extracts and metabolite profiling

Streptomyces sp. MBT84 was grown confluently on MM agar plates with and without 100 μM catechol (n = 6). After five days of growth, the agar plates were cut into small pieces, soaked overnight in ethyl acetate (EtOAc) to extract the metabolites, evaporated at room temperature, and dissolved in MeOH to a final concentration of 20 mg/mL. 10 μL was spotted onto a sterile filter disc placed onto a soft agar layer inoculated with *B. subtilis* 168. As controls, 10 μL MeOH and 6 μL of 1 mg/mL ampicillin solutions were used. Following incubation overnight at 37 °C (± 18 hours), the plates were examined visually for halos.

For liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis, the dry extracts were dissolved in MeOH to a final concentration of 1 mg/mL. LC-MS/MS acquisition was performed using Shimadzu Nexera X2 ultra high-performance liquid chromatography (UPLC) system, with attached photodiode array detector (PDA), coupled to Shimadzu 9030 QTOF mass spectrometer, equipped with a standard electrospray ionisation (ESI) source unit, in which a calibrant delivery system (CDS) is installed. A total of 2 μL was injected into a Waters Acquity HSS C_{18} column (1.8 µm, 100 Å, 2.1 × 100 mm). The column was maintained at 30 °C, and run at a flow rate of 0.5 mL/min, using 0.1% formic acid in H_2O, and 0.1% formic acid in acetonitrile (ACN) as solvents A and B, respectively. The gradient used was 5% B for 1 min, 5–85% B for 9 min, 85–100% B for 1 min, and 100% B for 4 min. The column was re-equilibrated to 5% B for 3 min before the next run was started. The PDA acquisition was performed in the range of 200-600 nm, at 4.2 Hz, with 1.2 nm slit width. The flow cell was maintained at 40 °C.

All the samples were analysed in positive polarity, using data dependent acquisition mode. In this regard, full scan MS spectra (m/z 100–1700, scan rate 10 Hz, ID enabled) were followed by two data dependent MS/MS spectra (m/z 100–1700, scan rate 10 Hz, ID disabled) for the two most intense ions per scan. The ions were fragmented using collision induced dissociation (CID) with fixed collision energy (CE 20 eV), and excluded for 1 s before being re-selected for fragmentation. The parameters used for the ESI source were: interface voltage 4 kV, interface temperature 300 °C, nebulizing gas flow 3 L/min, and drying gas flow 10 L/min.

LC-MS based comparative metabolomics

Raw data obtained from LC-MS analysis were converted to mzXML centroid files using Shimadzu LabSolutions Postrun Analysis. The files were imported into Mzmine 2.53 for data processing [281]. Unless stated otherwise, *m/z* tolerance was set to 0.002 *m/z* or 10.0 ppm, RT tolerance was set to 0.05 min, noise level was set to 2.0E2 and the minimum absolute intensity was set to 5.0E2. Raw data were cropped to RT 0.5–12 min. Mass ion peaks were detected (positive polarity, mass detector: centroid) and their chromatograms were built using

ADAP chromatogram builder[282] (minimum group size in number of scans: 10; group intensity threshold: 2.0E2). The detected peaks were smoothed (filter width: 9), and the chromatograms were deconvoluted (algorithm: local minimum search; chromatographic threshold: 85%; search minimum in RT range: 0.05; minimum relative height: 1%; minimum ratio of peak top/edge: 2; peak duration: 0.03–2.00 min). The detected peaks were deisotoped (monotonic shape; maximum charge: 2; representative isotope: most intense). Peak lists from different extracts were aligned (weight for RT = weight for *m/z* = 20; compare isotopic pattern with a minimum score of 50%). Missing peaks detected in at least one of the samples were filled with the gap filling algorithm (RT tolerance: 0.1 min). Among the peaks, we identified fragments (maximum fragment peak height: 50%), adducts ($[M+Na]^+$, $[M+K]^+$, $[M+NH_4]^+$, maximum relative adduct peak height: 3000%) and complexes (Ionization method: $[M+H]^+$, maximum complex height: 50%). Duplicate peaks were filtered. Artifacts caused by detector ringing were removed (*m/z* tolerance: 1.0 *m/z* or 1000.0 ppm). The aligned peaks were exported to a MetaboAnalyst file.

In Excel, features that were not consistently present with an intensity higher than 8000 in all samples were removed from the file. Additionally, all features that originate from the culture medium were removed by retaining only features with an average peak intensity of at least 50 times greater in the bacterial extracts than in the culture medium extracts. The resulting peak list was uploaded to MetaboAnalyst [271] for statistical analysis. Log transformation with pareto scaling was applied to the data. Differences with a twofold change and an FDR-adjusted *p*-value < 0.1 were considered statistically significant. Based on these criteria, volcano plots and heat maps were generated.

MS/MS-based molecular networking and dereplication

MS/MS raw data (obtained from Shimadzu 9030 QTOF MS) were converted to a 32-bit mzML file using MSConvert (ProteoWizard) and a molecular network was assembled using the online Global Natural Product Social Molecular Networking (GNPS) tool [220]. Both the precursor ion and the MS/MS fragment ion mass tolerance were set to 0.02 Da. The minimum cosine score was set to 0.7 and the minimum matched peaks set to 5. The MSCluster algorithm was run with a minimum cluster size of 2 spectra. The spectra in the network were searched against the GNPS spectral libraries. For this, the precursor ion and the MS/MS fragment ion mass tolerance were set to 0.5 Da. Matches between network spectra and library spectra required a minimum score of 0.7 and at least 6 matched peaks. Cytoscape 3.5.1 was used for visualization of the generated molecular networks. The edge thickness was set to represent the cosine score, with thicker lines indicating higher similarity between nodes. LC-MS/MS data were deposited in the MassIVE Public GNPS data set (MSV000087784). The molecular networking job in GNPS can be found at https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=78cfa392f9c94ac0a110dc682a2d8e6f.

The annotated MS/MS spectra were deposited in the GNPS spectral library for galtamycin C (CCMSLIB00006675753) and the new compound **8** (CCMSLIB00006675754).

At-line nanofractionation and bioactivity assessment

Liquid chromatography separation, subsequent at-line nanofractionation and parallel mass spectrometry analysis were performed as previously described [272]. Each sample was measured in duplicate. The extracts were dissolved in MeOH to a final concentration of 4 mg/mL, and 10 µL was injected into Waters Acquity UPLC system with attached PDA, equipped with XBridge Peptide BEH C_{18} column (5 μm, 300 Å, 4.6 × 100 mm). The flow rate used was 0.6 mL/min. Solvent A was 0.1% formic acid in H_2O and solvent B was 0.1% formic acid in ACN. The gradient used was 0–50% B for 20 min, 50–90% for 5 min, 90% for 5min, 90–100% for 1 min and 100% for 10 min. After the flow was split in a 1:9 ratio, the smaller fraction was fed to a Thermo Instruments MS system (LTQ Orbitrap XL, Bremen, Germany) equipped with an ESI source. The following ESI parameters were used: capillary voltage 5 V, spray voltage 3.5 kV, capillary temperature 300 °C, auxiliary gas flow rate 10 arbitrary units, and sheath gas flow rate 50 arbitrary units. MS spectra were acquired in the Orbitrap in positive mode at a mass range of 100-2000 *m/z*, and FT resolution of 30,000. The larger fraction eluted from the UPLC was sent to the chip-based nano-electrospray ionization source/fractionation robot (NanoMate Triversa, Advion BioSciences). The delay was determined as 0.3 min by injection 1 μL of 0.5 mg/mL ampicillin dissolved in water. Dried plates with the nanofractionated extract were tested for antibacterial activity in a resazurin reduction assay against *B. subtilis* 168 following the method described by Mladic et al. (2018) [272]. The bioactivity chromatograms were plotted in GraphPad Prism 9 software.

Large-scale fermentation and isolation of metabolites 7 and 8

Streptomyces sp. MBT84 was grown on MM agar plates supplemented with 50 μM catechol, 1% glycerol and 0.5% mannitol at 30 °C for five days. Agar plates were cut into small pieces and soaked in EtOAc to extract metabolites as described earlier. The solvent was subsequently evaporated under reduced pressure at 40 °C to obtain 1.6 g crude extract. This extract was adsorbed onto 1.6 g silica gel (pore size 60 Å, 70-230 mesh, Sigma Aldrich), and loaded on a silica column, followed by gradient elution using mixtures of n-hexane, EtOAc, and MeOH. One of the fractions that eluted with 50% EtOAc : 50% n-hexane was combined with the fraction that eluted with 75% EtOAc : 25% n-hexane, and reconstituted in acetonitrile. This fraction was further purified using a Waters preparative HPLC system comprised of 1525 pump, 2707 autosampler, and 2998 PDA detector. The pooled fraction was injected into a SunFire C_{18} column (10 μm, 100 Å, 19 × 150 mm). The column was run at a flow rate of 12.0 mL/min, using solvent A (dH_2O) and solvent B (acetonitrile), and a gradient of 70-100% B over 20 min to yield **7** (3.4 mg) and **8** (1.9 mg).

Galtamycin D (8): yellow amorphous powder; UV (LC-MS) λmax 222, 266, and 441 nm; HRESIMS *m/z* 579.1866 $[M+H]^+$ (calcd for $C_{31}H_{31}O_{11}$, 579.1861); ^1H and ^{13}C NMR data (Table S5).

NMR measurements

NMR measurements for the purified compounds were recorded on Bruker Ascend 850 NMR spectrometer (Bruker BioSpin GmbH), equipped with a 5 mm cryoprobe. The samples were measured in DMSO-d_6 in a 3 mm NMR tube through the use of an adapter. The spectra were referenced using the solvent residual peak and processed in MestReNova software.

Genome sequencing, assembly, and annotation

MBT84 was grown in 25 mL of YEME supplemented with 0.5% glycine and 5 mM $MgCl_2$ and cultivated at 30 °*C* with 200 rpm shaking speed. Genomic DNA was isolated by phenol-chloroform extraction as described previously[227]. Genome sequencing was performed using Pacbio Sequel RSII at DNA link Sequencing Lab – South Korea. Raw sequences were demultiplexed with Lima 1.9.0 to produce CCS reads which were converted to fastq with bam2fastq 1.3.0. The assembly was performed with Flye 2.5 [261]. The genome has been deposited at GenBank under accession number JAHTGP000000000. Biosynthetic gene clusters were identified using the genome mining tool antiSMASH 6.0 [217].

Proteomics sample preparation

MM agar plates with and without 100 μM catechol were covered with cellophane. MBT84 spores were spread on the plates using glass beads and incubated for five days at 30 °*C*. Biomass was scraped off and snap-frozen in liquid nitrogen, lysed in a precooled TissueLyser adaptor (Qiagen, The Netherlands) and proteins extracted using lysis buffer [4% SDS, 100 mM tris-HCl (pH 7.6), 50 mM EDTA]. Protein extracts were prepared as described in Chapter 4

Proteomics of catechol-grown cultures (proteomining)

For the analysis of the expression of biosynthetic enzymes in response to catechol, we applied natural product proteomining [249]. Samples were prepared from *Streptomyces* surface-grown cultures grown with or without catechol. Desalted peptide solutions were injected into Waters nanoAcquity UPLC system equipped with a Waters HSS T3 C_{18} (1.8 µm, 100 Å, 75 µm X 250 mm). A gradient from 1% to 40% acetonitrile in water (with added 0.1% FA) over 110 min was applied. Online MS/MS analysis was done using a Waters Synapt G2-Si HDMS mass spectrometer with a $UDMS^E$ method set up as described previously [283]. [Glu1]-fibrinopeptide B was used as a lock mass compound and sampled every 30 s. Raw data from all samples were first analysed using the vender software ProteinLynx

Global SERVER (PLGS, version 3.0.3, waters, USA). The resulting dataset was imported in ISOQuant version 1.8 [283] for label-free quantification. TOP3 quantification was filtered to remove identifications that had an average value of less than 4000, in all replicates. This led to the removal of 152 protein quantification results. Log_2 fold changes were calculated, and proteins were considered significantly altered in expression when FDR-adjusted $p < 0.1$ were obtained. Volcano plots were made from filtered data, with the biosynthetic gene clusters color-coded. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [265] partner repository with the dataset identifier PXD030319.

Proteomics of *Streptomyces* sp. MBT84 expressing catechol dioxygenases

Heterologous expression of catechol dioxygenases

Constructs for the heterologous expression of catechol 1,2-dioxygenase (C12O) and catechol 2,3-dioxygenase (C23O) were constructed as follows: The *catA1* gene encoding C12O was amplified from the genomic DNA of *Pseudomonas putida* KT2440 using the primer pair SF14_catA1_F and catA1_T0_R. The *xylE* gene encoding C23O was amplified from promoter-probe vector pIJ4083 using the primer pair SF14_xylE_F and xylE_T0_R. The forward primers contain the sequence of the strong constitutive promoter SF14[279] including RBS and the reverse primers contain a t0 terminator sequence. The PCR products SF14-*catA1*-T0 and SF14-*xylE*-T0 were placed into the BamHI / XbaI site of the multi-copy vector pWHM3 to create vectors pGWS1519 and pGWS1520 respectively. These constructs were introduced into *Streptomyces sp.* MBT84 via protoplast transformation, whereby the empty vector was used as a control. An overview of the constructs and oligonucleotides is presented in Table S1 and Table S2. For all experiments, the growth medium was supplemented with 10 μg/mL thiostrepton.

Analysis of the recombinant strains by quantitative proteomics

Quantitative proteomics was used to analyse the expression of catechol dioxygenases and the expression profile of BGC1.4 in recombinant strains of *Streptomyces* sp. MBT84 as described previously [216]. Briefly, the desalted peptide solution was separated on an UltiMate 3000 RSLCnano system (Thermo Scientific) set in a trap-elute configuration, coupled to QExactive HF (Thermo Scientific) mass spectrometer. The LC system used a Waters nanoEase M/Z Symmetry C_{18} trap column (5 μm, 100 Å, 180 μm × 20 mm) for peptide loading/retention, and Waters nanoEase M/Z HSS T3 C_{18} analytical column (1.8 μm, 100 Å, 75 μm × 250 mm) for peptide separation. The MS was operated in positive mode with data dependent acquisition. Raw LC-MS/MS files were analysed using MaxQuant software (v1.6.17.0)[264] with label

free quantification (LFQ) method applied. Proteins were considered significantly altered in expression when FDR-adjusted $p < 0.1$ were obtained. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [265] partner repository with the dataset identifier PXD030484.

CRISPRi technology

To knock down the expression of genes of the angucycline BGC, we applied CRISPRi RNA interference technology. The CRISPRi system was modified from pCRISPR-dCas9 [284] by expressing Cas9 from the constitutive *gapdh* promoter, using vector pSET152 that integrates at the øC31 attachment site on the *S. coelicolor* chromosome [285]. For this, a 20 nt spacer sequence was introduced into the sgRNA scaffold by PCR using forward primers KSβ_TF or KSβ_NT2F together with reverse primer SgTermi_R_B. The PCR products were cloned into pGWS1370 [286] via NcoI / BamHI restriction sites. The resulting constructs pGWS1516 (targeting template strand of KSβ, control) and pGWS1517 (targeting non-template strand of KSβ) were then introduced into *Streptomyces sp.* MBT84 via protoplast transformation[227]. An overview of the constructs and oligonucleotides is presented in Table S1 and S2.

Supplementary information for Chapter 5

Table S1. Overview of the primers used in this study

Primer name	sequence
SgTermi_R_B	CTAGGGATCCCAAAAAACCCCTCAAGACCCGTTTAGAGGCCCCAAGGGGTTATGCTAGTTACGCCTACGTAAAAAAAGCACCGACTCGGTGCC
KSβ_TF	CATGCCATGGGTTCGAGATGGGTGTGATCAGTTTTAGAGCTAGAAATAGC
KSβ_NT2F	CATGCCATGGGATGTCGCGGGGCAGTGTCCGTTTTAGAGCTAGAAATAGC
SF14_catA1_F	CATGGGATCCTAATGAGTTACGTAGACCTACGCCTTGACCTTGATGAGGCGGCGTGAGCTACAATCAATACTCGATTAGAATTCAAGGGAGAGAACATGACCGTGAAAATTTCCCAC
catA1_T0_R	CATGTCTAGATGGACTCACAAAGAAAAAACGCCCGGTGTGCAAGACCGAGCGTTCTGAACAATCAGCCCTCCTGCAACGCCCG
SF14_xylE_F	CATGGGATCCTAATGAGTTACGTAGACCTACGCCTTGACCTTGATGAGGCGGCGTGAGCTACAATCAATACTCGATTAGAATTCAAGGGAGAGAACATGAACAAAGGTGTAATGCGA
xylE_T0_R	CATGTCTAGATGGACTCACAAAGAAAAAACGCCCGGTGTGCAAGACCGAGCGTTCTGAACAATCAGGTCAGCACGGTCATGAA

Table S2. Overview of the plasmids used in this study

Plasmid	Description	Reference
pWHM3	Unstable *E. coli* / *Streptomyces* shuttle vector with high copy number	[287]
GWS1370	pGWS1369 (pSET152 lacking its NcoI site) containing sgRNA scaffold (no spacer) and Pgapdh-dCas9	[286]
GWS1516	GWS1370 containing a spacer targeting the template strand of KSβ	This work
GWS1517	GWS1370 containing a spacer targeting the non-template strand of KSβ	This work
GWS1519	pWHM3 containing *catA1* behind SF14 promoter and a t0 terminator	This work
GWS1520	pWHM3 containing *xylE* behind SF14 promoter and a t0 terminator	This work

Table S3. Genomic features of *Streptomyces* sp. MBT84

	***Streptomyces* sp. MBT84**
Number of contigs	3
Largest contig	9,591,886
Total length	10,344,466
N50	9,591,886
CDS	9,713
rRNAs	18
tRNAs	90

Table S4. Biosynthetic gene clusters identified by antiSMASH 6.0 [217]

Protocluster	Type	Most similar known cluster
Contig_1		
1	RiPP-like	Informatipeptin (57%)
2	NAPAA	-
3 *	terpene	Hopene (92%)
4 *	T2PKS, oligosaccharide	Saquayamycin A (87%)
5	T3PKS	-
6	siderophore	Grincamycin (8%)
7 *	terpene	Geosmin (100%)
8	RiPP-like	-
9 *	siderophore	-
10	terpene	Albaflavenone (100%)
11 *	T2PKS	Spore pigment (83%)
12	siderophore	Desferrioxamine B (66%)
13	RiPP-like	-
14	ectoine	Ectoine (100%)
15	NAPAA	Chalcomycin (9%)
16	T3PKS	Herboxidiene (8%)
17	RRE-containing	
18	NRPS, betalactone	Cyclomarin D (17%)
19	Melanin	Melanin (42%)
20	NRPS, NAPAA, RiPP-like	Stenothricin (18%)
Contig_2		
21	Bacteriocin	-
Contig_3		
-		

* Expressed BGCs: ≥ 1 core biosynthetic protein was identified

Table S5. Comparison of BGC4 to saquayamycin cluster

	Protein locus tag	Size (aa)	Putative function	Homologue in saquayamycin BGC0001769	%identity*
1	_08725	534	Multidrug-efflux transporter	sqnZ	98.50
2	_08730	246	Hypothetical protein	sqnAA	61.02
3	_08735	372	NADH oxidase	Not found	-
4	_08740	526	Multidrug resistance protein	sqnB	99.24
5	_08745	199	FMN-dependent NADPH-azoreductase	sqnC	100.00
6	_08750	227	Tetracycline repressor protein class E	sqnD	96.38
7	_08755	283	Hypothetical protein	sqnE	97.88
8	_08760	492	Anhydrotetracycline monooxygenase	sqnF	98.98
9	_08765	602	Tetracenomycin F2 cyclase	sqnBB	99.07
10	_08770	426	polyketide putative beta-ketoacyl synthase 1	sqnH	99.28
11	_08775	408	polyketide putative beta-ketoacyl synthase 2	sqnI	98.77
12	_08780	89	polyketide synthase acyl carrier protein	sqnJ	98.88
13	_08785	261	Putative ketoacyl reductase	sqnK	100.00
14	_08790	311	Putative polyketide cyclase	sqnL	99.04
15	_08795	665	Anhydrotetracycline monooxygenase	sqnM	98.95
16	_08800	404	putative MFS-type transporter EfpA	sqnN	99.26
17	_08805	430	Aclacinomycin-T 2-deoxy-L-fucose transferase	sqnG1	99.53
18	_08810	404	L-noviosyl transferase	sqnG2	99.26
19	_08815	193	dTDP-4-dehydrorhamnose 3,5-epimerase	sqnS1	100.00
20	_08820	376	L-noviosyl transferase	sqnG3	99.47
21	_08825	355	Glucose-1-phosphate thymidylyltransferase	sqnS2	99.72
22	_08830	327	dTDP-glucose 4,6-dehydratase	sqnS3	99.08
23	_08835	353	dTDP-6-deoxy-L-talose 4-dehydrogenase (NAD(P)(+))	sqnS4	97.79
24	_08840	434	GDP-perosamine synthase	sqnS5	99.54
25	_08845	254	dTDP-glucose 4,6-dehydratase	sqnS6	98.82
26	_08850	323	Hypothetical protein	sqnO	98.45
27	_08855	466	Hypothetical protein	sqnS7	99.57
28	_08860	318	Glucose--fructose oxidoreductase	sqnS8	98.11
29	_08865	241	4'-phosphopantetheinyl transferase psf-1	sqnCC	97.88
30	_08870	530	putative propionyl-CoA carboxylase beta chain 5	sqnP	99.62
31	_08875	221	Hypothetical protein	sqnDD	96.82
32	_08880	528	Aclacinomycin-N/aclacinomycin-A oxidase	sqnQ	99.81
33	_08885	145	Hypothetical protein	sqnEE	99.31
34	_08890	76	Hypothetical protein	Not found	-
35	_08895	254	Transcriptional regulatory protein YycF	sqnR	100.00
36	_08900	639	2-oxoglutarate oxidoreductase subunit KorA	sqnT	99.51
37	_08905	364	2-oxoglutarate oxidoreductase subunit KorB	sqnU	99.16

* Sequences were aligned using protein blast

Table S6. ^{1}H and ^{13}C NMR data of 8 in DMSO-*d*6 at 298 K

Position	δC, type	δH, mult. (*J* in Hz)
1	195.8, C	
2	50.4, CH_2	3.01, d (15.7) 2.82, dd (15.7, 2.6)
3	75.8, C	
4	33.4, CH_2	3.57, dd (18.1, 2.4) 2.94, dd (18.1, 2.9)
4a	138.3, C	
5	136.7, C	
6	114.8, CH	8.16, s
6a	130.3, C	
7	187.3, C	
7a	115.6, C	
8	158.1, C	
9	137.6, C	
10	133.8, CH	7.95, d (8.0)
11	119.2, CH	7.84, d (8.0)
11a	131.7, C	
12	187.4, C	
12a	117.1, C	
12b	160.5, C	
13	24.9, CH_3	1.45, s
8-OH		12.92, s
12b-OH		12.89, s
Sugar A, β-D-olivose		
1A	70.5, CH	5.01, dd (11.2, 2.2)
2A	35.9, CH_2	2.26, m 1.62, m
3A	75.7, CH	3.88, m
4A	73.6, CH	3.52, dd (9.1, 9.1)
5A	73.5, CH	3.62, m
6A	17.4, CH_3	1.27, d (6.1)
Sugar B, α-l-cinerulose B		
1B	90.5, CH	5.23, d (2.7)
2B	70.9, CH	4.35, m
3B	39.8, CH_2	2.90, dd (17.4, 2.8) 2.48, dd (17.4, 3.7)
4B	208.7, C	
5B	77.0, CH	4.71, q (6.8)
6B	16.1, CH_3	1.24, d (6.8)

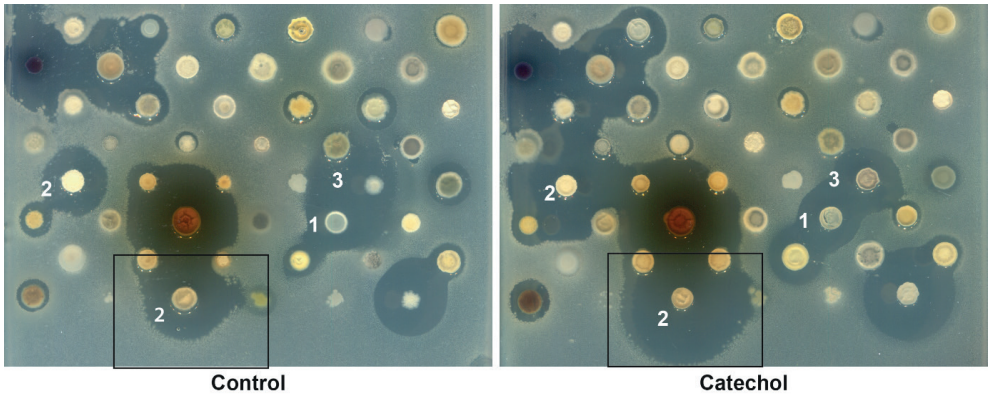


Figure S1. Catechol changes the bioactivity profile of different *Streptomyces*. *Streptomyces* sp. were spotted on MM with and without 100 μM catechol. Both inhibition (1) and promotion (2) of antibiotic production were observed. Additionally, we observed that catechol influenced interactions between different strains illustrated by the change in halo shape (3). Catechol significantly enhanced the bioactivity of *Streptomyces* sp. MBT84 (black square) (n = 3).

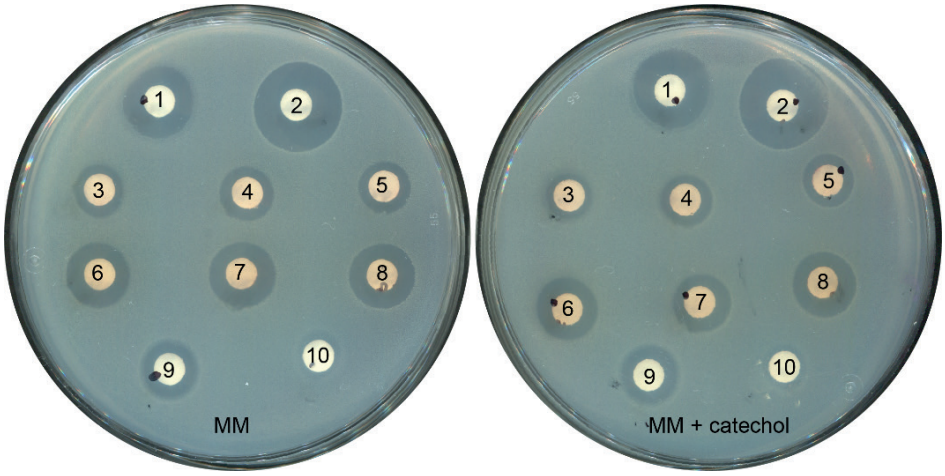


Figure S2. Addition of catechol to growth medium does not affect the susceptibility of *B. subtilis* 168 to different antibiotics and crude extracts of MBT84. MM glycerol mannitol agar plates with and without 100 μM catechol were overlaid with LB soft agar containing *B. subtilis* 168. Diffusion discs contain: 10 μL of (1) 0.5 mg/mL apramycin, (2) 0.5 mg/mL ampicillin, (3-5) crude extracts of MBT84 grown without catechol (20 mg/mL), (6-8) crude extracts of MBT84 grown in presence of catechol (20 mg/mL), (9) 1.5 mg/mL kanamycin, and (10) 5 mg/mL chloramphenicol. Experiment was conducted in duplicate with 2 replicates. No differences between the growth conditions were observed.

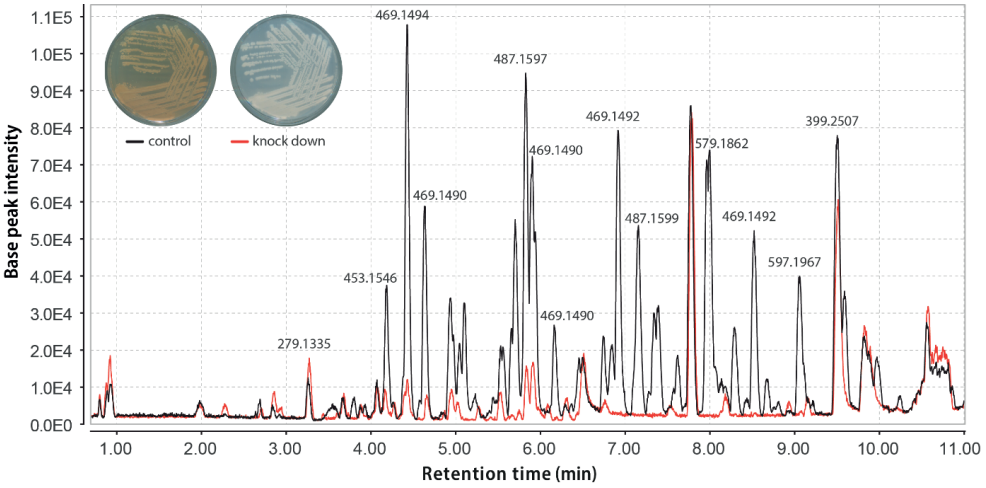


Figure S3. BGC4 specifies the angucyclines glycosides produced by *Streptomyces* sp. MBT84. Knock-down of the gene encoding KSβ of BGC4 in *Streptomyces* sp. MBT84 using CRISPRi resulted in almost complete inhibition of pigment production compared to the control strain harboring a construct that targets the template strand. The LC-MS chromatogram overlay of the crude extracts of the knock-down and control strains shows that inhibition of the expression of the KSβ resulted in decreased production of the majority of the metabolites. Three independent transformants were tested.

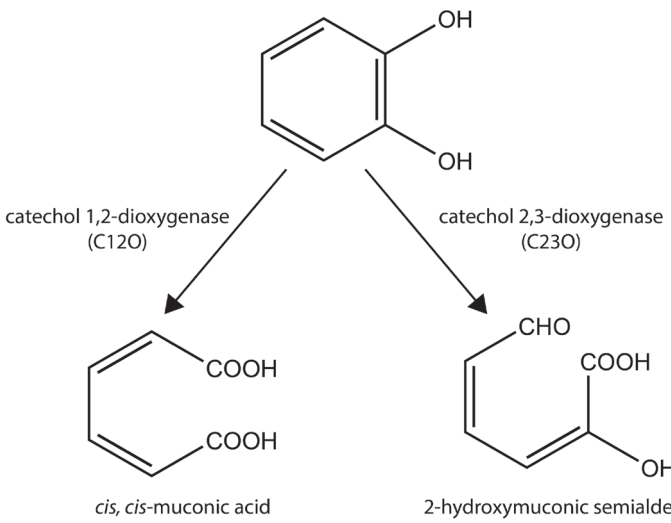


Figure S4. Degradation of catechol by catechol 1,2-dioxygenase and catechol 2,3-dioxygenase

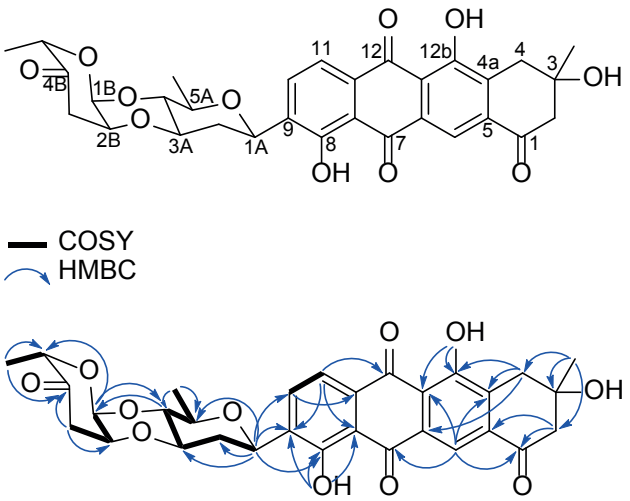


Figure S5. 2D structure (Top) and key COSY and HMBC correlations (bottom) of compound 8

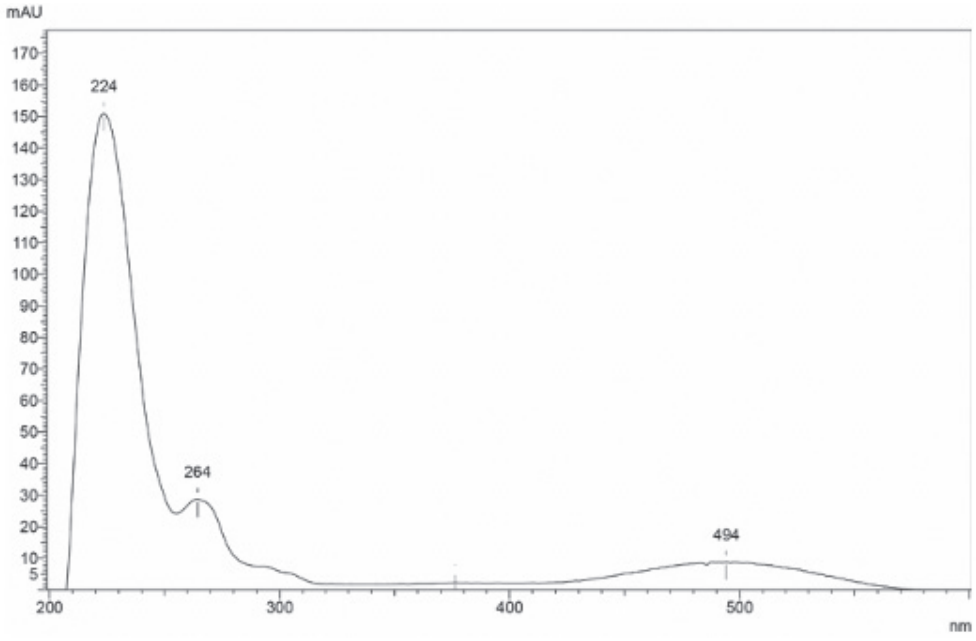


Figure S6. UV spectrum of 7

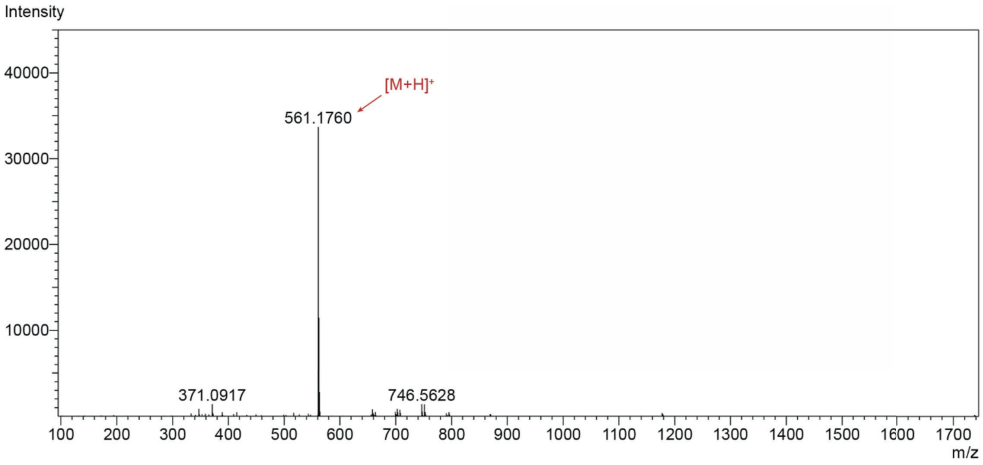


Figure S7. (+)-HR-ESI-MS spectrum of 7

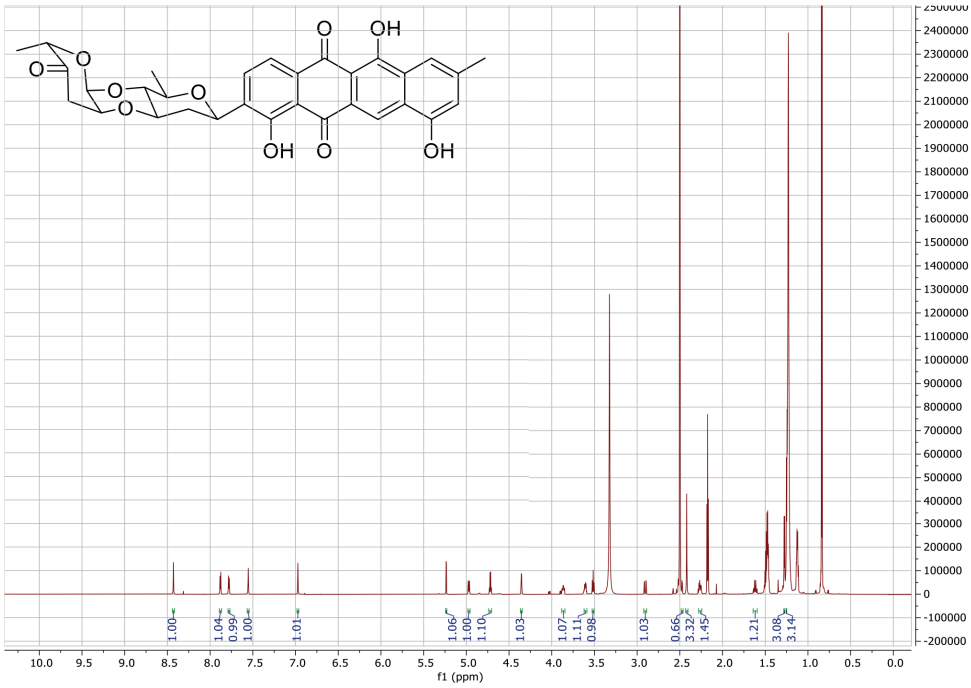


Figure S8. ^{1}H NMR spectrum of 7 (850 MHz, in DMSO-d_6)

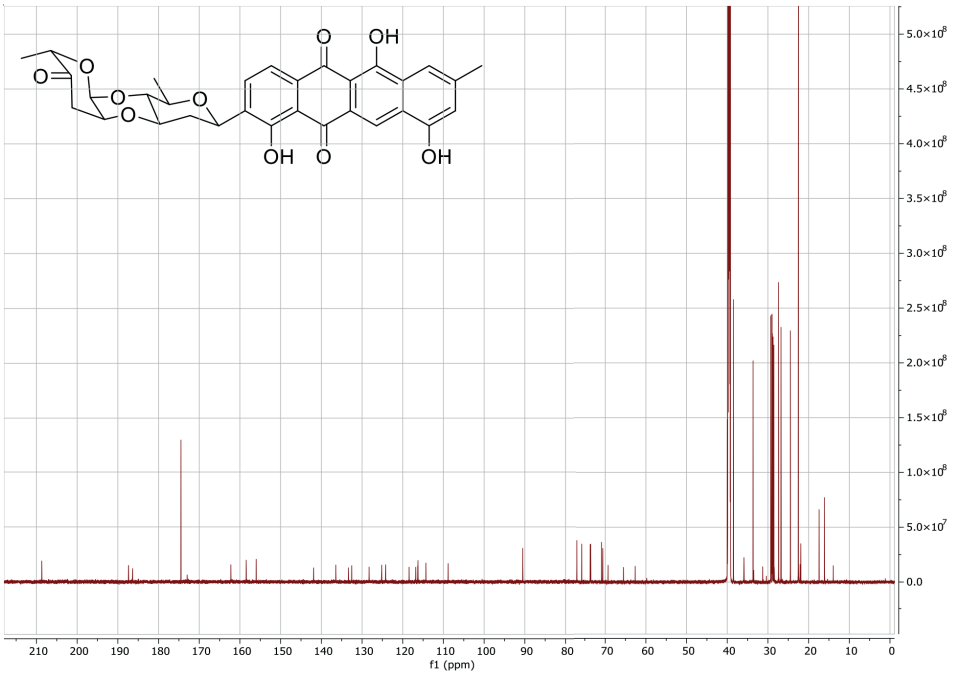


Figure S9. ^{13}C NMR spectrum of 7 (213 MHz, in DMSO-d_6)

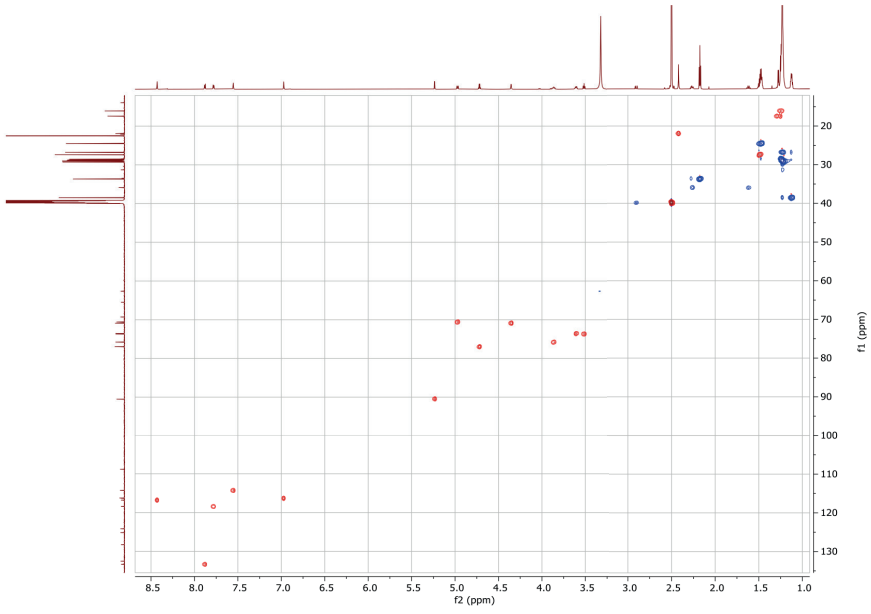


Figure S10. Multiplicity-edited HSQC spectrum of 7 (850 MHz, in DMSO-d_6)

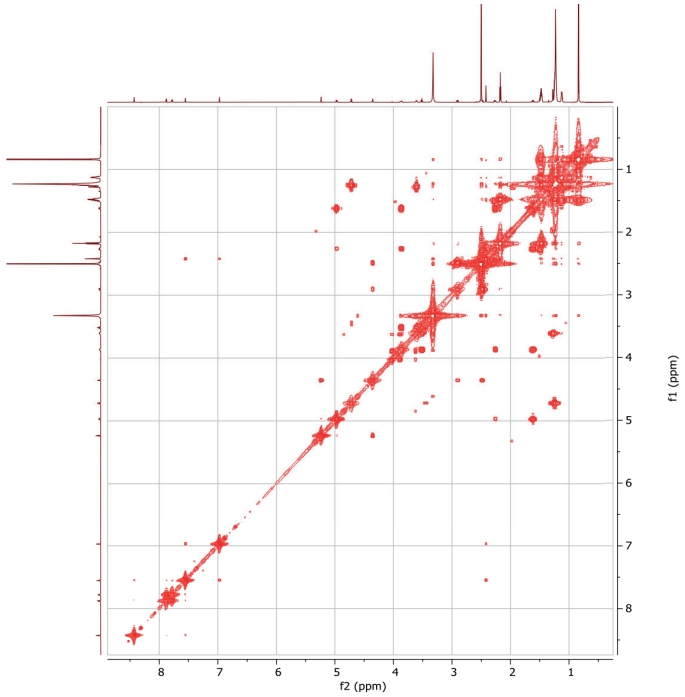


Figure S11. ^{1}H–^{1}H COSY spectrum of 7 (850 MHz, in DMSO-d_6)

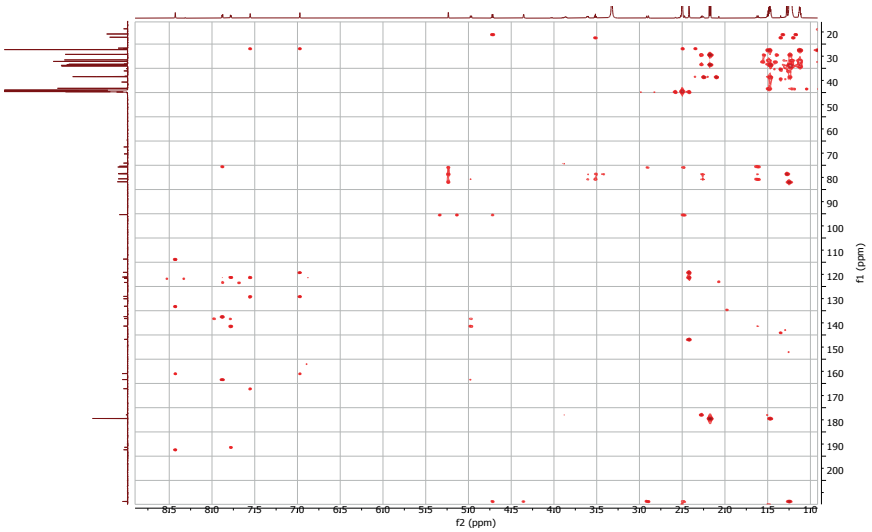


Figure S12. HMBC spectrum of 7 (850 MHz, in DMSO-d_6)

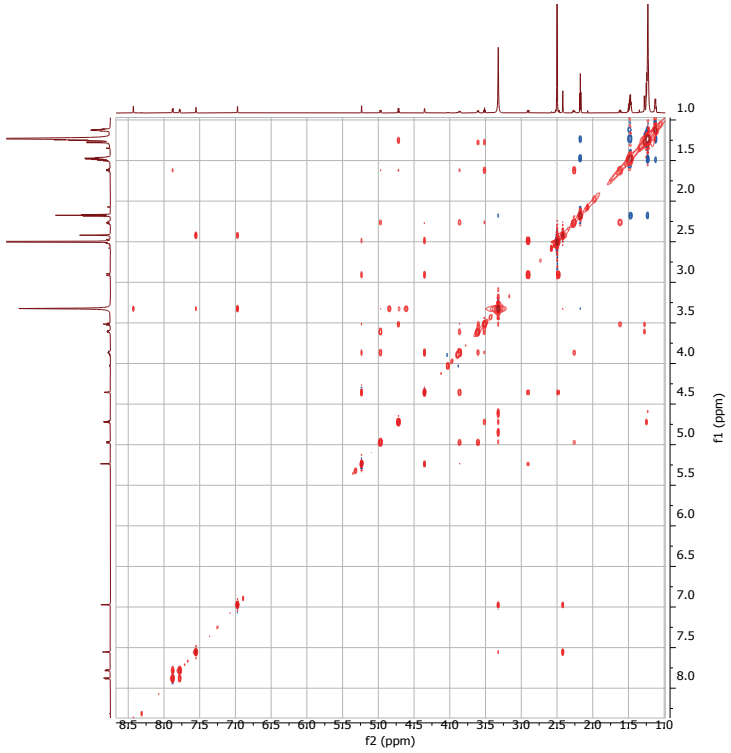


Figure S13. NOESY spectrum of 7 (850 MHz, in DMSO-d_6)

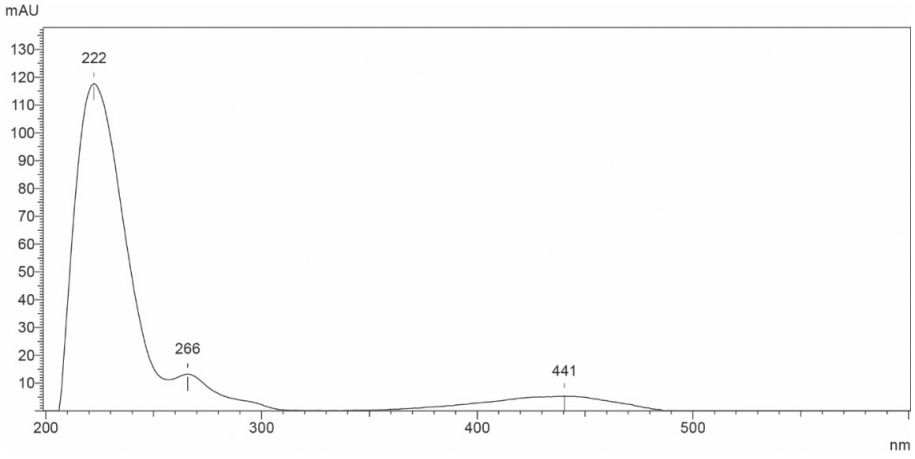


Figure S14. UV spectrum of 8

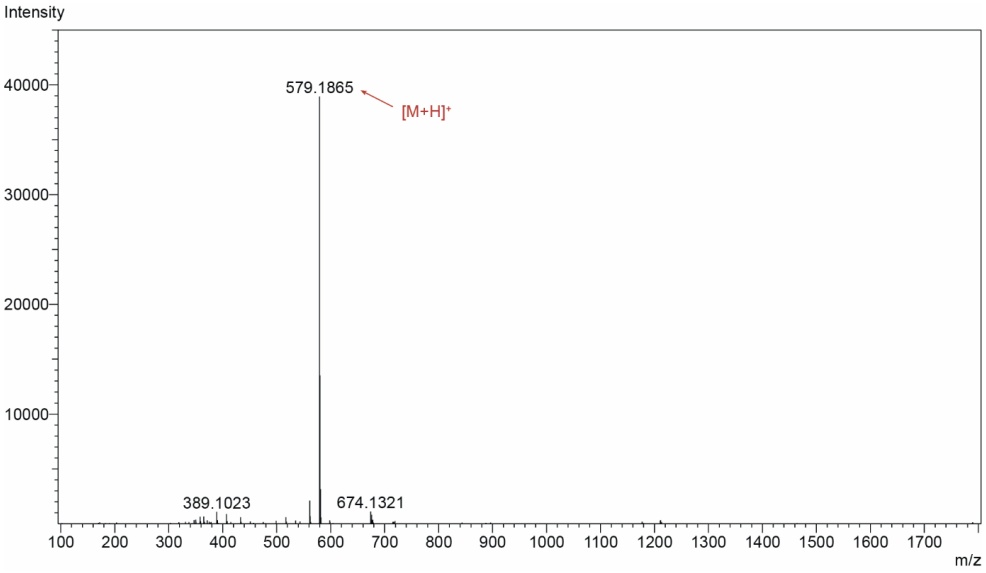


Figure S15. (+)-HR-ESI-MS spectrum of 8

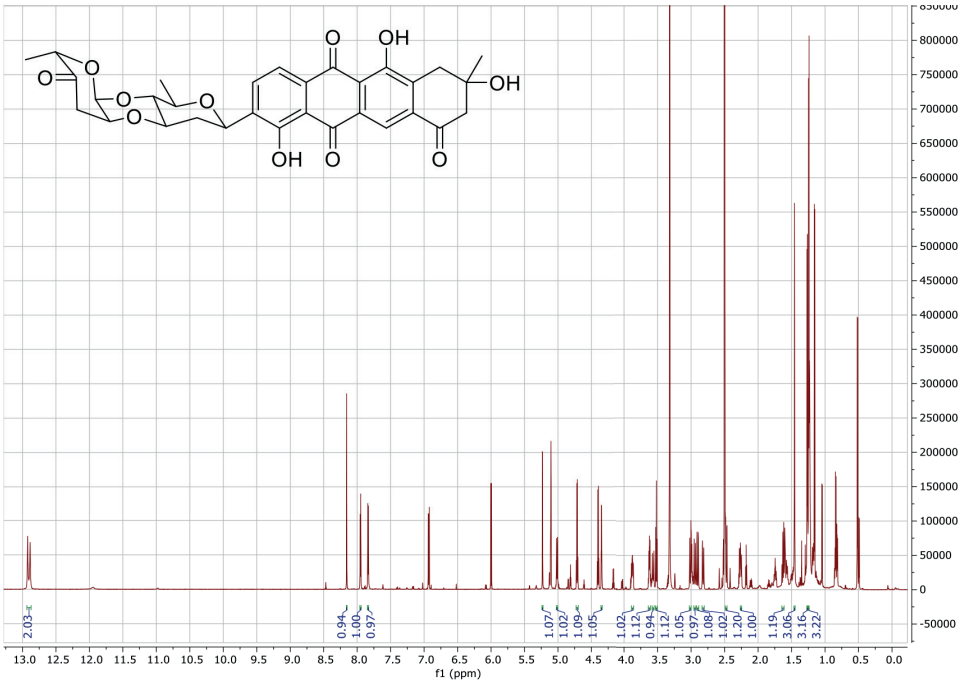


Figure S16. ^{1}H NMR spectrum of 8 (850 MHz, in DMSO-d_6)

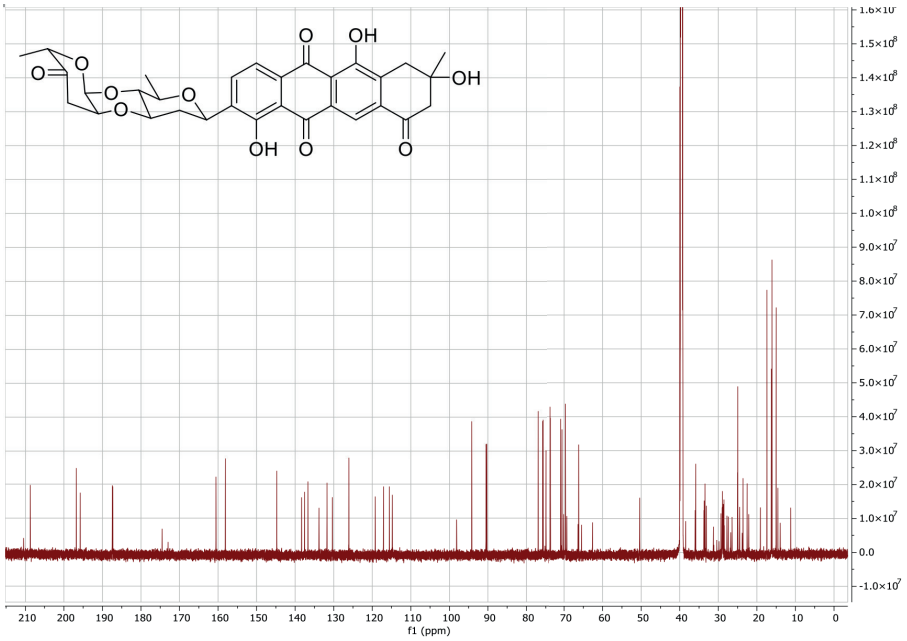


Figure S17. ^{13}C NMR spectrum of 8 (213 MHz, in DMSO-d_6)

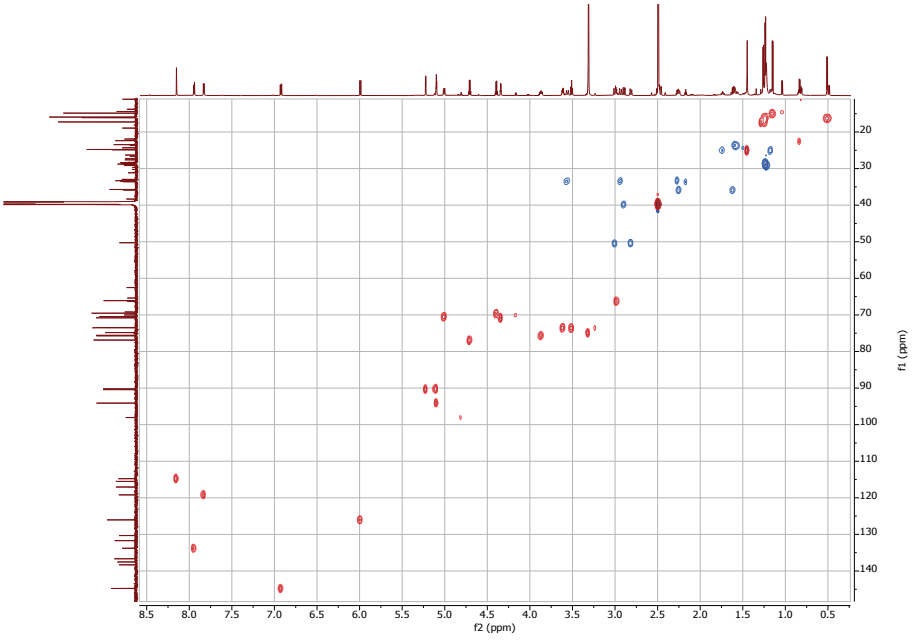


Figure S18. Multiplicity-edited HSQC spectrum of 8 (850 MHz, in DMSO-d_6)

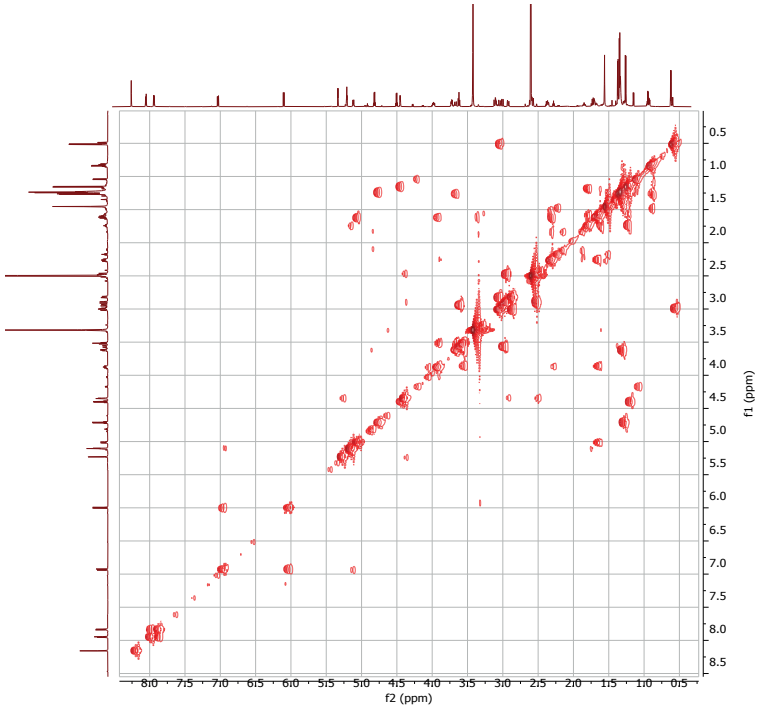


Figure S19. ^{1}H–^{1}H COSY spectrum of 8 (850 MHz, in DMSO-d_6)

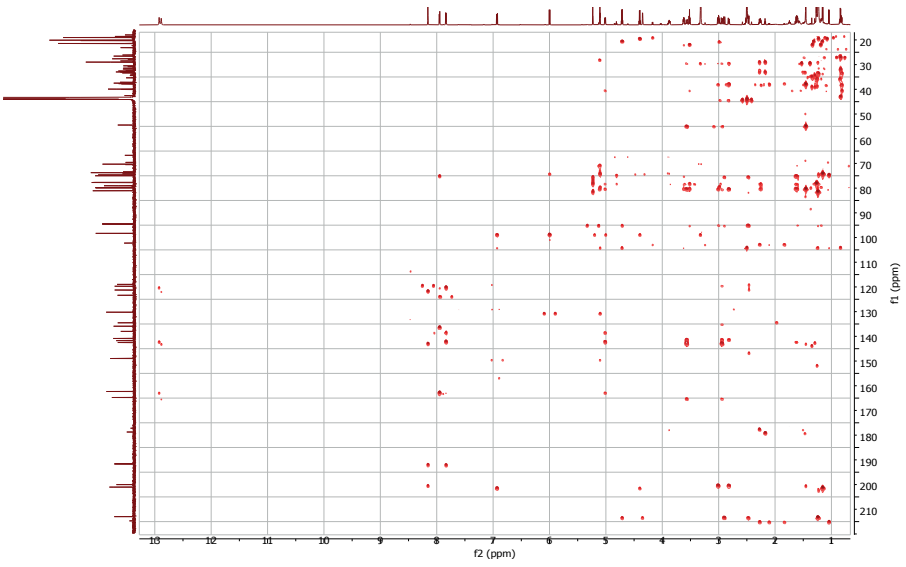


Figure S20. HMBC spectrum of 8 (850 MHz, in DMSO-d_6)

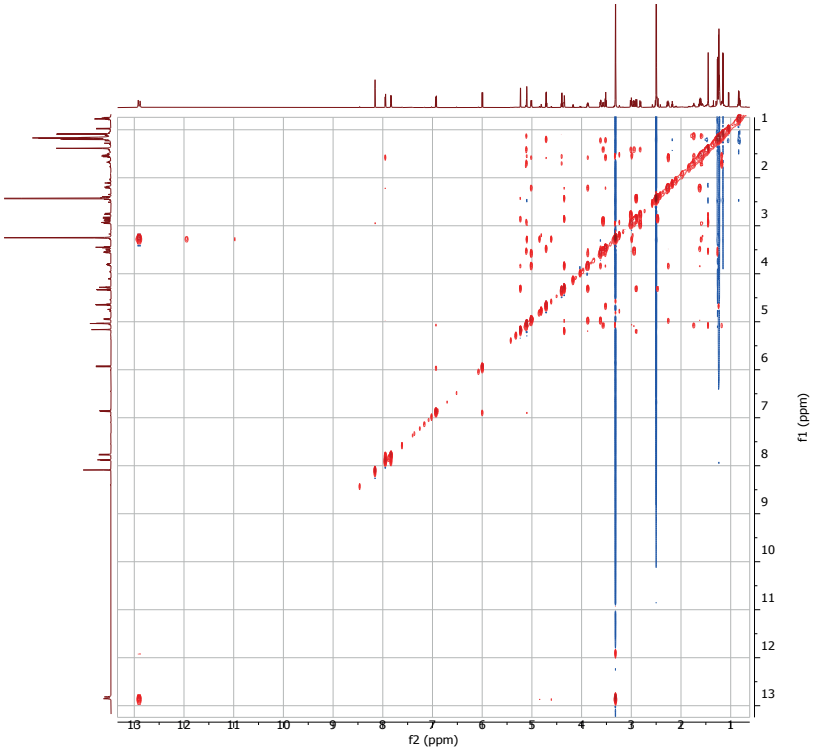


Figure S21. NOESY spectrum of 8 (850 MHz, in DMSO-d_6)

Bioactive *Pseudonocardia* within the microbiome of zebrafish

Doris A. van Bergeijk, Robert Whelan, Hannah Augustijn, Victor Carrión, Bjørn Koch, Gilles P. van Wezel, Herman P. Spaink

Abstract

Filamentous Actinobacteria are well known for their production of bioactive specialised metabolites. Many gifted producers have been isolated from the microbiomes of eukaryotic organisms, such as plants, marine organisms, and insects. Actinobacteria play an important role in the protection of these organisms against pathogens. Yet, the bioactive and functional potential of filamentous Actinobacteria within the microbiome of their eukaryotic hosts remains poorly characterised. Zebrafish larvae are excellent systems to characterise the impact of host-microbe interactions in health and disease. In this study, we investigated antibiotic-producing Actinobacteria associated with this model system. We analysed the gut microbiome of adult zebrafish and isolated Actinobacteria from zebrafish larvae, belonging to the less common actinobacterial genera *Microbacterium, Dermacoccus, Kocuria,* and *Pseudonocardia.* Genome sequencing of *Pseudonocardia* sp. ZF1 revealed much uncharacterised biosynthetic potential and high similarity to ant-associated *Pseudonocardia*. *Pseudonocardia* sp. ZF1 had antimicrobial activity against both Gram-positive and Gram-negative bacteria. Further research is required to identify the bioactive metabolites produced by the *Pseudonocardia* isolate and characterise is functional role *in vivo*.

Introduction

The bacterial world is home to a great number of specialised chemists, many of which belong to the phylum Actinobacteria [1,27]. Their specialised metabolites display an extraordinary chemical diversity, allowing Actinobacteria to thrive in a wide range of ecosystems [2]. Within these environments, siderophores facilitate the uptake of iron, compatible solutes protect against osmotic stress, and antibacterial metabolites function as signalling molecules or as chemical weaponry to outcompete other organisms [42]. Importantly, the beneficial effects provided by these metabolites are not limited to the producer, but can also provide advantage to the many multicellular organisms that host Actinobacteria [45-47]. For example, antibiotic-producing Actinobacteria are consistently associated with disease-suppressive soils, in which plants suffer less from soil-borne pathogens [47]. Another well characterised example is the fungus-farming ant system where *Pseudonocardia* reside in the cuticle of ants and protect the fungal cultivar against different pathogenic fungi through the production of bioactive metabolites [288-290].

The interactions that take place between resident bacteria, host and invading pathogens can shape microbial enzymatic potential and select for efficacious and relevant bioactivity [291]. It is therefore not surprising that microbiomes are a promising source of bioactive molecules. Sequencing of the human microbiome revealed the presence of many biosynthetic gene clusters (BGCs) across different bacterial phyla, including Actinobacteria [181,193]. Additionally, several novel bioactive molecules have been isolated from host-associated bacteria such as the novel antifungal agent cyphomycin produced by insect-associated *Streptomyces* [43], the anthracycline antibiotic keyicin from *Micromonospora* isolated from the sea squirt microbiome [292], and the antibiotic lugdunin from the human nasal commensal *Staphylococcus lugdunensis* [99]. Notably, human nasal carriage of *S. lugdunensis* was associated with reduced nasal carriage of *Staphylococcus aureus*, illustrating a potential role for antibiotic-producing commensal bacteria in the protection against (opportunistic) pathogens in humans [99].

Within the phylum Actinobacteria, most interest has been directed towards the health-promoting effects of the probiotic *Bifidobacteria* within the animal and human microbiome [293-295]. Yet, much remains to be discovered about the functional and bioactive potential of antibiotic-producing Actinobacteria within the animal microbiome. Moreover, a growing body of evidence shows that host-associated signalling molecules play a role in the activation of specialised metabolite production [296-300]. For example, the human opioid dynorphin stimulates the production of the virulence factor pyocyanin in *Pseudomonas aeruginosa* [242], and plant- and animal-associated stress hormones stimulate antibiotic and siderophore production in *Streptomyces* [44,301] (Chapter 4). We hypothesise that upon host stress, such as infection,

host stress molecules might serve as a 'cry for help' and activate production of specialised metabolites by Actinobacteria that can aid in the protection against invading pathogens [2]. To test this hypothesis, models are needed to study whether such interactions occur *in vivo* and whether Actinobacteria can provide protection against invading pathogens.

Over the past years, zebrafish larvae have proven to be a relevant model for microbiome research *in vivo* [302-305]. Potentially, they can serve as a model to explore the functional role of Actinobacteria within the microbiome of animals and to study the involvement of host-microbe interactions in the activation of specialised metabolite production. For this purpose, we analysed the presence of Actinobacteria within the gut microbiome of adult zebrafish and isolated Actinobacteria from zebrafish larvae, including a bioactive *Pseudonocardia* sp. with much uncharacterised biosynthetic potential. This study provides a first step towards exploring the functional role of antibiotic-producing Actinobacteria within the animal microbiome. Further research is required to identify the bioactive metabolites produced by the *Pseudonocardia* isolate and test its protective effect *in vivo*.

Results

Actinobacteria in the zebrafish microbiome

We aimed to isolate zebrafish-associated Actinobacteria to provide a first step towards using zebrafish as an *in vivo* model to explore the bioactive and functional potential of Actinobacteria within the animal microbiome. To get an overview of the genera that are present in the gut microbiome of zebrafish, 16S rRNA amplicon sequencing was performed. For this, the intestine of three adult zebrafish were removed and homogenised, and total DNA extracted. The DNA samples were analysed by 16S rRNA amplicon sequencing of the V3-V4 hypervariable region using the MiSeq Illumina sequencing platform. Similar to previous reports, the community was numerically dominated by Proteobacteria (Fig. 1A). Actinobacteria comprised on average 0.97% of the filtered reads (Fig. 1B). Sequence variants were assigned to the genera *Collinsella, Corynebacterium, Dermacoccus, Slackia, Bifidobacterium, Rothia*, *Nesterenkonia, Kocuria, Brevibacterium, Actinomyces*, *Marmicola, Cutibacterium, Lawsonella, Micrococcus, Turicella*, *Pseudonocardia,* and *Rhodococcus*.

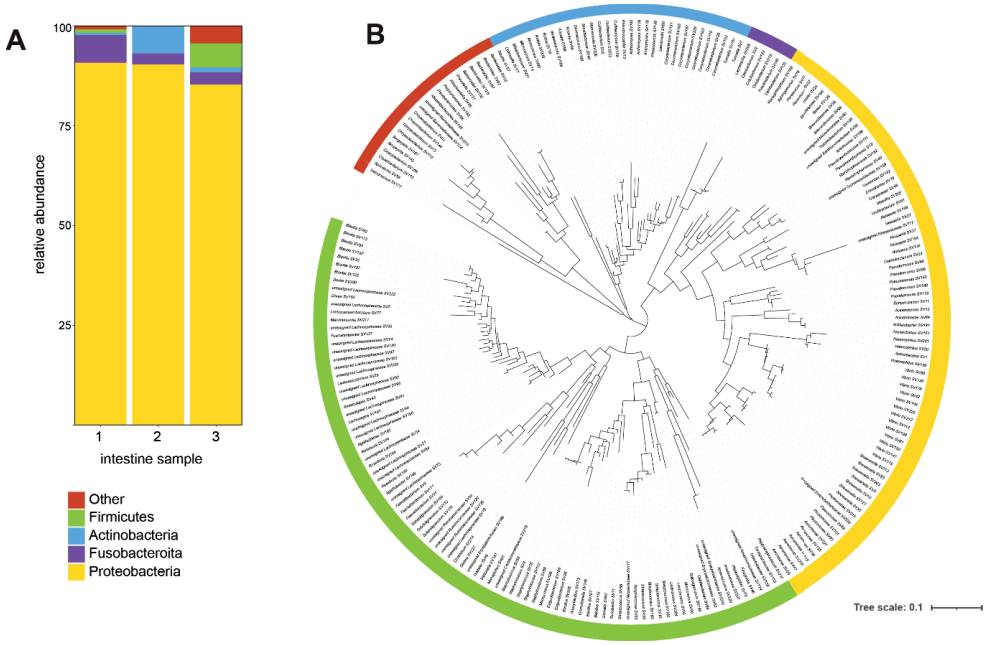


Figure 1. 16s rRNA taxonomic profiling of the adult zebrafish intestinal microbiome. (A) Taxonomic composition bar plot illustrating the phylum-level composition of the adult intestinal microbiota. (B) Midpoint-rooted maximum likelihood phylogenetic tree of adult zebrafish gut microbiota amplicon sequence variants.

For the isolation of Actinobacteria, zebrafish larvae at 5 days post fertilization (dpf) were washed in sterile PBS and tissue was homogenised. The homogenate was then serially diluted and plated onto media selective for Actinobacteria. Bacteria were isolated based on several morphological characteristics (non-motile, filamentous growth, and/or dry colony morphology) and grown on different media for phenotypic discrimination. Based on this, 12 strains were selected. Sequencing of the 16S rRNA gene revealed that six isolates belonged to the phylum Actinobacteria spanning five genera, namely *Dermacoccus, Kocuria, Microbacterium*, *Micrococcus*, and *Pseudonocardia* (Table 1). The other isolates belonged to the genera *Sphingomonas, Mesorhizobium*, and *Staphylococcus*. All of these isolated genera were detected in the amplicon sequencing data of the adult zebrafish gut, except *Microbacterium.*

Table 1. Isolated bacteria and their closest neighbour using (partial) 16S rRNA gene sequencing

Isolate	Top-hit EzBioCloud	similarity	Completeness
ZF0	*Staphylococcus warneri* ATCC 27836	100%	47.5
ZF1	*Pseudonocardia carboxydivorans* Y8	99.93%	99.9
ZF4	*Sphingomonas paucimobilis* NBRC 13935	100%	49.6
ZF9	*Micrococcus luteus* NCTC2665	100%	48.4
ZF22	*Sphingomonas sanguinis* NBRC 13937	99.86%	49.7
ZF23	*Sphingomonas yabuuchiae* GTC 868	100%	49.6
ZF27	*Microbacterium hominis* NBRC 15708	99.14%	48.4
ZF30	*Mesorhizobium terrae* NIBRBAC000500504	100%	49.8
ZF31	*Kocuria arsenatis* CM1E1	100%	48.2
ZF33	*Microbacterium hominis* NBRC 15708	99.14%	48.4
ZF37	*Sphingomonas sanguinis* NBRC 13937	99.86%	49.7
ZF50	*Dermacoccus nishinomiyaensis* DSM 20448	99.57%	48.5

Genomic analysis of *Pseudonocardia* sp. ZF1

The richness of bacterial genomes in terms of BGC diversity, has been correlated with key organismal features, including multicellularity [1]. Of the isolates, only *Pseudonocardia* sp. ZF1 exhibited filamentous growth and this strain was therefore selected for further characterisation. *Pseudonocardia* sp. ZF1 produced a cream-coloured aerial spore mass and orange substrate mycelia (Fig. 2A). No diffusible pigments were produced. Also in liquid, no diffusible pigments were produced and growth was characterised by the formation of a branched mycelium and pellets (Fig. 2B).

To gain insights into the relatedness to known *Pseudonocardia* isolates, we obtained the complete genome sequence using a combination of Nanopore and Illumina sequencing. Combined assembly resulted in a draft genome with a G+C content of 73.9% consisting of 3 contigs of 5.66 Mb, 494 kb, and 25.6 kb in length, all of which were predicted to be circular (Table S1, Fig. 3). The 5.66 Mb contig represents the genome sequence, while the smaller contigs (ZFp1 and ZFp2) likely represent extrachromosomal plasmids as they contained many genes for mobile genetic elements, such as transposases and integrases (Fig. 3).

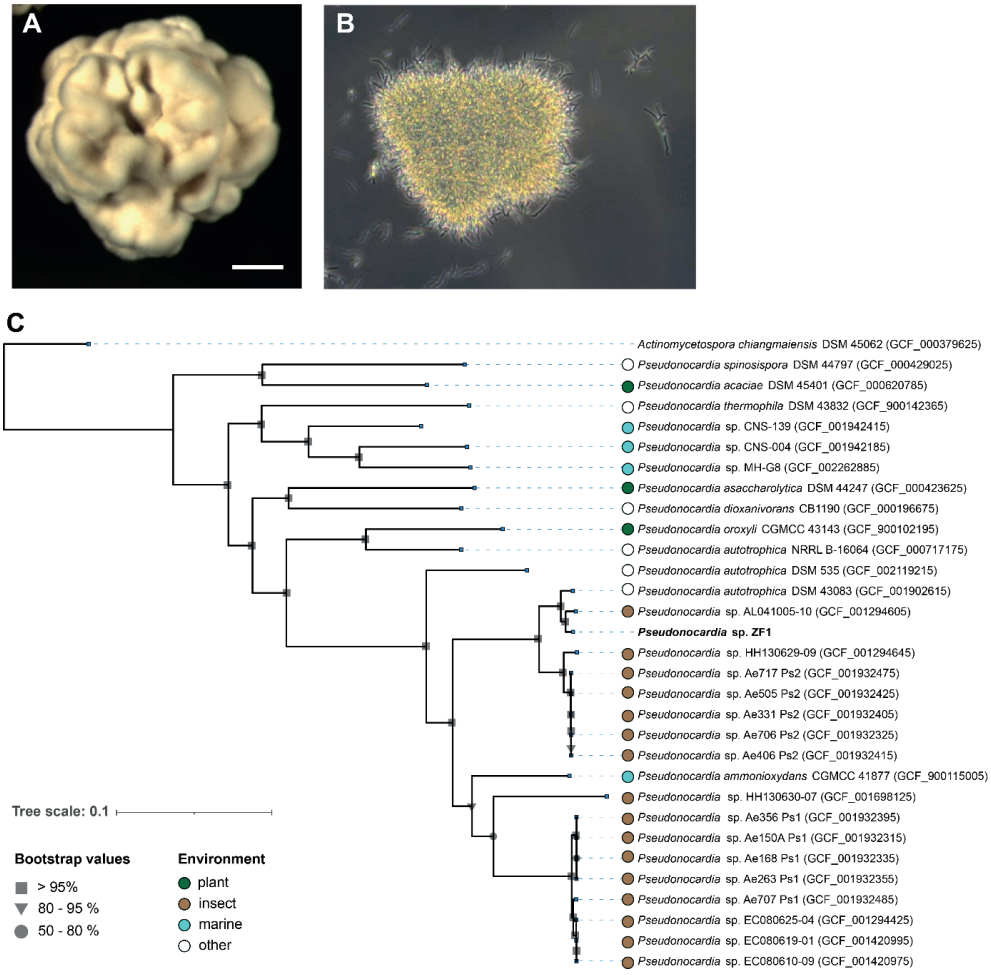


Figure 2. Phenotypic and taxonomic characterization of *Pseudonocardia* sp. ZF1. (A) *Pseudonocardia* sp. ZF1 produces cream-coloured colonies with an orange substrate mycelium when grown on ISP2 agar. Scale bar: 1 mm **(B)** In liquid ISP2, growth is characterised by the formation of a branched mycelium and pellets. **(C)** Genome-scale phylogenetic tree based on Multi-Locus Sequence Analysis of *Pseudonocardia* sp. ZF1 and different members of the genus *Pseudonocardia*. The tree was constructed using autoMLST [306]. *Actinomycetospora chiangmaiensis* DSM 45062 was used as an outgroup. Symbols indicate the bootstrap values in percentage of a total of 1000 bootstrap. Colours indicate from which environment the strains have been isolated.

A phylogenetic tree was constructed based on whole-genome sequences using autoMLST [306] (Fig. 2C). *Pseudonocardia* sp. ZF1 formed a monophyletic group with the fungus-growing ant-associated *Pseudonocardia* sp. AL041005-10 [48] and *Pseudonocardia autothrophica* DSM 43083 [307]. To assess the genomic distance between *Pseudonocardia* sp. ZF1 and its closest neighbour *Pseudonocardia* sp. AL041005-10, the average nucleotide identity (ANI) was calculated in an all-to-all genome comparison, revealing a similarity score of 97.21%. The species delineation threshold typically lies around 95% gene identity [232], suggesting that *Pseudonocardia* sp. ZF1 and *Pseudonocardia* sp. AL041005-10 belong to the same species. Comparative genome analysis of *Pseudonocardia* sp. ZF1 and *Pseudonocardia* sp. AL041005-10 indeed showed high similarity between the strains. No plasmid sequences similar to plasmids ZFp1 and ZFp2 were found in *Pseudonocardia* sp. AL041005-10 (Fig. S1).

Biosynthetic potential of zebrafish isolate *Pseudonocardia* sp. ZF1

Actinobacteria isolated from the microbiomes of different eukaryotic hosts have promising bioactive potential. To study the biosynthetic potential of *Pseudonocardia* sp. ZF1, the genome sequence was analysed using antiSMASH [217]. 15 BGCs were identified, the majority of which was located on the chromosome. Four BGCs were located on plasmid ZFp1, and no BGCs were identified on ZFp2 (Table S2). Only one BGC matches a known BGC with 100% similarity, specifying the compatible solute ectoine, which is commonly found in bacteria [223]. Notably, two-third of the BGCs had less than 25% homology to BGCs in the MIBiG database (Table S2).

We then wondered whether these BGCs were unique to *Pseudonocardia* sp. ZF1 or whether they were shared among *Pseudonocardia* spp. For this, the BGCs were analysed using cblaster [309], a tool for finding clusters of co-located sequences using BLAST searches. The majority of the BGCs were shared with other *Pseudonocardia* species, such as *Pseudonocardia alni* DSM44104, *Pseudonocardia* sp. ICBG1034, and *Pseudonocardia autotrophica* DSM 43083 (Table S2). However, for the BGCs located on plasmid ZFp1, almost no homology was found, also when the search was limited to only a small subset of genes of the predicted BGCs. In the comparative genome analysis of *Pseudonocardia* sp. ZF1 and its closest neighbour *Pseudonocardia* sp. AL041005-10, the BGCs on ZFp1 were not shared (Fig. S1). Notably, the best hit for BGC2.2 was found on plasmid pPA12743CP of *Pseudonocardia autotrophica* NBRC 12742, further validating ZFp1 most likely to be a plasmid.

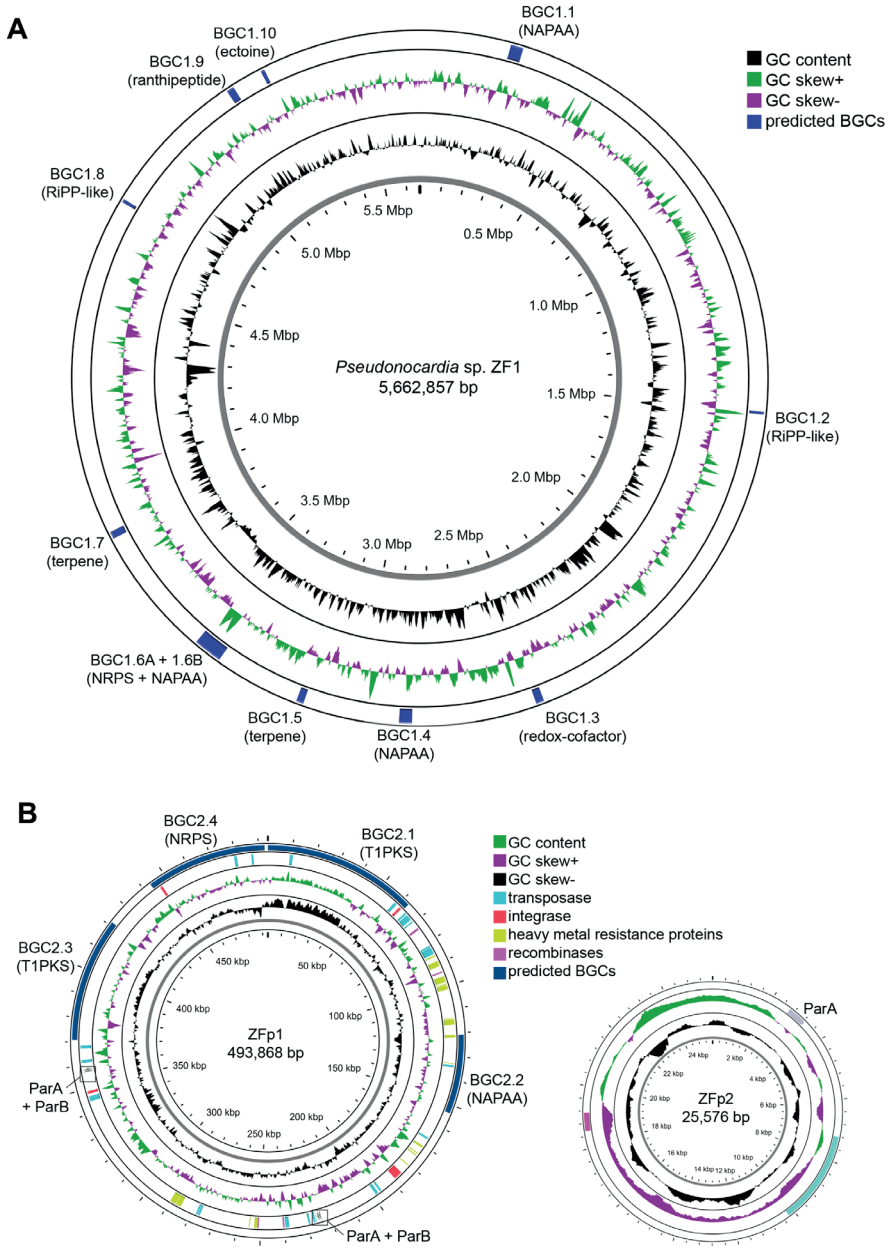


Figure 3. Schematic visualization of the circular chromosome of *Pseudonocardia* sp. ZF1 and the extrachromosomal elements. A-D) Representations of the chromosome (**A**) and plasmids ZFp1 and ZFp2 (**B**) were obtained using the CGview server [308]. In addition to GC content and CG skew, the location of the BGCs predicted by antiSMASH [217], and several plasmid elements are shown. Additionally, the ParA and ParB partition proteins are indicated. NAPAA: non-alpha poly-amino acids like e-Polylysin, NRPS: non-ribosomal peptide synthetase cluster, RiPP: ribosomally synthesised and post-translationally modified peptide product cluster, T1PKS: type 1 polyketide synthase.

Next, the antibiotic-producing potential of *Pseudonocardia* sp. ZF1 and the other isolated Actinobacteria was assessed. The isolates were grown on the nutrient-rich growth medium Nutrient Agar and the nutrient-poor Minimal Medium. Plates were inoculated with spots from glycerol stocks of the different isolates and after seven days of growth, plates were overlaid with soft agar containing the Gram-positive *Bacillus subtilis* 168 or the Gram-negative *Escherichia coli* ASD19. Of the isolates, only *Pseudonocardia* sp. ZF1 inhibited the growth of the tested indicator strains (Fig. 4). To identify the nature of the metabolites produced by *Pseudonocardia* ZF1, the strain was streaked on MM and NA agar plates, followed by ethyl acetate (EtOAc) extraction of the metabolites from the spent agar after seven days of growth. However, the crude extracts did not exhibit bioactivity. Also dereplication based on MS/MS spectra against the GNPS spectral library did not result in annotation of any known bioactive compounds [220]. Unpublished proteomics data indicates that three predicted BGCs are expressed when *Pseudonocardia* sp. ZF1 is grown on MM medium: BGC1.2 with no similarity to any known BGCs in the MIBiG repository, BGC1.6A with 60% similarity to the scabichelin BGC of *Streptomyces scabiei* 87.22, and the ectoine BGC1.10. Future research is needed to identify the bioactive metabolites and the cognate BGC, and analyse whether *Pseudonocardia* sp. ZF1 produces these molecules *in vivo.*

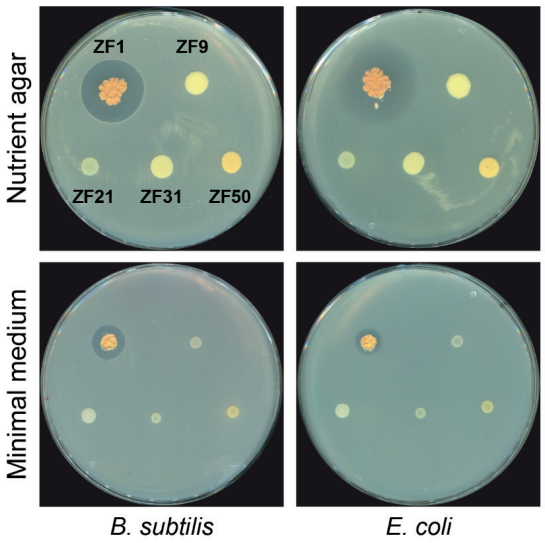


Figure 4. Antimicrobial activity of zebrafish isolates. After 4 days of growth, plates were overlaid with an indicator strain to test for antimicrobial activity (n = 3). Of the isolates, only *Pseudonocardia* sp. ZF1 inhibited the growth of *B. subtilis* and *E. coli.*

Discussion

As nature's medicine makers, Actinobacteria are welcome inhabitants of the microbiomes of eukaryotes, among others offering protection against invading pathogens. They likely do so in response to host-associated signaling molecules [2,44,301]. While a lot of attention is directed towards plant- and insect-Actinobacteria symbioses, the role of bioactive Actinobacteria within the microbiome of vertebrate organisms remains poorly characterised. Zebrafish larvae have been extensively used to study the role of the microbiome in health and disease. In this study, we explored the zebrafish microbiome for antibiotic-producing Actinobacteria to provide a first step towards using zebrafish as an *in vivo* model to study the bioactive and functional potential of Actinobacteria within animal microbiome.

16S rRNA amplicon sequencing revealed the presence of several actinobacterial genera commonly found in the human microbiome, such as *Bifidobacterium*, *Collinsella*, *Actinomyces*, *Rothia,* and *Corynebacterium* [310]. While several of these genera have received much attention with regard to their role in digestion and their abundance in correlation to health and disease, little is known about their specialised metabolites. Yet, a systematic analysis of the biosynthetic gene clusters in the human microbiome highlighted several BGCs that are harboured by members of *Corynebacterium. Actinomyces,* and *Rothia* [181]. Also, *Rothia* sp. isolated from the upper respiratory tract of weaned healthy piglets produce valinomycin *in vitro* and *in vivo* [311]. These findings highlight that much remains to be discovered about commensal Actinobacteria and their specialised metabolites in the animal microbiome. We failed to isolate members of these genera from zebrafish larvae. This is likely because of the isolation method used (aerobic and focused on filamentous Actinobacteria) and the young age of the zebrafish larvae. Future isolation studies will therefore focus on adult zebrafish and the use of different isolation methods and media suitable for both aerobic and anaerobic organisms.

Of the isolates, only *Pseudonocardia* sp. ZF1 exhibited filamentous growth, which has been linked to the propensity to produce bioactive metabolites and a high richness in BGC diversity. Therefore, the genome of *Pseudonocardia* sp. ZF1 was sequenced, resulting in three contigs, predicted to be circular. Contig 1 likely represents the chromosome, while the two smaller contigs (designated ZFp1 and ZFp2) may represent extrachromosomal plasmids. Indeed, different *Pseudonocardia* spp. carry one or multiple plasmids and may horizontally acquire BGCs first on their plasmids, which are later integrated in the chromosome [48,312,313]. Indeed, almost all predicted BGCs on contig 1 were shared with other *Pseudonocardia* spp., while the predicted BGCs on contig 2 were mostly unique to *Pseudonocardia* sp. ZF1. Due to the high levels of BGC diversity, plasmids of *Pseudonocardia* might be fruitful targets for drug discovery.

Pseudonocardia sp. ZF1 displays high similarity to *Pseudonocardia* sp. AL041005-10, isolated from the fungus-growing ant *Trachymyrmex cornetzi* [48]. The association of *Pseudonocardia* spp. with ants is one of the most well studied insect-microbe symbioses. Different *Pseudonocardia* sp. are present in the ant cuticle and aid in the protection of the ant's fungal garden against different fungal parasites [288-290]. Potentially, *Pseudonocardia* (or other bioactive Actinobacteria) can play a role in the protection of fish against fungal infections. Fungal infections by oomycetes, such as *Saprolegnia*, can cause severe losses of freshwater fish and their eggs in both nature and commercial fish farms [45,314]. Indeed, high abundance of specific Actinobacteria on salmon eggs has been correlated to a low incidence of *Saprolegnia* [45]. Of these Actinobacteria, *Streptomyces* isolates displayed antifungal activity against *Saprolegnia* while isolates of the genus *Frondihabitans* effectively inhibited its attachment to the salmon eggs, suggesting multiple protective mechanisms. It remains to be tested whether *Pseudonocardia* may also play a role in the protection of zebrafish against oomycetes.

Pseudonocardia spp. have mainly been isolated from soil, plants, marine sediments, and insects [315]. In this study, we have now isolated a fish-associated *Pseudonocardia* from zebrafish larvae. Using 16S rRNA amplicon sequencing we also confirmed a low abundance of *Pseudonocardia* within the intestine of adult zebrafish. This corresponds to several other studies analysing the microbiome composition of larvae and adult zebrafish [316-318]. Isolation of *Pseudonocardia* sp. ZF1 strengthens the finding that *Pseudonocardia* spp. are associated with zebrafish. The bioactivity of the isolate against Gram-positive and Gram-negative bacteria may suggest a protective role for *Pseudonocardia* sp. ZF1 against bacterial infections. Future studies are needed to identify the bioactive metabolites produced and to analyse whether the molecules are produced *in vivo*. Additionally, it would be of interest to test whether inoculation of zebrafish eggs and larvae with *Pseudonocardia* sp. ZF1 can protect against fungal and bacterial pathogen invasion.

Taken together, we show that antibiotic-producing *Pseudonocardia* with much uncharacterised potential are associated with zebrafish, providing a first step towards the use of zebrafish as an *in vivo* model to explore the functional potential of bioactive Actinobacteria within the animal microbiome.

Methods

16S rRNA gene amplicon sequencing of microbial content zebrafish intestine

Zebrafish were handled in compliance with animal welfare regulations and maintained according to standard protocols (http://zfn.org**).** The use of adult zebrafish was approved by the local animal welfare committee (DEC) of the University of Leiden (license number:

AVD1060020171767) and adhered to the international guidelines specified by the EU Animal Protection Directive 2010/63/EU. Adult zebrafish (n = 3) were euthanized using tricaine methanesulfonate (MS-222), the intestines were extracted surgically, transferred in one piece to a sterile Eppendorf tube containing 500 μL PBS, and cut lengthwise with sterilised scissors. Samples were vortexed thoroughly to free intestinal contents, after which the host tissue was removed. Next, samples were centrifuged at maximum speed, and supernatant was removed. DNA was extracted from the remaining pellet using phenol:chloroform extraction.

16S rRNA gene amplicon sequencing was performed on the V3-V4 hypervariable region of the microbial content of the adult zebrafish intestine using primers 341F (5'-CCTACGGGNGGCWGCAG-3') and 805R (5'-GACTACHVGGGTATCTAATCC-3'). 2 x 300 bp paired-end demultiplexed sequenced reads were generated using the Miseq sequencing platform by BaseClear, The Netherlands. Reads were quality filtered based on the sequence centres Illumina Chastity filtering and FASTQC quality control tool (v 0.11.5) The adult intestinal zebrafish data comprised of three samples, with the number of reads retrieved for each sample ranging from 9,173 to 92,725**.**

The Qiime2 (v2020.8) bioinformatics platform [319] was used to analyse the 16S rRNA sequencing data. DADA2 (v1.10.0) [320] was used to model and correct amplicon sequence errors, providing amplicon sequence variants (ASV) and associated abundances. The primers were trimmed, and the forward and reverse sequences were truncated to remove low-quality regions (positions 288 and 222 respectively). MAFFT (v7.471) [321] was used to perform multiple sequence alignment and FastTree (2.1.10) [322] was used to generate a phylogenetic tree. Taxonomy was assigned using a naïve Bayesian classifier and using the Silva v138 ribosomal RNA reference database [323], which was made Qiime2 compatible using the RESCRIPt package. ASVs assigned to the taxonomic labels chloroplast and mitochondria were removed from the analysis.

Isolation of Actinobacteria from zebrafish larvae

At five days post fertilization, 100 larvae of ABxTL wild-type zebrafish were washed in sterile PBS and transferred to a 1.5 mL SafeLock tube containing 100 μL sterile PBS and 5 zirconium oxide beads (1.0 mm diameter). The larvae were anesthetised for 2 min on ice, homogenised for 2 x 1 min in a tissue homogenizer (Bullet Blender model Blue-CE; Next Advance) and cooled on ice in between. For the isolation of Actinobacteria, 20 µL homogenate was serial diluted (10^{-1} – 10^{-6}) in sterile PBS, and 100 µL of the dilutions were plated on selective media containing nystatin (50 μg/mL) and nalidixic acid (10 μg/mL) for the inhibition of fungi and Gram-negative bacteria, respectively. The media were Actinomycete Isolation Agar (Difco), Czapek Dox (Difco), ISP2 medium (DSMZ #987) . ISP5 medium (DSMZ #993), Kuster's agar

[324], Trypton Yeast Extract agar (DSMZ #680), Minimal Medium [227] and Humic Acid agar [226]. Plates were incubated at 30 °C and isolates were obtained by loop-streaking on fresh agar media until pure. For stock preparation, isolates were grown in ISP2 liquid medium and stocks were prepared according to Kieser et al. [227] and stored at −80 °C.

16S rRNA-based bacterial identification

Genomic DNA was isolated and purified from overnight liquid-cultures as described previously [227]. The 16S rRNA genes of the isolates were amplified in polymerase chain reactions (PCR) using the universal primers 8F (5'-AGAGTTTGATCCTGGCTCAG-3') and 1492R (5'-GGTTACCTTGTTACGACTT-3'). Each PCR reaction (50 µL) contained 5 µL 5x PFU buffer, 5 µL DNTP mix (20 mM), 1.5 µL of each primer (10 µM) 5 µL DMSO, 1 µL PFU polymerase, and 1 µg gDNA in dH_2O. Amplification was achieved under the following conditions: an initial denaturation step of 3 min at 95 °C, followed by 35 cycles of denaturation (30 sec at 95 °C), annealing (30 sec at 56 °C), and extension (2 min at 72 °C), and a final extension step of 7 min at 72 °C. Sequencing was done by BaseClear, Leiden using Sanger sequencing with primer 8F. A 700 bp high quality region of the obtained sequence chromatograms was selected using Chromas Lite v2.01 and compared against the EZBioCloud 16S rRNA Database.

Genome sequencing

Strains were cultured in TSBS:YEME (50:50) at 30 °C with 200 rpm shaking speed [227]. Genomic DNA was isolated by phenol-chloroform extraction as described previously [227] and sent to be commercially sequenced at Future Genomics Technologies, The Netherlands. Genomes were sequenced using Illumina and Nanopore sequencing. Hybrid assembly was performed for each isolate using Unicycler (v0.4.0.7) [229]. Briefly, Unicycler performs a SPAdes assembly using the Illumina reads and then scaffolds the assembly graph using long reads. Unicycler polishes its final assembly with Illumina reads and uses Pilon to reduce the rate of small base-level errors [230]. The genome was annotated with the NCBI Prokaryotic Genome Annotation Pipeline and the genome sequence has been deposited at GenBank under accession number JAJNBY000000000.

Phylogenetic, bioactive potential and comparative genomics analysis

A phylogenetic tree was constructed based on whole-genome sequences using autoMLST [306]. AntiSMASH v6.0 [217] was used under default settings to predict BGCs from *Pseudonocardia* sp. ZF1 and its closest neighbour *Pseudonocardia* sp. AL041005-10. For the comparative genome analysis between these strains, BLASTP was used to identify similarity, where after MCScanX [235] was utilised to predict gene collinearity and synteny. Finally, the coordinates of the collinear regions and predicted BGCs were used as input for the visualizing tool Circos

v0.69-8 [236]. Cblaster [309] with default settings was used to compare the predicted BGCs of *Pseudonocardia* sp. ZF1 against NCBI sequence databases.

Antimicrobial activity assays

Strains were manually spotted (2 μL) on Nutrient Agar (Difco) and Minimal Medium agar plates (MM) supplemented with 0.5% mannitol and 1% glycerol (w/v) as non-repressing carbon sources. Both media were supplemented with 25 mM TES buffer. After four days of incubation, plates were overlaid with soft LB agar (1.8% w/v agar) containing one of the indicator strains pre-grown in liquid LB to exponential phase (OD600 = 0.4–0.6) and incubated overnight at 37 °C (± 18 hours). *Bacillus subtilis* 168 and *Escherichia coli* ASD19 were used as indicator strains for antimicrobial activity and were cultured in LB media at 37 °C.

Supplementary information for Chapter 6

Table S1. Genomic features of *Pseudonocardia* sp. ZF1

	Length (bp)	GC (%)	tRNAs	rRNAs
Contig 1	5,662,857	74.1	52	12
Contig 2	493,868	71.9		
Contig 3	25,576	70.4		

Table S2. Overview of BGCs predicted by antiSMASH 217 and analysis of homology using cblaster.

Proto-cluster	Type	Most similar known cluster	Top hit cblaster	Average identity
Contig 1				
1.1	NAPAA	CC-1065 (6%)	*Pseudonocardia pini* ICBG1122	87%
1.2	RiPP-like	-	*Pseudonocardia* sp. ICBG1034	99%
1.3	Redox-cofactor	Lankacidin C (13%)	*Pseudonocardia* sp. ICBG601	94%
1.4	NAPAA	Streptobacin (11%)	*Pseudonocardia antarctica* DSM 44749	99%
1.5	Terpene	SF2575 (6%)	*Pseudonocardia alni* DSM 44104	99%
1.6A*	NRPS	Scabichelin (60%)	*Pseudonocardia autotrophica* DSM 43083	100%
1.6B	NAPAA	Stenothricin (18%)	*Pseudonocardia alni* DSM 44104	99%
1.7	Terpene	Isorenieratene (28%)	*Pseudonocardia alni* DSM 44104	98%
1.8	RiPP-like	-	*Pseudonocardia* sp. ICBG1034	98%
1.9	Ranthipeptide	-	*Pseudonocardia autotrophica* DSM 43083	99%
1.10	Ectoine	Ectoine 100%	*Pseudonocardia* sp. AL041005-10	99%
Contig 2				
2.1*	T1PKS	Aculeximycin (41%)	No hits	-
2.2*	NAPAA	-	*Pseudonocardia autotrophica* NBRC 12743 plasmid pPA12743CP DNA	97%
2.3*	T1PKS	Coelimycin P1 (29%)	*Frankia* sp. BMG5.36	27%
2.4*	NRPS-like, NRPS	Paulomycin (5%)	*Pseudonocardia spinosispora* DSM 44797	49%

*Search was limited to only a small subset of genes of the predicted BGC as a search with the whole predicted cluster did not provide any hits

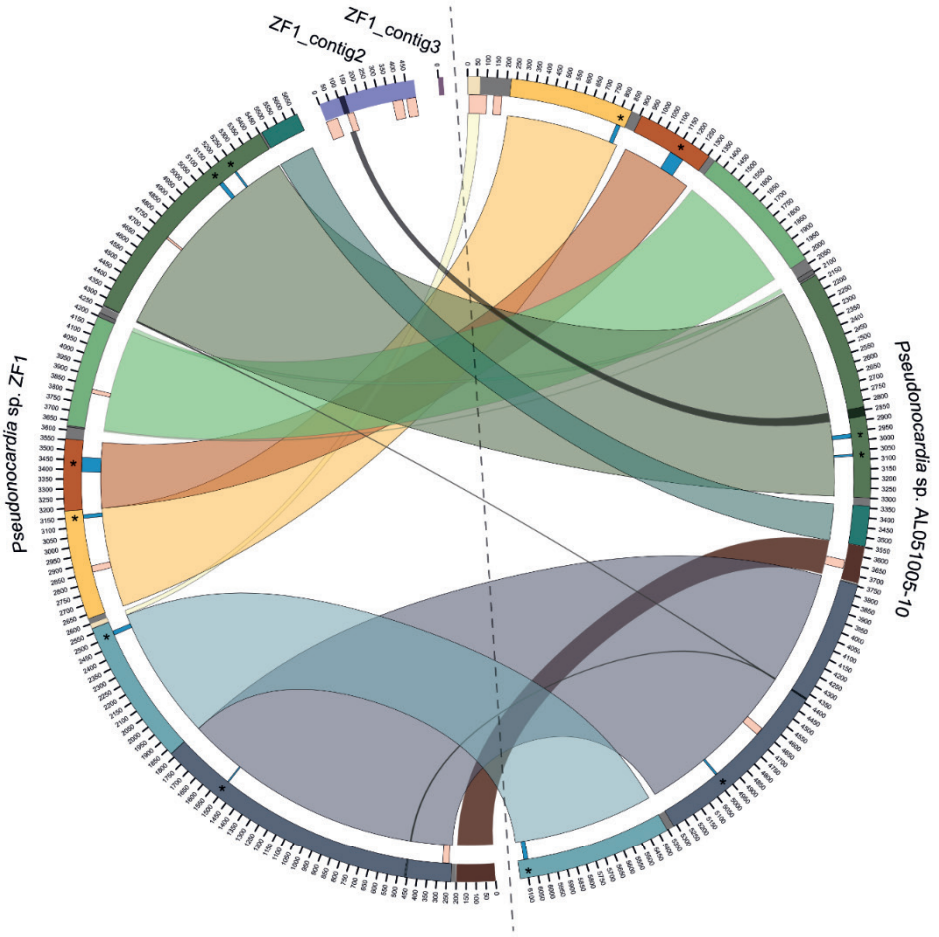


Figure S1. Synteny plot comparing *Pseudonocardia* sp. ZF1 to *Pseudonocardia* sp. AL41005-10. Genome sequences were compared by pairwise BLASTP and collinearity calculation method. Similarly coloured areas represent high similarity regions. The location of predicted BGCs by antiSMASH v6.0 are indicated in pink. Shared BGCs are indicated in blue and with an asterisk

7

Summarising general discussion

The phylum Actinobacteria represents one of the most diverse groups of bacteria in nature. Its members show a remarkable range of morphologies and are widely distributed across both terrestrial and aquatic ecosystems [2,49,50]. Within these different environments, they can be found as free-living organisms and as part of the microbiomes of a wide variety of eukaryotes, such as plants, insects, and mammals [42,43]. This remarkable ecological diversity is reflected in their metabolic potential as Actinobacteria are extremely versatile producers of chemically diverse metabolites that mediate important ecological functions [2,42]. Ectoines protect against osmotic stress [3], siderophores facilitate the uptake of iron [5], pigments have antioxidant activity [66], volatiles play a role in attracting organisms to promote spore dispersal [6], and gamma-butyrolactones function as regulatory signals [4]. Additionally, many metabolites exhibit antimicrobial activity and serve as biological weaponry to outcompete other organisms [7,8].

Due to their chemical and biological diversity, numerous actinobacterial natural products have been the starting point for drug discovery and have found use in human and veterinary medicine, agriculture, and biotechnology [9-11]. They include most of the clinically used antibiotics along with numerous chemotherapeutics, immunosuppressants, chelating agents, herbicides, and hydrolytic enzymes [9-11]. Importantly, the natural products we have discovered so far only present a small fraction of the metabolites that Actinobacteria can produce [27]. The genomes of many Actinobacteria are full of uncharacterised biosynthetic gene clusters (BGCs) that encode the cellular machinery required to produce specialised metabolites [10,16,19,20,241]. However, many BGCs are not expressed under laboratory conditions preventing the discovery of their cognate products [29,93,267]. Additionally, natural product-based research suffers from low return on investments due to the rediscovery of known molecules (known as replication) [239,240]. This leads to the question: how do we more efficiently exploit Actinobacteria as resource of chemical diversity?

In natural product research, the link between the ecological versatility of Actinobacteria and their extraordinary metabolic potential is often overlooked. Yet, the remarkable diversity of Actinobacteria is the result of millions of years of evolution during which ecological forces have shaped their specialised metabolite repertoire [38,102,104,105,109]. Within their environment, abiotic and biotic stresses generate selective pressure, driving horizontal gene transfer (HGT) and shaping the distribution and diversity of BGCs among bacteria. Moreover, just as ecological forces shape biosynthetic potential, they have shaped the regulation of specialised metabolism [34,42]. It is therefore likely that expression of many BGCs is tightly linked to specific challenges in the complex environment of Actinobacteria [32,34].

Understanding the ecological context in which Actinobacteria live can guide us towards gifted Actinobacteria in the search for chemical novelty and help us to identify the signals that activate

their specialised metabolism to unlock the full biosynthetic potential of Actinobacteria in the laboratory (Chapter 2) [42]. Importantly, specialised metabolites can also provide advantage to the many multicellular organisms that host Actinobacteria [45-47]. Understanding the ecological role of specialised metabolites within the microbial communities of plants, insects, and animals, may therefore guide us towards the use of antibiotic-producing Actinobacteria as probiotics in agriculture or human health.

In this thesis therefore different ecological approaches are used, aimed at accessing the full potential of Actinobacteria. We explored a unique environment for Actinobacteria and investigated their taxonomic and metabolic diversity (Chapter 3), we analysed the effect of human stress hormones on the antibiotic production of *Streptomyces (*Chapter 4), discovered the ubiquitous catechol moiety as elicitor of siderophore and angucycline production (Chapter 4 and 5), and studied zebrafish-associated Actinobacteria to provide a first step towards using the zebrafish as an *in vivo* model to explore the bioactive and functional potential of Actinobacteria within the animal microbiome (Chapter 6).

Exploring new environments: ancient microbiomes

The extraordinary diversity of Actinobacteria and their specialised metabolites is the result of millions of years of evolution. Yet, a huge proportion of the diversity of the Actinobacteria remains uncharacterised [27,41,200,221]. It is therefore likely that exploration of more diverse environments and taxa will provide us with new evolutionary insights, metabolic pathways, and chemistry. Additionally, much remains to be discovered about the evolutionary events that have shaped the specialised metabolite repertoire of Actinobacteria.

We were presented with a once-in-a-lifetime opportunity to explore a unique environment for ancient Actinobacteria that could potentially provide evolutionary insights: the ancient microbial community of an exceptionally well-preserved mammoth, estimated to be 28,000 years old [207,208]. In **Chapter 3**, a faecal sample was extracted from the gastrointestinal tracts of this extraordinary specimen with the aim to isolate filamentous Actinobacteria. Six strains belonging to five genera, namely *Micromonospora*, *Oerskovia*, *Saccharopolyspora, Sanguibacte*r, and *Streptomyces*, were recovered and their morphology, taxonomic profile, and biosynthetic potential was analysed. This revealed significant phylogenetic distance between the isolates and current known strains and much uncharacterised potential, yielding unexplored genomic information that is not yet present in current databases.

The isolation of Actinobacteria from ancient samples represents an opportunity to study evolutionary events, for example to identify ancient HGT occurrences [42,104]. When HGT leads

to acquisition of BGCs with similar functions, there may be strong evolutionary pressure for BGC loss. For example, analysis of the evolutionary history of two siderophore biosynthetic pathways in *Salinispora* revealed that acquisition of one pathway correlated to the loss of the other [109]. Such events may happen at large scale and could mask a large proportion of historical HGT events [105]. In this study, the major phylogenetic differences that were observed between the isolates and current strains, as well as the limited number of isolated strains, makes it hard to evaluate evolutionary differences on a genomic level. Yet, many ancient ecosystems await further exploration, as still a significant part of the biosphere contains permafrost, representing an important archive of evolutionary information and possibly lost biosynthetic potential.

The large phylogenetic distance between the mammoth isolates and current known strains and the high percentage of uncharacterised biosynthetic potential, even in the well-studied genus *Streptomyces*, emphasize that much of the diversity of Actinobacteria remains to be captured. Although sampling unique environments has demonstrated value to uncover rare taxa and novel chemistry [41,325-327], more guidance is needed to target specific environments. Many genes have a strong biogeographic signal, i.e. they are specific to a single habitat, and several studies suggest a connection between biogeography and biosynthetic diversity [178,328-330]. More broad scale characterisation of the microbial and biosynthetic diversity in different ecological niches is needed to point us towards specific locations or environments that are most likely to yield novel species and chemistry. Additionally, the majority of all species cannot be cultivated in the laboratory, stressing the need for new culture techniques to enable growth of these uncultured organisms. Co-cultivation, use of specific growth factors (e.g. siderophores), and the *in situ* diffusion chamber iChip have been used successfully to grow uncultured bacteria, and have even led to the discovery of the new antibiotic teixobactin [91,331,332]. Increasing our understanding of the ecological context in which (Actino)bacteria live can provide insights on how to bring nature to the laboratory and increase the success of isolation and screening campaigns.

While exploration of new environments remains an important strategy to find novel species and potentially novel chemistry, we should not forget about the vast reservoir of uncharacterised BGCs within the genomes of Actinobacteria that have been isolated up to this date. Even in the extensively studied *Streptomyces coelicolor* still ~40% of its BGCs remain uncharacterised. This is largely due to the fact that we lack the understanding that is required to activate BGC expression in the laboratory. To tackle this bottleneck in natural product research, more focus should be directed towards the elucidation of the regulatory mechanisms governing BGC expression and translating these insights into solutions to activate antibiotic production.

Animal stress hormones as elicitors of siderophore production

The difficulty to unlock the full biosynthetic potential of Actinobacteria in the laboratory, can be attributed to our limited knowledge of what activates their specialised metabolism in nature. In the laboratory, Actinobacteria are typically grown in isolation under stable conditions and with an abundance of nutrients. This does not properly mimic their competitive and rapidly changing environment in nature and likely explains why so many BGCs are not expressed during laboratory cultivation. Various studies have therefore focused on simulating naturally occurring conditions in the laboratory, for example by varying growth conditions and co-cultivation experiments, revealing a major impact of nutrient availability, pH, and microbial interactions on the level and timing of specialised metabolism [37,44,138,252,268].

Within their natural environments, Actinobacteria interact with different eukaryotic organisms whose signalling molecules may play a role in the activation of specialised metabolite production [2,42]. Indeed, plant stress hormones, such as jasmonates and salicylic acid, can increase the antibiotic activity of endophytic *Streptomyces* [44]. These hormones and other exudates are produced by plants under pathogenic stress and might represent a 'cry for help': their release may activate the production of bioactive metabolites by members of the plants microbiome and counteracts the pathogenic attack [2,44]. Similarly, animal hormones also influence bacteria, as exemplified by the human opioid dynorphin that stimulates production of pyocyanin in *Pseudomonas aeruginosa* [242]. In **Chapter 4**, we therefore explored the impact of the human catecholamine stress hormone adrenaline on the specialised metabolism of *Streptomyces*. Previous studies showed that catecholamines influence bacterial growth [243,244], biofilm formation [245], and horizontal gene transfer [246]. Our data show that catecholamines elicit siderophore production in *Streptomyces*, resulting in reduced growth of *B. subtilis*. To the best of our knowledge, elicitation of siderophore production by animal- and plant-associated molecules has not been reported in *Streptomyces* before.

We discovered that the increase in siderophore production is related to the iron chelating properties of the catechol moiety, which results in reduced iron availability. Although siderophores are not of major interest in terms of drug discovery, they play an important role in bacterial community dynamics [5]. In cocultivation experiments, many of the observed changes in bioactivity and metabolome production have been attributed to changes in iron availability and siderophore production [35,95,251,252]. It would therefore be interesting to explore the impact of plant- and animal-associated catechol compounds on the growth and metabolite production of Actinobacteria within a microbial community. Additionally, bacterial production of siderophores can promote plant growth and protect plants against fungal infection [257,333-335]. For example, bacteria producing siderophores that cannot be used by plant pathogens are linked to disease suppression in tomato plants *in vivo* [333,334]. It would be interesting to test whether *Streptomyces*

also respond to plant-associated catechol compounds (e.g., catechin and L-dopa) in soil and whether this affects plant health. Such experiments are important to provide insight into the ecological relevance of the response of *Streptomyces* to the catechol moiety.

Chemical elicitors in combination with omics-techniques

In **Chapter 5**, we further analysed the potential of catechol as elicitor by testing its effect on some of the well-characterised Actinobacteria in our collection. We showed that catechol promotes the expression of a BGC that produces different (novel) angucycline glycosides. Importantly, expression of catechol-degrading enzymes counteracted the eliciting effect of catechol, highlighting the importance of the catechol moiety in the response.

This study illustrates the power of chemical elicitors in combination with omics-techniques for the identification of bioactive metabolites, which can then be linked to their BGCs. Identifying the metabolites that are responsible for a specific bioactivity in complex mixtures is difficult and time consuming. Additionally, many studies that focus on the identification of novel natural products do not identify the responsible BGC. Therefore, many BGCs remain cryptic, i.e. they have not yet been linked to a natural product [42]. The combination of elicitor screening with technologies such as at-line nanofractionation [272], LC-MS analysis, GNPS networking [220], and quantitative proteomics [249] offers a powerful workflow to prioritise bioactive metabolites for isolation and assist in 'decrypting' cryptic BGCs (Figure 1).

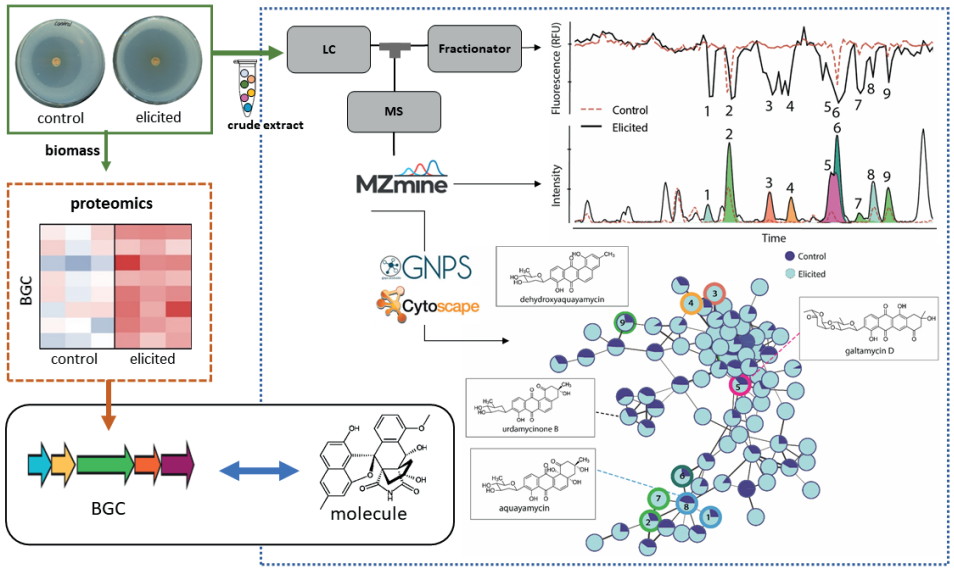


Figure 1. General workflow for drug discovery based on the work presented in Chapter 5

Lastly, the catechol moiety is ubiquitous in nature and part of a wide range of molecules associated with different kingdoms of life [5,247,269,270]. Potentially, focusing on conserved structural motives may represent a powerful approach to identify elicitors of specialised metabolism. The ubiquity of elicitors is also of interest from an ecological perspective as it may point towards conserved plant-microbe and animal-microbe interactions. Notably, many plant hormones show strong structural resemblance to human-associated molecules [336]. Therefore, the 'cry for help' that has been hypothesised to occur between plant and antibiotic-producing Actinobacteria, may also apply to human-microbe interactions. This interesting concept requires further investigation in an *in vivo* situation.

Towards a model to study host-Actinobacteria interactions *in vivo*.

The role of bioactive Actinobacteria within the microbiome of vertebrate organisms remains poorly characterised. As described above, recent studies show that plant and human stress hormones impact the bioactivity of Actinobacteria, pointing towards a 'cry for help' from host to microbe upon stress (**Chapter 4**) [44,301]. Models are needed to study whether such interactions occur *in vivo* and whether Actinobacteria can provide protection against invading pathogens.

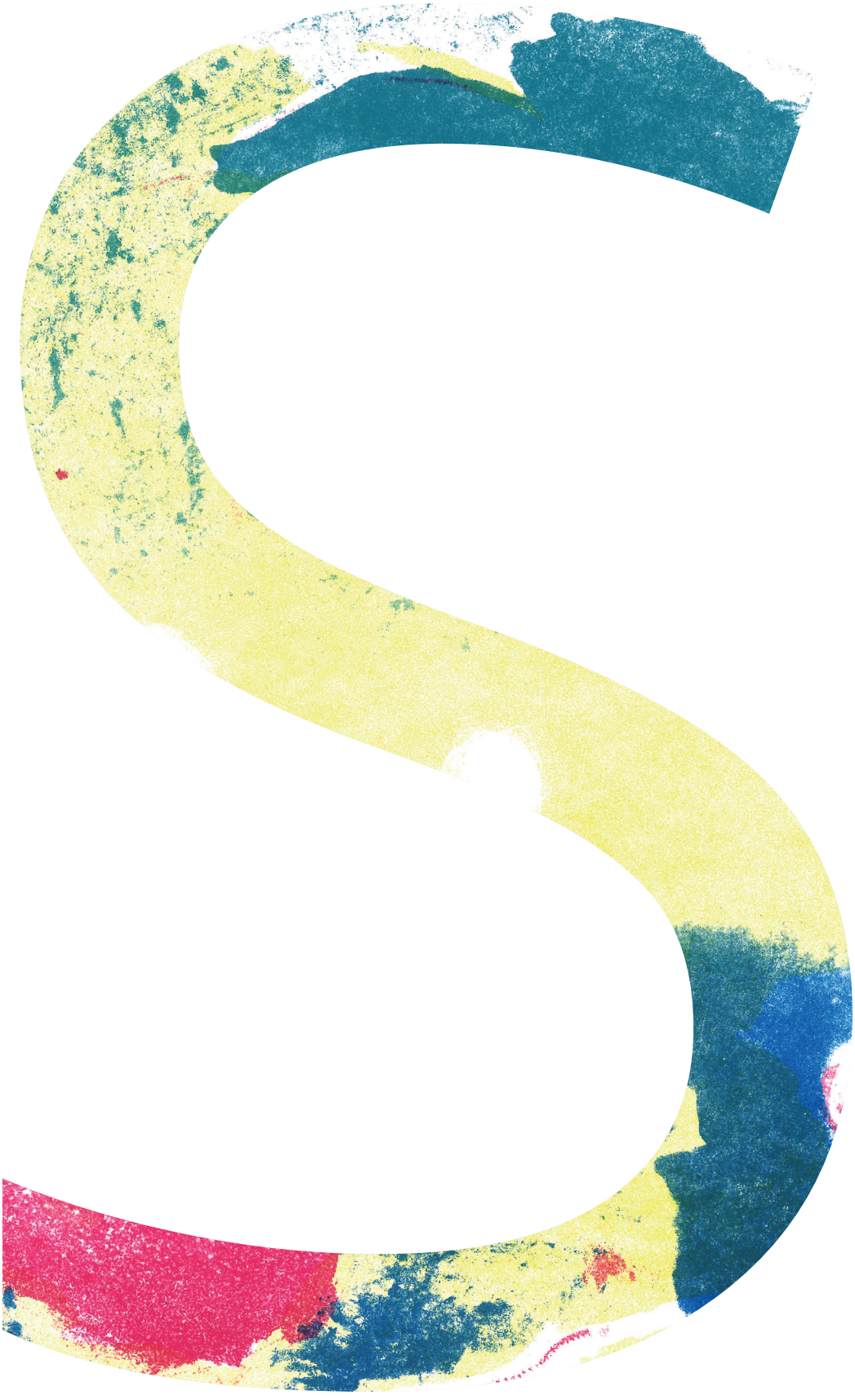
Zebrafish larvae are a relevant model system to characterise the role of the microbiome and host-microbe interactions in health and disease [302-305]. In **Chapter 6**, we analysed the gut microbiome of adult zebrafish and isolated Actinobacteria from zebrafish larvae, including a bioactive *Pseudonocardia* species. Although further research is required to identify the bioactive metabolites produced by the *Pseudonocardia* isolate and study its functional role *in vivo*, our data provides a first step towards the use of zebrafish as a model to study antibiotic-producing Actinobacteria within the animal microbiome. As a next step, the bioactive metabolites produced by *Pseudonocardia* sp. ZF1 and its cognate BGC should be identified. This will allow us to analyse whether the BGC is expressed *in vivo* and whether we can detect the molecule in the gut metabolome of zebrafish. Additionally, we aim to reintroduce *Pseudonocardia* sp. ZF1 into the microbiome and on the eggs of zebrafish to test whether the strain can aid in the protection against (fish) pathogens. If we succeed in the steps described above, the 'cry for help' hypothesis may be tested by exposing zebrafish to stressful conditions, such as pathogen infection, and analysing the presence of *Pseudonocardia*-associated metabolites. Likewise, also other antibiotic producing Actinobacteria may be tested to explore their potential as probiotics.

Outlook

As nature's medicine makers, Actinobacteria are welcome guests to the microbiomes of eukaryotic organisms, offering protection against invading pathogens. Similar to the human use of bacteria for antibiotics, many multicellular organisms deploy Actinobacteria and their

bioactive specialised metabolites for protection against invading pathogens. This interaction has inspired research into the use of Actinobacteria as probiotics, for example to protect crops against fungal infection. With the alarming increase in antimicrobial resistance, such approaches will become more and more important, also in the clinic and animal husbandry. Increasing our knowledge on the ecological role of Actinobacteria and their specialised metabolites in their native microbial communities and environments is essential to further explore the use of Actinobacteria as probiotics in human health care and agriculture.

Another intriguing approach to use antibiotic-producing bacteria to fight infection has been introduced by the *cry for help* hypothesis, which proposes that antibiotic production by Actinobacteria is only activated by the host under specific stressful conditions. Indeed, our work and that of others has revealed a role for plant- and animal-associated stress hormones in the activation of specialised metabolism in *Streptomyces*. An important matter to solve in the future is to identify the signal transduction pathways by which plant- and animal-associated signals (such as jasmonic acid and adrenaline) impact specialised metabolism, from internalization through metabolism to the activation of BGCs. Understanding the underlying mechanisms is of great interest as they may offer new tools to characterise the vast reservoir of cryptic BGCs and find novel antibiotics. Ultimately, these insights may be harnessed to engineer probiotic strains that only produce antibiotics on demand.



Nederlandse samenvatting

Ofschoon ze met het blote oog niet te zien zijn, zijn bacteriën essentieel voor al het leven op aarde; ze spelen een belangrijke rol in de stikstofkringloop, breken organisch materiaal af en zijn onderdeel van het microbioom van planten en dieren, waar ze een rol spelen in vitamineproductie, de voedselvertering en de bescherming tegen pathogenen. Feitelijk zijn bacteriën zeer getalenteerde scheikundigen: ze produceren een onvoorstelbare diversiteit aan bioactieve metabolieten en enzymen die niet alleen nuttig zijn voor de bacterie zelf, maar ook voor onder andere planten en dieren. Veel bacteriële metabolieten worden door de mens gebruikt als geneesmiddelen, zoals antibiotica, immunosuppressiva en antikanker medicijnen. Ook hebben veel bacteriële metabolieten en enzymen toepassing gevonden in de landbouw en biotechnologie.

Een groot deel van de bioactieve metabolieten die als geneesmiddel gebruikt worden, zijn afkomstig van leden van het fylum Actinobacteriën. Dit fylum vertegenwoordigt één van de meest diverse groepen van bacteriën, die een enorme verscheidenheid aan morfologieën aan de dag leggen en te vinden zijn in de meest uiteenlopende habitats. Binnen het fylum zijn het vooral de filamenteus groeiende Actinobacteriën zoals *Streptomyces, Micromonospora,* en *Amycolatopsis* die veel verschillende bioactieve metabolieten produceren. Dit heeft mogelijk te maken met de filamenteuze groei van deze bacteriën. Na ontkieming van sporen, worden lange draadvormige hyfen gevormd die zich vertakken en zo een complex netwerk van mycelium vormen. Door deze manier van groeien kunnen deze bacteriën zich niet zomaar verplaatsen bij gevaar. Door het produceren van veel verschillende bioactieve metabolieten kunnen ze toch overleven, ectoines beschermen bijvoorbeeld tegen osmotische stress, sideroforen vergemakkelijken de opname van ijzer, en pigmenten werken als antioxidanten en beschermen tegen radiatie. Nog belangrijker, met hun antimicrobiële activiteit dienen veel metabolieten als biologische wapens in de competitie met andere micro-organismen. Bioactieve metabolieten spelen dus een belangrijke rol om te overleven en we verwijzen daarom vaak naar deze moleculen als secundaire of gespecialiseerde metabolieten.

Deze gespecialiseerde metabolieten zijn door hun bioactiviteit niet alleen belangrijk voor de bacterie zelf maar ook voor ons. De ontdekking van hun antimicrobiële werking door onder andere Alexander Fleming en Selman Waksman bracht een revolutie teweeg in de behandeling van infectieziekten. In het zogenaamde “gouden tijdperk van de ontdekking van antibiotica” van ± 1940 – 1960 werden door middel van het “high-throughput screenen” van bacteriën en schimmels meer dan 1000 microbiële metabolieten met antibacteriële of schimmelwerende activiteit geïdentificeerd, waaronder veel van de antibiotica die vandaag de dag nog steeds in de kliniek wordt gebruikt. Helaas heeft het gebruik van antibiotica door de

mens over de jaren heen geleid tot het ontstaan van multiresistentie in veel ziekteverwekkers. Daardoor vormen bacteriële infecties opnieuw een grote bedreiging voor de gezondheid en is er urgente behoefte aan nieuwe antibiotica. Echter, het succes van “high-throughput screenen” is drastisch afgenomen, voornamelijk door “replicatie”, oftewel het steeds maar weer herontdekken van al bekende moleculen. In de afgelopen 40 jaar zijn slechts drie nieuwe antibioticaklassen ontdekt en veel farmaceutische bedrijven hebben het onderzoek naar antibiotica stopgezet.

Toch weten we sinds de ontwikkeling van de genoom “sequencing” technieken begin 2000 dat bacteriën nog veel meer moleculen kunnen maken dan we tot nu toe gevonden hebben. De genen die coderen voor de eiwitten die nodig zijn voor de biosynthese van metabolieten liggen namelijk geclusterd op het genoom en vormen herkenbare biosynthetische genclusters (BGCs). Door genoom sequentiebepaling en bioinformaticavoorspellingen werd het mogelijk om deze BGCs op het genoom te identificeren en wat bleek: in de genomen van bacteriën die al tientallen jaren uitvoerig bestudeerd waren, werden talrijke niet gekarakteriseerde BGCs gevonden. Zo was van *Streptomyces coelicolor* A3(2), het modelorganisme voor antibioticaproductie, bekend dat het vier antibiotica kon produceren. Toen het genoom van stam geanalyseerd werd, kwamen er nog eens 18 andere BGCs aan het licht en hernieuwde inspanningen leidden vervolgens tot de ontdekking van een vijfde antibioticum.

Vandaag de dag wordt geschat dat we slechts zo’n 3% van alle bestaande BGCs hebben gekarakteriseerd. Een belangrijke reden hiervoor is dat we nog maar een klein deel van de bacteriële wereld onderzocht hebben, en in een beperkt aantal ecosystemen. Daarnaast komt een groot deel van de BGCs niet tot expressie wanneer de bacteriën opgekweekt worden in het laboratorium. We verwijzen naar deze BGCs als “stil” of “slapend”: ze zijn wel aanwezig, maar de bijbehorende metabolieten worden onder laboratoriumomstandigheden niet geproduceerd. Dat zoveel BGCs niet tot expressie komen in het laboratorium geeft aan dat we nog niet begrijpen onder welke omstandigheden bacteriën bepaalde gespecialiseerde metabolieten maken.

Om hier meer inzicht in te krijgen, moeten we de ecologische context waarin Actinobacteriën gespecialiseerde metabolieten maken beter begrijpen (**Hoofdstuk 2**). Want hoewel ze in het laboratorium niet het hele arsenaal aan moleculen produceren, gebruiken zij die metabolieten zeer waarschijnlijk wel in hun natuurlijke omgeving. Bovendien heeft de ecologie een belangrijke rol gespeeld bij de evolutie van de diversiteit en verspreiding van BGCs en speelt omgeving dus ook een rol bij het zoeken naar de meest getalenteerde producenten van bioactieve metabolieten. Ten slotte kan inzicht in de ecologische rol van gespecialiseerde

metabolieten binnen de microbiomen van planten, insecten en dieren, ons ook leiden naar het gebruik van antibioticaproducerende Actinobacteriën als probiotica in bevordering van de landbouw of de menselijke gezondheid.

Als onderdeel van het Netherlands Centre for One Health (NCOH) onderzoeksthema '*tackling antimicrobial resistance'* richt dit proefschrift zich daarom op verschillende ecologische strategieën voor het vinden van nieuwe Actinobacteriën, waaronder *Streptomyces.*

Verkennen van nieuwe omgevingen: een oeroud milieu

De buitengewone diversiteit van de Actinobacteriën en hun gespecialiseerde metabolieten is het resultaat van miljoenen jaren evolutie. Toch is een groot deel van deze diversiteit nog steeds niet gekarakteriseerd, onder andere omdat maar een beperkt aantal ecosystemen bestudeerd is. Dit terwijl het zeer waarschijnlijk is dat ecologische factoren binnen een habitat een grote impact hebben op de diversiteit aan BGCs die in een bacterie voorkomt. De afgelopen jaren is de aandacht daarom verschoven naar meer diverse omgevingen, waaronder extreme habitats (zoals diepzee-sedimenten en woestijnen) en de microbiomen van planten, dieren en mensen. Dit heeft geleid tot de ontdekking van veel interessante bacteriën en nieuwe moleculen. Het verkennen van nieuwe habitats is daarom een interessante strategie die nieuwe evolutionaire inzichten, metabolische routes en chemie kan opleveren.

Wij kregen de uitzonderlijke mogelijkheid om naar Actinobacteriën te zoeken in het darmkanaal van een goed bewaarde mammoet van ruim 28.000 jaar oud. De oudheid van de mammoet gaf namelijk de mogelijkheid om oeroude Actinobacteriën te bestuderen, iets wat nog maar heel weinig gedaan is. **Hoofdstuk 3** beschrijft de extractie van een fecesmonster uit de darm van de mammoet en de isolatie van verschillende soorten Actinobacteriën uit dit monster. De geïsoleerde bacteriestammen werden gekweekt en de genomen geanalyseerd voor taxonomische vergelijkingen en om het biosynthetisch potentieel van de bacteriën te voorspellen. De mammoetbacteriën produceren een aantal bekende bioactieve moleculen, maar hun genomen zitten ook vol met onbekende BGCs waarvan de bijbehorende metabolieten nog ontdekt moeten worden. Uit de taxonomische analyse kwam verder naar voren dat de mammoetstammen grote verschillen vertonen ten opzichte van de bacteriën die tot nu toe zijn onderzocht. Dit komt deels doordat er van sommige bacteriesoorten nog maar relatief weinig genoomsequenties beschikbaar zijn, maar mogelijk speelt de oudheid van de stammen en/of de bijzondere omgeving waar ze vandaan komen hier ook een rol in. Wat de rol van de habitat en de oudheid van de stammen is bij de grote verschillen tussen de mammoetbacteriën en al bekende bacteriën is een vraag die nog beantwoord moet worden. Het is wel duidelijk dat het isoleren en genoomanalyses van de Actinobacteriën uit het darmkanaal van een mammoet waardevolle informatie heeft opgeleverd.

Dierlijke stresshormonen stimuleren de productie van sideroforen

Hoewel de exploratie van nieuwe omgevingen een belangrijke strategie blijft om nieuwe bacteriën en mogelijk nieuwe chemische stoffen te vinden, moeten we niet vergeten dat de genomen van Actinobacteriën die tot op heden geïsoleerd zijn een enorm reservoir aan onbekende BGCs bevatten. Dit is grotendeels te wijten aan het feit dat veel BGCs niet tot expressie komen in het laboratorium. Om dit knelpunt in de zoektocht naar nieuwe bioactieve metabolieten aan te pakken, moeten we begrijpen hoe en wanneer de productie van gespecialiseerde metabolieten geactiveerd wordt. Deze inzichten kunnen we dan vertalen naar oplossingen om antibioticaproductie in het laboratorium te activeren en daarmee hopelijk nieuwe bioactieve moleculen te vinden.

In het laboratorium worden bacteriën onder stabiele omstandigheden gekweekt, met een overvloed aan voedingsstoffen en in afwezigheid van concurrerende organismen. Dit is geen goede afspiegeling van de competitieve en snel veranderende natuurlijke omgeving waarin de bacteriën normaal leven. Verschillende studies hebben zich daarom toegespitst op het nabootsen van natuurlijk voorkomende omstandigheden in het laboratorium, bijvoorbeeld door het variëren van groeicondities en co-cultivatie experimenten (het samen groeien van meerdere bacteriën). Uit deze studies blijkt dat de beschikbaarheid van nutriënten, pH, en microbiële interacties grote invloed hebben op de productie van gespecialiseerde metabolieten.

In hun natuurlijke omgeving zijn Actinobacteriën onderdeel van de microbiomen van verschillende eukaryote organismen zoals planten en dieren. De afgelopen jaren is er daarom ook steeds meer aandacht voor de interactie tussen gastheer en bacterie en het effect daarvan op de productie van gespecialiseerde metabolieten. Zo stimuleren plant stresshormonen, zoals jasmonaten en salicylzuur, de antibiotica productie van streptomyceten en heeft het menselijke opioïde dynorfine een positief effect op de productie van pyocyanine in *Pseudomonas aeruginosa*. In **hoofdstuk 4** is onderzocht of het menselijke stresshormoon adrenaline als elicitor kan werken. Adrenaline behoort samen met dopamine en noradrenaline tot de catecholamine neurotransmitters. De naam catecholamine is afgeleid van de moleculaire structuur van deze groep hormonen, die bestaat uit een benzeenring met twee hydroxylgroepen (de catecholgroep). Het onderzoek in Hoofdstuk 4 laat zien dat adrenaline de groei en antibioticaproductie van *Streptomyces* beïnvloedt. Van de stammen die getest werden, produceerde *Streptomyces* sp. MBT42 veel meer antibiotica in aanwezigheid van adrenaline. We ontdekten dat MBT42 niet alleen op adrenaline reageerde, maar ook op de andere catecholamines, dopamine en noradrenaline, en op verschillende plant-geassocieerde metabolieten met een catecholgroep. De catecholunit bleek de sleutel te zijn voor het activeren van de productie van sideroforen in *Streptomyces*. Deze toename houdt vermoedelijk verband

met de ijzerbindende eigenschappen van de catecholgroep. Doordat deze ijzer bindt, is er minder van dit essentiële metaal beschikbaar voor *Streptomyces* en dit is een signaal om meer sideroforen te gaan maken. Interessant is dat de siderofoor die door MBT42 gemaakt wordt, de groei van *Bacillus subtilis* remt en daardoor een soort antibacteriële activiteit heeft.

De waarde van het combineren van elicitoren met "omics" technieken

Hoofdstuk 5 gaat dieper in op de potentie van catechol als elicitor door het effect ervan te testen op een aantal goed gekarakteriseerde Actinobacteriën. Hier sprong de respons van *Streptomyces* sp. MBT84 er uit: catechol stimuleerde de antibiotica activiteit en de productie van een geel pigment. Met behulp van verschillende zogenoemde "omics" technieken (genomics, proteomics en metabolomics) werd de respons verder geanalyseerd. Daarbij kwam naar voren dat catechol de expressie bevordert van een BGC dat verschillende (nieuwe) angucyclineglycosiden produceert. Deze bioactieve metabolieten zijn zeer bekend door hun antikanker activiteit. Sommige bacteriën kunnen catechol afbreken en daarom werd ook onderzocht wat het effect was van het tot expressie brengen van catechol-afbrekende enzymen in *Streptomyces* sp. MBT84. In aanwezigheid van deze enzymen bleek *Streptomyces* sp. MBT84 niet langer te reageren op catechol, wat het belang van de catecholstructuur in de respons benadrukt.

Deze studie illustreert de waarde van het gebruik van elicitoren in combinatie met "omics" technieken voor de identificatie van bioactieve metabolieten en het linken van deze metabolieten met hun BGCs. Het identificeren van de bioactieve metabolieten is vaak moeilijk en tijdrovend omdat bij extractie van metabolieten zeer complexe mengsels ontstaan van veel verschillende moleculen. Bovendien wordt in veel studies die zich richten op het vinden van nieuwe bioactieve metabolieten de verantwoordelijke BGC niet geïdentificeerd. Daarom blijven veel BGCs cryptisch, wat wil zeggen dat ze nog niet in verband zijn gebracht met een metaboliet. De combinatie van elicitorscreening met technologieën zoals "at-line nanofractionering", LC-MS-analyse, GNPS-netwerken en kwantitatieve proteomics biedt een krachtig werkproces om bioactieve metabolieten te prioriteren voor isolatie en te helpen bij het "ontcijferen" van cryptische BGCs

Op zoek naar een model om gastheer-Actinobacterie interacties *in vivo* te bestuderen.

De rol van bioactieve Actinobacteriën binnen het microbioom van dieren en mensen blijft slecht gekarakteriseerd. Daarnaast laten verschillende studies en de bevindingen in hoofdstuk 4 zien dat dierlijke stresshormonen het gespecialiseerde metabolisme van bacteriën kunnen beïnvloeden. Om meer inzicht te krijgen in de rol die antibioticaproducerende Actinobacteriën kunnen spelen in het dierlijke en/of menselijke microbioom en mogelijke gastheer-microbe interacties te kunnen onderzoeken, zijn *in vivo* modellen nodig.

Zebravislarven zijn een relevant modelsysteem om de rol van het microbioom en gastheer-microbe interacties in gezondheid en ziekte te karakteriseren. Over de antibioticaproducerende Actinobacteriën in het zebravismicrobioom is echter nog weinig bekend. In **Hoofdstuk 6** is daarom het darmmicrobioom van zebravissen geanalyseerd en zijn verschillende Actinobacteriën uit zebravislarven geïsoleerd, waaronder de bioactieve *Pseudonocardia* sp. ZF1. Door de genoomsequentie te bepalen werden veel ongekarakteriseerde BGCs geïdentificeerd, waarvan er een aantal op een plasmide lag. Verder onderzoek is nodig om de bioactieve metabolieten die deze *Pseudonocardia* produceert te identificeren. Dan kan geanalyseerd worden of deze metabolieten ook in de darm van de zebravis geproduceerd worden. Daarnaast zou het interessant zijn om te testen of de *Pseudonocardia* bescherming kan bieden tegen (vis)pathogenen. Hoewel er dus nog veel onderzocht moet worden, is met de isolatie van Actinobacteriën, waaronder de *Pseudonocardia*, een eerste stap naar het gebruik van de zebravis als *in vivo* model gezet.

Toekomstbeeld

Als medicijnmakers van de natuur, zijn Actinobacteriën welkome gasten in de microbiomen van planten en dieren waar ze bescherming bieden tegen binnendringende ziekteverwekkers. Vergelijkbaar met het menselijke gebruik van antibiotica, zetten veel eukaryoten Actinobacteriën en hun bioactieve metabolieten in als bescherming tegen binnendringende pathogenen. Mogelijk kunnen wij Actinobacteriën op een vergelijkbare manier gebruiken in de landbouw, veehouderij en de kliniek ter bescherming tegen infecties, bijvoorbeeld door Actinobacteriën die metabolieten produceren tegen plantenpathogenen toe te voegen aan het microbioom van onze gewassen of door antibioticaproducerende Actinobacteriën als probiotica te consumeren. Met de alarmerende toename van antimicrobiële resistentie zullen dergelijke benaderingen steeds belangrijker worden. Meer kennis over de ecologische rol van Actinobacteriën en hun gespecialiseerde metabolieten in hun natuurlijke omgevingen en in microbiomen is essentieel voor verder onderzoek naar het gebruik van Actinobacteriën als probiotica in de gezondheidszorg en de landbouw.

Een andere benadering van het gebruik van antibiotica-producerende bacteriën om infecties te bestrijden is geïntroduceerd door de “cry for help” hypothese, waarbij de antibioticaproductie van Actinobacteriën alleen door de gastheer wordt geactiveerd onder specifieke stressvolle omstandigheden. Uit ons werk en dat van anderen blijkt inderdaad dat plant- en diergeassocieerde stresshormonen een rol spelen bij de activering van het gespecialiseerde metabolisme in *Streptomyces*. Een belangrijke kwestie die in de toekomst moet worden opgelost is het identificeren van de signaaltransductieroutes waarlangs plant- en diergeassocieerde signalen (zoals jasmonzuur en adrenaline) het gespecialiseerde metabolisme beïnvloeden, van het transporteren van het signaal tot de activering van specifieke BGCs. Inzicht in de onderliggende mechanismen is van groot belang om nieuwe mogelijkheden te creëren om antibioticaproductie in het laboratorium te activeren en daarmee nieuwe moleculen te vinden. Uiteindelijk kunnen deze inzichten mogelijk benut worden om probiotische stammen te ontwikkelen die alleen antibiotica produceren in reactie op een infectie.

References

1 Cimermancic, P. *et al.* Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **158**, 412-421 (2014).

2 van der Meij, A., Worsley, S. F., Hutchings, M. I. & van Wezel, G. P. Chemical ecology of antibiotic production by actinomycetes. *FEMS Microbiol Rev* **41**, 392-416 (2017).

3 Sadeghi, A. *et al.* Diversity of the ectoines biosynthesis genes in the salt tolerant *Streptomyces* and evidence for inductive effect of ectoines on their accumulation. *Microbiol Res* **169**, 699-708 (2014).

4 Takano, E. Gamma-butyrolactones: *Streptomyces* signalling molecules regulating antibiotic production and differentiation. *Curr Opin Microbiol* **9**, 287-294 (2006).

5 Kramer, J., Ozkaya, O. & Kummerli, R. Bacterial siderophores in community and host interactions. *Nat Rev Microbiol* **18**, 152-163 (2020).

6 Becher, P. G. *et al.* Developmentally regulated volatiles geosmin and 2-methylisoborneol attract a soil arthropod to *Streptomyces* bacteria promoting spore dispersal. *Nat Microbiol* **5**, 821-829 (2020).

7 Wright, E. S. & Vetsigian, K. H. Inhibitory interactions promote frequent bistability among competing bacteria. *Nat Commun* **7**, 11274 (2016).

8 Abrudan, M. I. *et al.* Socially mediated induction and suppression of antibiosis during bacterial coexistence. *Proc Natl Acad Sci U S A* **112**, 11054-11059 (2015).

9 Berdy, J. Bioactive microbial metabolites. *J Antibiot (Tokyo)* **58**, 1-26 (2005).

10 Katz, L. & Baltz, R. H. Natural product discovery: past, present, and future. *J Ind Microbiol Biotechnol* **43**, 155-176 (2016).

11 Newman, D. J. & Cragg, G. M. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *J Nat Prod* **83**, 770-803 (2020).

12 Fleming, A. On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of *B. influenzae*. 1929. *Bull World Health Organ* **79**, 780-790 (2001).

13 Sakula, A. Selman Waksman (1888-1973), discoverer of streptomycin: a centenary review. *Br J Dis Chest* **82**, 23-31 (1988).

14 Berdy, J. Thoughts and facts about antibiotics: where we are now and where we are heading. *J Antibiot (Tokyo)* **65**, 385-395 (2012).

15 Murray, C. J. L. *et al.* Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *Lancet* **399**, 629-655 (2022).

16 Baltz, R. H. Renaissance in antibacterial discovery from actinomycetes. *Curr Opin Pharmacol* **8**, 557-563 (2008).

17 Cooper, M. A. & Shlaes, D. Fix the antibiotics pipeline. *Nature* **472**, 32 (2011).

18 Hutchings, M. I., Truman, A. W. & Wilkinson, B. Antibiotics: past, present and future. *Curr Opin Microbiol* **51**, 72-80 (2019).

19 Martin, M. F. & Liras, P. Organization and expression of genes involved in the biosynthesis of antibiotics and other secondary metabolites. *Annu Rev Microbiol* **43**, 173-206 (1989).

20 Bentley, S. D. *et al.* Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**, 141-147 (2002).

21 Ikeda, H. *et al.* Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat Biotechnol* **21**, 526-531 (2003).

22 Wright, L. F. & Hopwood, D. A. Identification of the antibiotic determined by the SCP1 plasmid of *Streptomyces coelicolor* A3(2). *J Gen Microbiol* **95**, 96-106 (1976).

23 Wright, L. F. & Hopwood, D. A. Actinorhodin is a chromosomally-determined antibiotic in *Streptomyces coelicolor* A3(2). *J Gen Microbiol* **96**, 289-297 (1976).

24 Feitelson, J. S., Malpartida, F. & Hopwood, D. A. Genetic and biochemical characterization of the red gene cluster of *Streptomyces coelicolor* A3(2). *J Gen Microbiol* **131**, 2431-2441 (1985).

25 Chong, P. P. *et al.* Physical identification of a chromosomal locus encoding biosynthetic genes for the lipopeptide calcium-dependent antibiotic (CDA) of *Streptomyces coelicolor* A3(2). *Microbiology (Reading)* **144 (Pt 1)**, 193-199 (1998).

26 Gomez-Escribano, J. P. *et al.* Structure and biosynthesis of the unusual polyketide alkaloid coelimycin P1, a metabolic product of the cpk gene cluster of *Streptomyces coelicolor* M145. *Chemical Science* **3**, 2716-2720 (2012).

27 Gavriilidou, A. *et al.* Compendium of specialized metabolite biosynthetic diversity encoded in bacterial genomes. *Nat Microbiol* **7**, 726-735 (2022).

28 Medema, M. H. & Fischbach, M. A. Computational approaches to natural product discovery. *Nat Chem Biol* **11**, 639-648 (2015).

29 Rutledge, P. J. & Challis, G. L. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat Rev Microbiol* **13**, 509-523 (2015).

30 Moon, K., Xu, F. & Seyedsayamdost, M. R. Cebulantin, a cryptic lanthipeptide antibiotic uncovered using bioactivity-coupled HiTES. *Angew Chem Int Ed Engl* **58**, 5973-5977 (2019).

31 Moon, K., Xu, F., Zhang, C. & Seyedsayamdost, M. R. Bioactivity-HiTES unveils cryptic antibiotics encoded in actinomycete bacteria. *ACS Chem Biol* **14**, 767-774 (2019).

32 van der Heul, H. U., Bilyk, B. L., McDowall, K. J., Seipke, R. F. & van Wezel, G. P. Regulation of antibiotic production in Actinobacteria: new perspectives from the post-genomic era. *Nat Prod Rep* **35**, 575-604 (2018).

33 Bibb, M. J. Regulation of secondary metabolism in streptomycetes. *Curr Opin Microbiol* **8**, 208-215 (2005).

34 Hoskisson, P. A. & Fernández-Martínez, L. T. Regulation of specialised metabolites in Actinobacteria - expanding the paradigms. *Environ Microbiol Rep* **10**, 231-238 (2018).

35 Traxler, M. F., Watrous, J. D., Alexandrov, T., Dorrestein, P. C. & Kolter, R. Interspecies interactions stimulate diversification of the *Streptomyces coelicolor* secreted metabolome. *mBio* **4 ,** e00459-13 (2013).

36 Rigali, S. *et al.* Feast or famine: the global regulator DasR links nutrient stress to antibiotic production by *Streptomyces*. *EMBO Rep* **9**, 670-675 (2008).

37 Zhu, H., Sandiford, S. K. & van Wezel, G. P. Triggers and cues that activate antibiotic production by actinomycetes. *J Ind Microbiol Biotechnol* **41**, 371-386 (2014).

38 Smanski, M. J., Schlatter, D. C. & Kinkel, L. L. Leveraging ecological theory to guide natural product discovery. *J Ind Microbiol Biotechnol* **43**, 115-128 (2016).

39 NCOH. *About Netherlands Centre for One Health*, <https://ncoh.nl/about>

40 Chevrette, M. G. & Handelsman, J. Needles in haystacks: reevaluating old paradigms for the discovery of bacterial secondary metabolites. *Nat Prod Rep* **38**, 2083-2099 (2021).

41 Sayed, A. M. *et al.* Extreme environments: microbiology leading to specialized metabolites. *J Appl Microbiol* **128**, 630-657 (2020).

42 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).

43 Chevrette, M. G. *et al.* The antimicrobial potential of *Streptomyces* from insect microbiomes. *Nat Commun* **10**, 516 (2019).

44 van der Meij, A. *et al.* Inter- and intracellular colonization of *Arabidopsis* roots by endophytic Actinobacteria and the impact of plant hormones on their antimicrobial activity. *Antonie Van Leeuwenhoek* **111**, 679-690 (2018).

45 Liu, Y. *et al.* Deciphering microbial landscapes of fish eggs to mitigate emerging diseases. *ISME J* **8**, 2002-2014 (2014).

46 Scott, J. J. *et al.* Bacterial protection of beetle-fungus mutualism. *Science* **322**, 63 (2008).

47 Mendes, R. *et al.* Deciphering the rhizosphere microbiome for disease-suppressive bacteria. *Science* **332**, 1097-1100 (2011).

48 Sit, C. S. *et al.* Variable genetic architectures produce virtually identical molecules in bacterial symbionts of fungus-growing ants. *Proc Natl Acad Sci U S A* **112**, 13150-13154 (2015).

49 Whitman, W. B. *et al. Bergey's Manual of Systematic Bacteriology: Volume 5: The Actinobacteria*. (Springer Science & Business Media, 2012).

50 Barka, E. A. *et al.* Taxonomy, physiology, and natural products of Actinobacteria. *Microbiol Mol Biol Rev* **80**, 1-43 (2016).

51 Hopwood, D. A. *Streptomyces in nature and medicine: the antibiotic makers*. (Oxford University Press, 2007).

52 Vrancken, K. & Anne, J. Secretory production of recombinant proteins by *Streptomyces*. *Future Microbiol* **4**, 181-188 (2009).

53 Newman, D. J. & Cragg, G. M. Natural products as sources of new drugs over the last 25 years. *J Nat Prod* **70**, 461-477 (2007).

54 Davies, J. & Davies, D. Origins and evolution of antibiotic resistance. *Microbiol Mol Biol Rev* **74**, 417-433 (2010).

55 Payne, D. J., Gwynn, M. N., Holmes, D. J. & Pompliano, D. L. Drugs for bad bugs: confronting the challenges of antibacterial discovery. *Nat Rev Drug Discov* **6**, 29-40 (2007).

56 Baltz, R. H. Antimicrobials from actinomycetes: back to the future. *Microbe* **2**, 125-131 (2007).

57 van den Berg, M. A. *et al.* Genome sequencing and analysis of the filamentous fungus *Penicillium chrysogenum*. *Nat Biotechnol* **26**, 1161-1168 (2008).

58 Bode, H. B., Bethe, B., Hofs, R. & Zeeck, A. Big effects from small changes: possible ways to explore nature's chemical diversity. *Chembiochem* **3**, 619-627 (2002).

59 Romano, S., Jackson, S. A., Patry, S. & Dobson, A. D. W. Extending the "One Strain Many Compounds" (OSMAC) principle to marine microorganisms. *Mar Drugs* **16** (2018).

60 Wu, C. *et al.* Lugdunomycin, an angucycline-derived molecule with unprecedented chemical architecture. *Angew Chem Int Ed Engl* **58**, 2809-2814 (2019).

61 Wright, E. S. & Vetsigian, K. H. Inhibitory interactions promote frequent bistability among competing bacteria. *Nat Commun* **7**, 11274 (2016).

62 Hibbing, M. E., Fuqua, C., Parsek, M. R. & Peterson, S. B. Bacterial competition: surviving and thriving in the microbial jungle. *Nat Rev Microbiol* **8**, 15-25 (2010).

63 Traxler, M. F. & Kolter, R. Natural products in soil microbe interactions and evolution. *Nat Prod Rep* **32**, 956-970 (2015).

64 Traxler, M. F., Watrous, J. D., Alexandrov, T., Dorrestein, P. C. & Kolter, R. Interspecies interactions stimulate diversification of the *Streptomyces coelicolor* secreted metabolome. *MBio* **4,** e00459-13 (2013).

65 Guerinot, M. L. Microbial iron transport. *Annu Rev Microbiol* **48**, 743-772 (1994).

66 Li, C., Ji, C. & Tang, B. Purification, characterisation and biological activity of melanin from Streptomyces sp. *FEMS Microbiol Lett* **365** (2018).

67 Sadeghi, A. *et al.* Diversity of the ectoines biosynthesis genes in the salt tolerant *Streptomyces* and evidence for inductive effect of ectoines on their accumulation. *Microbiol Res* **169**, 699-708 (2014).

68 Kroiss, J. *et al.* Symbiotic streptomycetes provide antibiotic combination prophylaxis for wasp offspring. *Nat Chem Biol* **6**, 261-263 (2010).

69 Raaijmakers, J. M. & Mazzola, M. Diversity and natural functions of antibiotics produced by beneficial and plant pathogenic bacteria. *Annu Rev Phytopathol* **50**, 403-424 (2012).

70 Doroghazi, J. R. & Metcalf, W. W. Comparative genomics of actinomycetes with a focus on natural product biosynthetic genes. *BMC Genomics* **14**, 611 (2013).

71 Cimermancic, P. *et al.* Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **158**, 412-421 (2014).

72 Sriswasdi, S., Yang, C. C. & Iwasaki, W. Generalist species drive microbial dispersion and evolution. *Nat Commun* **8**, 1162 (2017).

73 Shimkets, L. J. in *Bacterial genomes: physical structure and analysis* (eds F. J. de Bruijn, J. R. Lupski, & G. M. Weinstock), 5-11 (Chapman & Hall, 1998).

74 Fraser, C. M. *et al.* The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**, 397-403 (1995).

75 Salem, H. *et al.* Drastic genome reduction in an herbivore's pectinolytic symbiont. *Cell* **171**, 1520-1531.e1513 (2017).

76 Davies, J. Millennium bugs. *Trends Cell Biol* **9**, M2-5 (1999).

77 Claessen, D., Rozen, D. E., Kuipers, O. P., Sogaard-Andersen, L. & van Wezel, G. P. Bacterial solutions to multicellularity: a tale of biofilms, filaments and fruiting bodies. *Nat Rev Microbiol* **12**, 115-124 (2014).

78 Flärdh, K. & Buttner, M. J. *Streptomyces* morphogenetics: dissecting differentiation in a filamentous bacterium. *Nat Rev Microbiol* **7**, 36-49 (2009).

79 Chater, K. F. & Losick, R. in *Bacteria as multicellular organisms* (eds J. A. Shapiro & M. Dworkin), 149–182 (New York: Oxford University Press, 1997).

80 Merrick, M. J. A morphological and genetic mapping study of bald colony mutants of *Streptomyces coelicolor*. *J Gen Microbiol* **96**, 299-315. (1976).

81 Hopwood, D. A., Wildermuth, H. & Palmer, H. M. Mutants of *Streptomyces coelicolor* defective in sporulation. *J Gen Microbiol* **61**, 397-408 (1970).

82 Bibb, M. J. Regulation of secondary metabolism in streptomycetes. *Curr Opin Microbiol* **8**, 208-215 (2005).

83 van der Heul, H. U., Bilyk, B. L., McDowall, K. J., Seipke, R. F. & van Wezel, G. P. Regulation of antibiotic production in Actinobacteria: new perspectives from the post-genomic era. *Nat Prod Rep* **35**, 575-604 (2018).

84 Manteca, A. A death round affecting a young compartmentalized mycelium precedes aerial mycelium dismantling in confluent surface cultures of *Streptomyces antibioticus*. *Microbiology* **151**, 3689-3697 (2005).

85 Manteca, A., Mader, U., Connolly, B. A. & Sanchez, J. A proteomic analysis of *Streptomyces coelicolor* programmed cell death. *Proteomics* **6**, 6008-6022 (2006).

86 Tenconi, E., Traxler, M. F., Hoebreck, C., van Wezel, G. P. & Rigali, S. Production of prodiginines is part of a programmed cell death process in *Streptomyces coelicolor*. *Front Microbiol* **9**, 1742 (2018).

87 Gomez-Escribano, J. P. *et al.* Structure and biosynthesis of the unusual polyketide alkaloid coelimycin P1, a metabolic product of the cpk gene cluster of *Streptomyces coelicolor* M145. *Chem Sci* **3**, 2716-2720 (2012).

88 Ohnishi, Y. *et al.* Genome sequence of the streptomycin-producing microorganism *Streptomyces griseus* IFO 13350. *J Bacteriol* **190**, 4050-4060 (2008).

89 Wu, C. *et al.* Expanding the chemical space for natural products by *Aspergillus-Streptomyces* co-cultivation and biotransformation. *Sci Rep* **5**, 10868 (2015).

90 Challis, G. L. & Hopwood, D. A. Synergy and contingency as driving forces for the evolution of multiple secondary metabolite production by *Streptomyces* species. *Proc Natl Acad Sci U S A* **100**, 14555-14561 (2003).

91 Ling, L. L. *et al.* A new antibiotic kills pathogens without detectable resistance. *Nature* **517**, 455-459 (2015).

92 Wilson, M. C. *et al.* An environmental bacterial taxon with a large and distinct metabolic repertoire. *Nature* **506**, 58-62 (2014).

93 Amos, G. C. A. *et al.* Comparative transcriptomics as a guide to natural product discovery and biosynthetic gene cluster functionality. *Proc Natl Acad Sci U S A* **114**, E11121-E11130 (2017).

94 Machado, H., Tuttle, R. N. & Jensen, P. R. Omics-based natural product discovery and the lexicon of genome mining. *Curr Opin Microbiol* **39**, 136-142 (2017).

95 Martinet, L. *et al.* A single biosynthetic gene cluster is responsible for the production of bagremycin antibiotics and ferroverdin iron chelators. *mBio* **10,** e01230-19 (2019).

96 Seipke, R. F., Kaltenpoth, M. & Hutchings, M. I. *Streptomyces* as symbionts: an emerging and widespread theme? *FEMS Microbiol Rev* **36**, 862-876 (2012).

97 Jensen, P. R., Williams, P. G., Oh, D.-C., Zeigler, L. & Fenical, W. Species-specific secondary metabolite production in marine actinomycetes of the genus *Salinispora*. *Appl Environ Microbiol* **73**, 1146-1152 (2007).

98 Yang, A. *et al.* Nitrosporeusines A and B, unprecedented thioester-bearing alkaloids from the Arctic *Streptomyces nitrosporeus*. *Org Lett* **15**, 5366-5369 (2013).

99 Zipperer, A. *et al.* Human commensals producing a novel antibiotic impair pathogen colonization. *Nature* **535**, 511-516 (2016).

100 Ziemert, N. *et al.* Diversity and evolution of secondary metabolism in the marine actinomycete genus *Salinispora*. *Proc Natl Acad Sci U S A* **111**, E1130-E1139 (2014).

101 Adamek, M. *et al.* Comparative genomics reveals phylogenetic distribution patterns of secondary metabolites in *Amycolatopsis* species. *BMC Genomics* **19**, 426 (2018).

102 Andam, C. P., Choudoir, M. J., Vinh Nguyen, A., Sol Park, H. & Buckley, D. H. Contributions of ancestral inter-species recombination to the genetic diversity of extant *Streptomyces* lineages. *ISME J* **10**, 1731-1741 (2016).

103 Tidjani, A. *et al.* Massive gene flux drives genome diversity between sympatric *Streptomyces* conspecifics. *MBio* **10**, e01533-19 (2019).

104 McDonald, B. R. & Currie, C. R. Lateral gene transfer dynamics in the ancient bacterial genus *Streptomyces. mBio* **8**, e00644-17 (2017).

105 Medema, M. H., Cimermancic, P., Sali, A., Takano, E. & Fischbach, M. A. A systematic computational analysis of biosynthetic gene cluster evolution: lessons for engineering biosynthesis. *PLoS Comp Biol* **10**, e1004016 (2014).

106 Joynt, R. & Seipke, R. F. A phylogenetic and evolutionary analysis of antimycin biosynthesis. *Microbiology* **164**, 28-39 (2018).

107 Chevrette, M. G. *et al.* Taxonomic and metabolic incongruence in the ancient genus *Streptomyces*. *Front Microbiol* **10**, 2170 (2019).

108 Navarro-Munoz, J. C. *et al.* A computational framework to explore large-scale biosynthetic diversity. *Nat Chem Biol* **16**, 60-68 (2020).

109 Bruns, H. *et al.* Function-related replacement of bacterial siderophore pathways. *ISME J* **12**, 320-329 (2018).

110 Jensen, P. R. Natural products and the gene cluster revolution. *Trends Microbiol* **24**, 968-977 (2016).

111 Chater, K. F. & Chandra, G. The evolution of development in *Streptomyces* analysed by genome comparisons. *FEMS Microbiol Rev* **30**, 651-672 (2006).

112 Ventura, M. *et al.* Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. *Microbiol Mol Biol Rev* **71**, 495-548 (2007).

113 Choulet, F. *et al.* Evolution of the terminal regions of the *Streptomyces* linear chromosome. *Mol Biol Evol* **23**, 2361-2369 (2006).

114 Bilyk, B., Horbal, L. & Luzhetskyy, A. Chromosomal position effect influences the heterologous expression of genes and biosynthetic gene clusters in *Streptomyces albus* J1074. *Microb Cell Fact* **16**, 5 (2017).

115 Letzel, A.-C. *et al.* Genomic insights into specialized metabolism in the marine actinomycete *Salinispora*. *Environ Microbiol* **19**, 3660-3673 (2017).

116 Ghinet, M. G. *et al.* Uncovering the prevalence and diversity of integrating conjugative elements in Actinobacteria. *PLoS One* **6**, e27846 (2011).

117 Kinashi, H., Shimaji, M. & Sakai, A. Giant linear plasmids in *Streptomyces* which code for antibiotic biosynthesis genes. *Nature* **328**, 454-456 (1987).

118 Medema, M. H. *et al.* The sequence of a 1.8-mb bacterial linear plasmid reveals a rich evolutionary reservoir of secondary metabolic pathways. *Genome Biol Evol* **2**, 212-224 (2010).

119 Mochizuki, S. *et al.* The large linear plasmid pSLA2-L of *Streptomyces rochei* has an unusually condensed gene organization for secondary metabolism. *Mol Microbiol* **48**, 1501-1510 (2003).

120 Huang, J. *et al.* Cross-regulation among disparate antibiotic biosynthetic pathways of *Streptomyces coelicolor*. *Mol Microbiol* **58**, 1276-1287 (2005).

121 McLean, T. C., Hoskisson, P. A. & Seipke, R. F. Coordinate regulation of antimycin and candicidin biosynthesis. *mSphere* **1**, e00305-16 (2016).

122 Liu, G., Chater, K. F., Chandra, G., Niu, G. & Tan, H. Molecular regulation of antibiotic biosynthesis in *Streptomyces*. *Microbiol Mol Biol Rev* **77**, 112-143 (2013).

123 Wietzorrek, A. & Bibb, M. A novel family of proteins that regulates antibiotic production in streptomycetes appears to contain an OmpR-like DNA-binding fold. *Mol Microbiol* **25**, 1181-1184. (1997).

124 Autret, S., Nair, R. & Errington, J. Genetic analysis of the chromosome segregation protein Spo0J of *Bacillus subtilis*: evidence for separate domains involved in DNA binding and interactions with Soj protein. *Mol Microbiol* **41**, 743-755 (2001).

125 Gramajo, H. C., Takano, E. & Bibb, M. J. Stationary-phase production of the antibiotic actinorhodin in *Streptomyces coelicolor* A3(2) is transcriptionally regulated. *Mol Microbiol* **7**, 837-845 (1993).

126 Tomono, A., Tsai, Y., Yamazaki, H., Ohnishi, Y. & Horinouchi, S. Transcriptional control by A-factor of *strR*, the pathway-specific transcriptional activator for streptomycin biosynthesis in *Streptomyces griseus*. *J Bacteriol* **187**, 5595-5604 (2005).

127 Lawlor, E. J., Baylis, H. A. & Chater, K. F. Pleiotropic morphological and antibiotic deficiencies result from mutations in a gene encoding a tRNA-like product in *Streptomyces coelicolor* A3(2). *Genes Dev* **1**, 1305-1310 (1987).

128 Fernandez-Moreno, M. A., Caballero, J. L., Hopwood, D. A. & Malpartida, F. The act cluster contains regulatory and antibiotic export genes, direct targets for translational control by the bldA tRNA gene of *Streptomyces*. *Cell* **66**, 769-780 (1991).

129 Willey, J. M. & Gaskell, A. A. Morphogenetic signaling molecules of the streptomycetes. *Chem Rev* **111**, 174-187 (2011).

130 Tahlan, K. *et al.* Initiation of actinorhodin export in *Streptomyces coelicolor*. *Mol Microbiol* **63**, 951-961 (2007).

131 Wang, L. *et al.* Autoregulation of antibiotic biosynthesis by binding of the end product to an atypical response regulator. *Proc Natl Acad Sci U S A* **106**, 8617-8622 (2009).

132 Willems, A. R. *et al.* Crystal structures of the *Streptomyces coelicolor* TetR-like protein ActR alone and in complex with actinorhodin or the actinorhodin biosynthetic precursor (S)-DNPA. *J Mol Biol* **376**, 1377-1387 (2008).

133 Francis, I. M., Jourdan, S., Fanara, S., Loria, R. & Rigali, S. The cellobiose sensor CebR is the gatekeeper of *Streptomyces scabies* pathogenicity. *MBio* **6**, e02018 (2015).

134 Nazari, B. *et al.* Chitin-induced gene expression involved in secondary metabolic pathways in *Streptomyces coelicolor* A3(2) grown in soil. *Appl Environ Microbiol* **79**, 707-713 (2012).

135 Craig, M. *et al.* Unsuspected control of siderophore production by N-acetylglucosamine in streptomycetes. *Environ Microbiol Rep* **4**, 512-521 (2012).

136 Świątek-Połatyńska, M. A. *et al.* Genome-wide analysis of *in vivo* binding of the master regulator DasR in *Streptomyces coelicolor* identifies novel non-canonical targets. *PLoS One* **10**, e0122479 (2015).

137 Urem, M., Swiatek-Polatynska, M. A., Rigali, S. & van Wezel, G. P. Intertwining nutrient-sensory networks and the control of antibiotic production in *Streptomyces*. *Mol Microbiol* **102**, 183-195 (2016).

138 Zhu, H. *et al.* Eliciting antibiotics active against the ESKAPE pathogens in a collection of actinomycetes isolated from mountain soils. *Microbiology* **160**, 1714-1725 (2014).

139 Hosaka, T. *et al.* Antibacterial discovery in actinomycetes strains with mutations in RNA polymerase or ribosomal protein S12. *Nat Biotechnol* **27**, 462-464 (2009).

140 Tanaka, Y. *et al.* Antibiotic overproduction by *rpsL* and *rsmG* mutants of various actinomycetes. *Appl Environ Microbiol* **75**, 4919-4922 (2009).

141 Bertrand, S. *et al.* Metabolite induction via microorganism co-culture: a potential way to enhance chemical diversity for drug discovery. *Biotechnol Adv* **32**, 1180-1204 (2014).

142 Hoshino, S., Wakimoto, T., Onaka, H. & Abe, I. Chojalactones A–C, cytotoxic butanolides isolated from *Streptomyces* sp. cultivated with mycolic acid containing bacterium. *Org Lett* **17**, 1501-1504 (2015).

143 Sugiyama, R. *et al.* 5-Alkyl-1,2,3,4-tetrahydroquinolines, new membrane-interacting lipophilic metabolites produced by combined culture of *Streptomyces nigrescens* and *Tsukamurella pulmonis*. *Org Lett* **17**, 1918-1921 (2015).

144 Hsiao, N. H., Gottelt, M. & Takano, E. Chapter 6. Regulation of antibiotic production by bacterial hormones. *Methods Enzymol* **458**, 143-157 (2009).

145 Albright, J. C. *et al.* Large-scale metabolomics reveals a complex response of *Aspergillus nidulans* to epigenetic perturbation. *ACS Chem Biol*, 1535-1541 (2015).

146 Craney, A., Ozimok, C., Pimentel-Elardo, S. M., Capretta, A. & Nodwell, J. R. Chemical perturbation of secondary metabolism demonstrates important links to primary metabolism. *Chem Biol* **19**, 1020-1027 (2012).

147 Niu, G., Chater, K. F., Tian, Y., Zhang, J. & Tan, H. Specialised metabolites regulating antibiotic biosynthesis in *Streptomyces* spp. *FEMS Microbiol Rev* **40**, 554-573 (2016).

148 Onaka, H., Mori, Y., Igarashi, Y. & Furumai, T. Mycolic acid-containing bacteria induce natural-product biosynthesis in *Streptomyces* species. *Appl Environ Microbiol* **77**, 400-406 (2011).

149 Schroeckh, V. *et al.* Intimate bacterial-fungal interaction triggers biosynthesis of archetypal polyketides in *Aspergillus nidulans*. *Proc Natl Acad Sci U S A* **106**, 14558-14563 (2009).

150 Sung, A. A., Gromek, S. M. & Balunas, M. J. Upregulation and identification of antibiotic activity of a marine-derived *Streptomyces* sp. via co-cultures with human pathogens. *Mar Drugs* **15**, 250 (2017).

151 Pérez, J. *et al. Myxococcus xanthus* induces actinorhodin overproduction and aerial mycelium formation by *Streptomyces coelicolor*. *Microb Biotechnol* **4**, 175-183 (2011).

152 Patin, N. V., Floros, D. J., Hughes, C. C., Dorrestein, P. C. & Jensen, P. R. The role of inter-species interactions in *Salinispora* specialized metabolism. *Microbiology* **164**, 946-955 (2018).

153 Ezaki, M. *et al.* Biphenomycin A production by a mixed culture. *Appl. Environ. Microbiol.* **58**, 3879-3882 (1992).

154 Kurosawa, K. *et al.* Rhodostreptomycins, antibiotics biosynthesized following horizontal gene transfer from *Streptomyces padanus* to *Rhodococcus fascians*. *J Amer Chem Soc* **130**, 1126-1127 (2008).

155 Traxler, M. F., Seyedsayamdost, M. R., Clardy, J. & Kolter, R. Interspecies modulation of bacterial development through iron competition and siderophore piracy. *Mol Microbiol* **86**, 628-644 (2012).

156 Onaka, H., Tabata, H., Igarashi, Y., Sato, Y. & Furumai, T. Goadsporin, a chemical substance which promotes secondary metabolism and morphogenesis in streptomycetes. I. Purification and characterization. *J Antibiot* **54**, 1036-1044 (2001).

157 Yang, Y. L., Xu, Y., Straight, P. & Dorrestein, P. C. Translating metabolic exchange with imaging mass spectrometry. *Nat Chem Biol* **5**, 885-887 (2009).

158 Currie, C. R., Scott, J. A., Summerbell, R. C. & Malloch, D. D. Fungus-growing ants use antibiotic-producing bacteria to control garden parasites. *Nature* **398**, 701-704 (1999).

159 Heine, D. *et al.* Chemical warfare between leafcutter ant symbionts and a co-evolved pathogen. *Nat Commun* **9**, 2208 (2018).

160 Spaepen, S. in *Principles of Plant-Microbe Interactions* (ed B. Lugtenberg), 247-256 (Springer, 2015).

161 Badri, D. V. & Vivanco, J. M. Regulation and function of root exudates. *Plant Cell Environ* **32**, 666-681 (2009).

162 Badri, D. V., Chaparro, J. M., Zhang, R., Shen, Q. & Vivanco, J. M. Application of natural blends of phytochemicals derived from the root exudates of *Arabidopsis* to the soil reveal that phenolic-related compounds predominantly modulate the soil microbiome. *J Biol Chem* **288**, 4502-4512 (2013).

163 Bulgarelli, D. *et al.* Revealing structure and assembly cues for *Arabidopsis* root-inhabiting bacterial microbiota. *Nature* **488**, 91-95 (2012).

164 Lebeis, S. L. *et al.* PLANT MICROBIOME. Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa. *Science* **349**, 860-864 (2015).

165 Blin, K. *et al.* antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res* **47**, W81-W87 (2019).

166 Skinnider, M. A., Merwin, N. J., Johnston, C. W. & Magarvey, N. A. PRISM 3: expanded prediction of natural product chemical structures from microbial genomes. *Nucleic Acids Res* **45**, W49-W54 (2017).

167 Kautsar, S. A. *et al.* MIBiG 2.0: a repository for biosynthetic gene clusters of known function. *Nucleic Acids Res* **48**, D454-D458 (2019).

168 Doroghazi, J. R. *et al.* A roadmap for natural product discovery based on large-scale genomics and metabolomics. *Nat Chem Biol* **10**, 963-968 (2014).

169 Alanjary, M. *et al.* The Antibiotic Resistant Target Seeker (ARTS), an exploration engine for antibiotic cluster prioritization and novel drug target discovery. *Nucleic Acids Res* **45**, W42-W48 (2017).

170 Culp, E. J. *et al.* Evolution-guided discovery of antibiotics that inhibit peptidoglycan remodelling. *Nature* **578**, 582-587 (2020).

171 van Heel, A. J. *et al.* BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins. *Nucleic Acids Res* **46**, W278-W281 (2018).

172 Du, C. & van Wezel, G. P. Mining for microbial gems: integrating proteomics in the postgenomic natural product discovery pipeline. *Proteomics* **18**, e1700332 (2018).

173 Goering, A. W. *et al.* Metabologenomics: correlation of microbial gene clusters with metabolites drives discovery of a nonribosomal peptide with an unusual amino acid monomer. *ACS Cent Sci* **2**, 99-108 (2016).

174 Soldatou, S., Eldjarn, G. H., Huerta-Uribe, A., Rogers, S. & Duncan, K. R. Linking biosynthetic and chemical space to accelerate microbial secondary metabolite discovery. *FEMS Microbiol Lett* **366** (2019).

175 Stokes, J. M. *et al.* A deep learning approach to antibiotic discovery. *Cell* **180**, 688-702.e13 (2020).

176 van der Hooft, J. J., Wandy, J., Barrett, M. P., Burgess, K. E. & Rogers, S. Topic modeling for untargeted substructure exploration in metabolomics. *Proc Natl Acad Sci U S A* **113**, 13738-13743 (2016).

177 Schorn, M. A. *et al.* Sequencing rare marine actinomycete genomes reveals high density of unique natural product biosynthetic gene clusters. *Microbiology* **162**, 2075-2086 (2016).

178 Charlop-Powers, Z. *et al.* Global biogeographic sampling of bacterial secondary metabolism. *Elife* **4**, e05048 (2015).

179 Charlop-Powers, Z. *et al.* Urban park soil microbiomes are a rich reservoir of natural product biosynthetic diversity. *Proc Natl Acad Sci U S A* **113**, 14811-14816 (2016).

180 Tiwari, K. & Gupta, R. K. Rare actinomycetes: a potential storehouse for novel antibiotics. *Crit Rev Biotechnol* **32**, 108-132 (2012).

181 Donia, M. S. *et al.* A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. *Cell* **158**, 1402-1414 (2014).

182 Culp, E. J. *et al.* Hidden antibiotics in actinomycetes can be identified by inactivation of gene clusters for common antibiotics. *Nat Biotechnol* **37**, 1149-1154 (2019).

183 Hiard, S. *et al.* PREDetector: a new tool to identify regulatory elements in bacterial genomes. *Biochem Biophys Res Commun* **357**, 861-864 (2007).

184 Tan, K., Moreno-Hagelsieb, G., Collado-Vides, J. & Stormo, G. D. A comparative genomics approach to prediction of new members of regulons. *Genome Res* **11**, 566-584 (2001).

185 Rigali, S., Anderssen, S., Naome, A. & van Wezel, G. P. Cracking the regulatory code of biosynthetic gene clusters as a strategy for natural product discovery. *Biochem Pharmacol* **153**, 24-34 (2018).

186 Carrion, V. J. *et al.* Pathogen-induced activation of disease-suppressive functions in the endophytic root microbiome. *Science* **366**, 606-612 (2019).

187 Duhrkop, K., Shen, H., Meusel, M., Rousu, J. & Bocker, S. Searching molecular structure databases with tandem mass spectra using CSI:FingerID. *Proc Natl Acad Sci U S A* **112**, 12580-12585 (2015).

188 Mohimani, H. *et al.* Dereplication of peptidic natural products through database search of mass spectra. *Nat Chem Biol* **13**, 30-37 (2017).

189 Mohimani, H. *et al.* Dereplication of microbial metabolites through database search of mass spectra. *Nat Commun* **9**, 4035 (2018).

190 Ernst, M. *et al.* MolNetEnhancer: enhanced molecular networks by integrating metabolome mining and annotation tools. *Metabolites,* **9**, 144 (2019).

191 Wilson, M. C. & Piel, J. Metagenomic approaches for exploiting uncultivated bacteria as a resource for novel biosynthetic enzymology. *Chem Biol* **20**, 636-647 (2013).

192 Smanski, M. J. *et al.* Synthetic biology to access and expand nature's chemical diversity. *Nat Rev Microbiol* **14**, 135-149 (2016).

193 Sugimoto, Y. *et al.* A metagenomic strategy for harnessing the chemical repertoire of the human microbiome. *Science* **366** (2019).

194 Smanski, M. J. *et al.* Functional optimization of gene clusters by combinatorial design and assembly. *Nat Biotechnol* **32**, 1241-1249 (2014).

195 Shomar, H. *et al.* Metabolic engineering of a carbapenem antibiotic synthesis pathway in *Escherichia coli*. *Nat Chem Biol* **14**, 794-800 (2018).

196 Kuypers, M. M. M., Marchant, H. K. & Kartal, B. The microbial nitrogen-cycling network. *Nat Rev Microbiol* **16**, 263-276 (2018).

197 Stursova, M., Zifcakova, L., Leigh, M. B., Burgess, R. & Baldrian, P. Cellulose utilization in forest litter and soil: identification of bacterial and fungal decomposers. *FEMS Microbiol Ecol* **80**, 735-746 (2012).

198 Tremaroli, V. & Backhed, F. Functional interactions between the gut microbiota and host metabolism. *Nature* **489**, 242-249 (2012).

199 Mendes, R., Garbeva, P. & Raaijmakers, J. M. The rhizosphere microbiome: significance of plant beneficial, plant pathogenic, and human pathogenic microorganisms. *FEMS Microbiol Rev* **37**, 634-663 (2013).

200 Lloyd, K. G., Steen, A. D., Ladau, J., Yin, J. & Crosby, L. Phylogenetically novel uncultured microbial cells dominate earth microbiomes. *mSystems* **3,** e00055-18 (2018).

201 Wibowo, M. C. *et al.* Reconstruction of ancient microbial genomes from the human gut. *Nature* **594**, 234-239 (2021).

202 D'Costa, V. M. *et al.* Antibiotic resistance is ancient. *Nature* **477**, 457-461 (2011).

203 Soldatou, S. *et al.* Comparative metabologenomics analysis of polar actinomycetes. *Mar Drugs* **19**, 103 (2021).

204 Millan-Aguinaga, N. *et al.* Awakening ancient polar Actinobacteria: diversity, evolution and specialized metabolite potential. *Microbiology (Reading)* **165**, 1169-1180 (2019).

205 Wetterich, S. *et al.* Ice Complex formation on Bol'shoy Lyakhovsky Island (New Siberian Archipelago, East Siberian Arctic) since about 200 ka. *Quaternary Research* **92**, 530-548 (2019).

206 Nikolskiy, P. A., Sulerzhitsky, L. D. & Pitulko, V. V. Last straw versus Blitzkrieg overkill: Climate-driven changes in the Arctic Siberian mammoth population and the Late Pleistocene extinction problem. *Quaternary Science Reviews* **30**, 2309-2328 (2011).

207 Grigoriev, S. E. *et al.* A woolly mammoth (*Mammuthus primigenius*) carcass from Maly Lyakhovsky Island (New Siberian Islands, Russian Federation). *Quaternary International* **445**, 89-103 (2017).

208 Kornienko, I. V. *et al.* Complete mitochondrial genome of a woolly mammoth (*Mammuthus primigenius*) from Maly Lyakhovsky Island (New Siberian Islands, Russia) and its phylogenetic assessment. *Mitochondrial DNA B Resour* **3**, 596-598 (2018).

209 Goncharov, A. *et al.* Draft Genome Sequence of *Enterococcus faecium* Strain 58m, Isolated from Intestinal Tract Content of a Woolly Mammoth, *Mammuthus primigenius*. *Genome Announc* **4** (2016).

210 Battistuzzi, F. U., Feijao, A. & Hedges, S. B. A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. *BMC Evol Biol* **4**, 44 (2004).

211 Barco, R. A. *et al.* A genus definition for bacteria and archaea based on a standard genome relatedness index. *mBio* **11**, e02475-19 (2020).

212 Hohmann, C. *et al.* Caboxamycin, a new antibiotic of the benzoxazole family produced by the deep-sea strain Streptomyces sp. NTK 937. *J Antibiot (Tokyo)* **62**, 99-104 (2009).

213 Ma, L. *et al.* Complete genome sequence of Streptomyces sp. SCSIO 03032 isolated from Indian Ocean sediment, producing diverse bioactive natural products. *Mar Genomics* **55**, 100803 (2021).

214 Rottig, A. *et al. Streptomyces jeddahensis* sp. nov., an oleaginous bacterium isolated from desert soil. *Int J Syst Evol Microbiol* **67**, 1676-1682 (2017).

215 Cornell, C. R., Marasini, D. & Fakhr, M. K. Draft Genome Sequences of Megaplasmid-Bearing *Streptomyces* sp. Strains BF-3 and 4F, Isolated from the Great Salt Plains of Oklahoma. *Genome Announc* **6**, e00208-18 (2018).

216 Zhang, Z. *et al.* Antibiotic production in *Streptomyces* is organized by a division of labor through terminal genomic differentiation. *Sci Adv* **6**, eaay5781 (2020).

217 Blin, K. *et al.* antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res* **49**, W29-W35 (2021).

218 Avalos, M., Garbeva, P., Raaijmakers, J. M. & van Wezel, G. P. Production of ammonia as a low-cost and long-distance antibiotic strategy by *Streptomyces* species. *ISME J* **14**, 569-583 (2020).

219 Avalos, M., van Wezel, G. P., Raaijmakers, J. M. & Garbeva, P. Healthy scents: microbial volatiles as new frontier in antibiotic research? *Curr Opin Microbiol* **45**, 84-91 (2018).

220 Wang, M. *et al.* Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol* **34**, 828-837 (2016).

221 Sunagawa, S. *et al.* Ocean plankton. Structure and function of the global ocean microbiome. *Science* **348**, 1261359 (2015).

222 Ferrari, G. *et al.* Assessing metagenomic signals recovered from Lyuba, a 42,000-year-old permafrost-preserved Woolly Mammoth calf. *Genes (Basel)* **9** (2018).

223 Czech, L. *et al.* Role of the extremolytes ectoine and hydroxyectoine as stress protectants and nutrients: genetics, phylogenomics, biochemistry, and structural analysis. *Genes (basel)* **9,** 177 (2018).

224 Belknap, K. C., Park, C. J., Barth, B. M. & Andam, C. P. Genome mining of biosynthetic and chemotherapeutic gene clusters in *Streptomyces* bacteria. *Sci Rep* **10**, 2003 (2020).

225 Joynt, R. & Seipke, R. F. A phylogenetic and evolutionary analysis of antimycin biosynthesis. *Microbiology (Reading)* **164**, 28-39 (2018).

226 Hayakawa, M. & Nonomura, H. Humic acid-vitamin agar, a new medium for the selective isolation of soil actinomycetes. *Journal of Fermentation Technology* **65**, 501-509 (1987).

227 Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F. & Hopwood, D. A. *Practical Steptomyces Genetics*. (John Innes Foundation, 2000).

228 Kuester, E. & Williams, S. T. Selection of Media for Isolation of Streptomycetes. *Nature* **202**, 928-929 (1964).

229 Wick, R. R., Judd, L. M., Gorrie, C. L. & Holt, K. E. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* **13**, e1005595 (2017).

230 Walker, B. J. *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**, e112963 (2014).

231 Asnicar, F. *et al.* Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. *Nat Commun* **11**, 2500 (2020).

232 Jain, C., Rodriguez, R. L., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* **9**, 5114 (2018).

233 Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068-2069 (2014).

234 Kanz, C. *et al.* The EMBL Nucleotide Sequence Database. *Nucleic Acids Res* **33**, D29-33 (2005).

235 Wang, Y. *et al.* MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res* **40**, e49 (2012).

236 Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res* **19**, 1639-1645 (2009).

237 Liu, M. & Douthwaite, S. Activity of the ketolide telithromycin is refractory to Erm monomethylation of bacterial rRNA. *Antimicrob Agents Chemother* **46**, 1629-1633 (2002).

238 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504 (2003).

239 Kolter, R. & van Wezel, G. P. Goodbye to brute force in antibiotic discovery? *Nat Microbiol* **1**, 15020 (2016).

240 Lewis, K. Platforms for antibiotic discovery. *Nat Rev Drug Discov* **12**, 371-387 (2013).

241 van der Aart, L. T. *et al.* Polyphasic classification of the gifted natural product producer *Streptomyces roseifaciens* sp. nov. *Int J Syst Evol Microbiol* **69**, 899-908 (2019).

242 Zaborina, O. *et al.* Dynorphin activates quorum sensing quinolone signaling in *Pseudomonas aeruginosa*. *PLoS Pathog* **3**, e35 (2007).

243 Lyte, M. & Ernst, S. Catecholamine induced growth of Gram negative bacteria. *Life Sci* **50**, 203-212 (1992).

244 Belay, T., Aviles, H., Vance, M., Fountain, K. & Sonnenfeld, G. Catecholamines and in vitro growth of pathogenic bacteria: enhancement of growth varies greatly among bacterial species. *Life Sci* **73**, 1527-1535 (2003).

245 Sandrini, S., Alghofaili, F., Freestone, P. & Yesilkaya, H. Host stress hormone norepinephrine stimulates pneumococcal growth, biofilm formation and virulence gene expression. *BMC Microbiol* **14**, 180 (2014).

246 Peterson, G., Kumar, A., Gart, E. & Narayanan, S. Catecholamines increase conjugative gene transfer between enteric bacteria. *Microb Pathog* **51**, 1-8 (2011).

247 Bernatoniene, J. & Kopustinskiene, D. M. The role of catechins in cellular responses to oxidative stress. *Molecules* **23**, 965 (2018).

248 Gubbens, J. *et al.* Intertwined precursor supply during biosynthesis of the catecholate-hydroxamate siderophores qinichelins in *Streptomyces* sp. MBT76. *ACS Chem Biol* **12**, 2756-2766 (2017).

249 Gubbens, J. *et al.* Natural product proteomining, a quantitative proteomics platform, allows rapid discovery of biosynthetic gene clusters for different classes of natural products. *Chem Biol* **21**, 707-718 (2014).

250 Li, B. & Walsh, C. T. Identification of the gene cluster for the dithiolopyrrolone antibiotic holomycin in *Streptomyces clavuligerus*. *Proc Natl Acad Sci U S A* **107**, 19731-19735 (2010).

251 Jones, S. E. *et al. Streptomyces* volatile compounds influence exploration and microbial community dynamics by altering iron availability. *mBio* **10**, e00171-19 (2019).

252 Lee, N. *et al.* Iron competition triggers antibiotic biosynthesis in *Streptomyces coelicolor* during coculture with *Myxococcus xanthus*. *ISME J* **14**, 1111-1124 (2020).

253 Perraud, Q. *et al.* Opportunistic use of catecholamine neurotransmitters as siderophores to access iron by *Pseudomonas aeruginosa*. *Environ Microbiol* (2020).

254 Beasley, F. C., Marolda, C. L., Cheung, J., Buac, S. & Heinrichs, D. E. *Staphylococcus aureus* transporters Hts, Sir, and Sst capture iron liberated from human transferrin by Staphyloferrin A, Staphyloferrin B, and catecholamine stress hormones, respectively, and contribute to virulence. *Infect Immun* **79**, 2345-2355 (2011).

255 Dichtl, S. *et al.* Dopamine is a siderophore-like iron chelator that promotes *Salmonella enterica* Serovar Typhimurium virulence in mice. *mBio* **10**, e02624-18 (2019).

256 Li, W. *et al.* Norepinephrine represses the expression of toxA and the siderophore genes in *Pseudomonas aeruginosa*. *FEMS Microbiol Lett* **299**, 100-109 (2009).

257 Kloepper, J. W., Leong, J., Teintze, M. & Schroth, M. N. Enhanced plant growth by siderophores produced by plant growth-promoting rhizobacteria. *Nature* **286**, 885-886 (1980).

258 Chan, A. N. *et al.* Role for dithiolopyrrolones in disrupting bacterial metal homeostasis. *Proc Natl Acad Sci U S A* **114**, 2717-2722 (2017).

259 Li, B., Wever, W. J., Walsh, C. T. & Bowers, A. A. Dithiolopyrrolones: biosynthesis, synthesis, and activity of a unique class of disulfide-containing antibiotics. *Nat Prod Rep* **31**, 905-923 (2014).

260 Schwyn, B. & Neilands, J. B. Universal chemical assay for the detection and determination of siderophores. *Anal Biochem* **160**, 47-56 (1987).

261 Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* **37**, 540-546 (2019).

262 Wessel, D. & Flugge, U. I. A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal Biochem* **138**, 141-143 (1984).

263 Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat Protoc* **2**, 1896-1906 (2007).

264 Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367-1372 (2008).

265 Perez-Riverol, Y. *et al.* The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* **47**, D442-D450 (2019).

266 Nett, M., Ikeda, H. & Moore, B. S. Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Nat Prod Rep* **26**, 1362-1384 (2009).

267 Covington, B. C., Xu, F. & Seyedsayamdost, M. R. A natural product chemist's guide to unlocking silent biosynthetic gene clusters. *Annu Rev Biochem* **90**, 763-788 (2021).

268 Westhoff, S., Kloosterman, A. M., van Hoesel, S. F. A., van Wezel, G. P. & Rozen, D. E. Competition sensing changes antibiotic production in *Streptomyces*. *mBio* **12,** e02729-20 (2021).

269 Soares, A. R. *et al.* The role of L-DOPA in plants. *Plant Signal Behav* **9**, e28275 (2014).

270 Ladino-Orjuela, G., Gomes, E., da Silva, R., Salt, C. & Parsons, J. R. Metabolic pathways for degradation of aromatic hydrocarbons by bacteria. *Rev Environ Contam Toxicol* **237**, 105-121 (2016).

271 Xia, J., Psychogios, N., Young, N. & Wishart, D. S. MetaboAnalyst: a web server for metabolomic data analysis and interpretation. *Nucleic Acids Res* **37**, W652-660 (2009).

272 Mladic, M. *et al.* Detection and identification of antibacterial proteins in snake venoms using at-line nanofractionation coupled to LC-MS. *Toxicon* **155**, 66-74 (2018).

273 Frank, A. M. *et al.* Clustering millions of tandem mass spectra. *J Proteome Res* **7**, 113-122 (2008).

274 Peng, A. *et al.* Angucycline glycosides from an intertidal sediments strain *Streptomyces* sp. and their cytotoxic activity against hepatoma carcinoma cells. *Mar Drugs* **16**, 470 (2018).

275 Kharel, M. K. *et al.* Angucyclines: biosynthesis, mode-of-action, new natural products, and synthesis. *Nat Prod Rep* **29**, 264-325 (2012).

276 Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173-1183 (2013).

277 Tong, Y., Charusanti, P., Zhang, L., Weber, T. & Lee, S. Y. CRISPR-Cas9 based engineering of actinomycetal genomes. *ACS Synth Biol* **4**, 1020-1029 (2015).

278 Ingram, C., Brawner, M., Youngman, P. & Westpheling, J. xylE functions as an efficient reporter gene in *Streptomyces* spp.: use for the study of galP1, a catabolite-controlled promoter. *J Bacteriol* **171**, 6617-6624 (1989).

279 Labes, G., Bibb, M. & Wohlleben, W. Isolation and characterization of a strong promoter element from the *Streptomyces ghanaensis* phage I19 using the gentamicin resistance gene (aacC1) of Tn 1696 as reporter. *Microbiology (Reading)* **143 (Pt 5)**, 1503-1512 (1997).

280 Xu, F., Nazari, B., Moon, K., Bushin, L. B. & Seyedsayamdost, M. R. Discovery of a cryptic antifungal compound from *Streptomyces albus* J1074 using high-throughput elicitor screens. *J Am Chem Soc* **139**, 9203-9212 (2017).

281 Pluskal, T., Castillo, S., Villar-Briones, A. & Oresic, M. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **11**, 395 (2010).

282 Myers, O. D., Sumner, S. J., Li, S., Barnes, S. & Du, X. One step forward for reducing false positive and false negative compound identifications from mass spectrometry metabolomics data: new algorithms for constructing extracted ion chromatograms and detecting chromatographic peaks. *Anal Chem* **89**, 8696-8703 (2017).

283 Distler, U. *et al.* Drift time-specific collision energies enable deep-coverage data-independent acquisition proteomics. *Nat Methods* **11**, 167-170 (2014).

284 Tong, Y., Charusanti, P., Zhang, L., Weber, T. & Lee, S. Y. CRISPR-Cas9 Based Engineering of Actinomycetal Genomes. *ACS Synth Biol* **4**, 1020-1029 (2015).

285 Ultee, E. *et al.* Teichoic acids anchor distinct cell wall lamellae in an apically growing bacterium. *Commun Biol* **3**, 314 (2020).

286 Zhang, L. *et al.* An alternative and conserved cell wall enzyme that can substitute for the lipid II synthase MurG. *mBio* **12**, e03381-20 (2021).

287 Vara, J., Lewandowska-Skarbek, M., Wang, Y. G., Donadio, S. & Hutchinson, C. R. Cloning of genes governing the deoxysugar portion of the erythromycin biosynthesis pathway in *Saccharopolyspora erythraea* (*Streptomyces erythreus*). *J Bacteriol* **171**, 5872-5881 (1989).

288 Currie, C. R., Bot, A. N. M. & Boomsma, J. J. Experimental evidence of a tripartite mutualism: bacteria protect ant fungus gardens from specialized parasites. *Oikos* **101**, 91-102 (2003).

289 Oh, D. C., Poulsen, M., Currie, C. R. & Clardy, J. Dentigerumycin: a bacterial mediator of an ant-fungus symbiosis. *Nat Chem Biol* **5**, 391-393 (2009).

290 Bruner-Montero, G. *et al.* Symbiont-mediated protection of *Acromyrmex* leaf-cutter ants from the entomopathogenic fungus *Metarhizium anisopliae*. *mBio* **12**, e0188521 (2021).

291 Chevrette, M. G. & Currie, C. R. Emerging evolutionary paradigms in antibiotic discovery. *J Ind Microbiol Biotechnol* **46**, 257-271 (2019).

292 Adnani, N. *et al.* Coculture of marine invertebrate-associated bacteria and interdisciplinary technologies enable biosynthesis and discovery of a new antibiotic, keyicin. *ACS Chem Biol* **12**, 3093-3102 (2017).

293 Cho, I. & Blaser, M. J. The human microbiome: at the interface of health and disease. *Nat Rev Genet* **13**, 260-270 (2012).

294 Fukuda, S. *et al.* Bifidobacteria can protect from enteropathogenic infection through production of acetate. *Nature* **469**, 543-547 (2011).

295 Henrick, B. M. *et al.* Bifidobacteria-mediated immune system imprinting early in life. *Cell* **184**, 3884-3898 e3811 (2021).

296 Zhalnina, K. *et al.* Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly. *Nat Microbiol* **3**, 470-480 (2018).

297 Bulgarelli, D. *et al.* Revealing structure and assembly cues for *Arabidopsis* root-inhabiting bacterial microbiota. *Nature* **488**, 91-95 (2012).

298 Badri, D. V., Chaparro, J. M., Zhang, R., Shen, Q. & Vivanco, J. M. Application of natural blends of phytochemicals derived from the root exudates of *Arabidopsis* to the soil reveal that phenolic-related compounds predominantly modulate the soil microbiome. *J Biol Chem* **288**, 4502-4512 (2013).

299 Kaltenpoth, M. *et al.* Partner choice and fidelity stabilize coevolution in a Cretaceous-age defensive symbiosis. *Proc Natl Acad Sci U S A* **111**, 6359-6364 (2014).

300 Zhang, M. M., Poulsen, M. & Currie, C. R. Symbiont recognition of mutualistic bacteria by *Acromyrmex* leaf-cutting ants. *ISME J* **1**, 313-320 (2007).

301 van Bergeijk, D. A. *et al.* The ubiquitous catechol moiety elicits siderophore and angucycline production in *Streptomyces*. *Commun Chem* **5**, 14 (2022).

302 Burns, A. R. & Guillemin, K. The scales of the zebrafish: host-microbiota interactions from proteins to populations. *Curr Opin Microbiol* **38**, 137-141 (2017).

303 Douglas, A. E. Simple animal models for microbiome research. *Nat Rev Microbiol* **17**, 764-775 (2019).

304 Flores, E. M., Nguyen, A. T., Odem, M. A., Eisenhoffer, G. T. & Krachler, A. M. The zebrafish as a model for gastrointestinal tract-microbe interactions. *Cell Microbiol* **22**, e13152 (2020).

305 Stagaman, K., Sharpton, T. J. & Guillemin, K. Zebrafish microbiome studies make waves. *Lab Anim (NY)* **49**, 201-207 (2020).

306 Alanjary, M., Steinke, K. & Ziemert, N. AutoMLST: an automated web server for generating multi-locus species trees highlighting natural product potential. *Nucleic Acids Res* **47**, W276-W282 (2019).

307 Grumaz, C. *et al.* Draft genome sequence of *Pseudonocardia autotrophica* strain DSM 43083, an efficient producer of peroxidases for lignin modification. *Genome Announc* **5** (2017).

308 Grant, J. R. & Stothard, P. The CGView Server: a comparative genomics tool for circular genomes. *Nucleic Acids Res* **36**, W181-184 (2008).

309 Gilchrist, C. L. M. *et al.* cblaster: a remote search tool for rapid identification and visualization of homologous gene clusters. *Bioinformatics Advances* **1**, vbab016 (2021).

310 Rajilic-Stojanovic, M. & de Vos, W. M. The first 1000 cultured species of the human gastrointestinal microbiota. *FEMS Microbiol Rev* **38**, 996-1047 (2014).

311 Gaiser, R. F., M.et al. A mammalian commensal of the oropharyngeal cavity produces antibiotic and antiviral valinomycin *in vivo* (preprint). *Research square* (2021).

312 Goldstein, S. L. & Klassen, J. L. *Pseudonocardia* symbionts of fungus-growing ants and the evolution of defensive secondary metabolism. *Front Microbiol* **11**, 621041 (2020).

313 Parra, J., Soldatou, S., Rooney, L. M. & Duncan, K. R. *Pseudonocardia abyssalis* sp. nov. and *Pseudonocardia oceani* sp. nov., two novel actinomycetes isolated from the deep Southern Ocean. *Int J Syst Evol Microbiol* **71** (2021).

314 van den Berg, A. H., McLaggan, D., Diéguez-Uribeondo, J. & van West, P. The impact of the water moulds *Saprolegnia diclina* and *Saprolegnia parasitica* on natural ecosystems and the aquaculture industry. *Fungal Biology Reviews* **27**, 33-42 (2013).

315 Riahi, H. S., Heidarieh, P. & Fatahi-Bafghi, M. Genus *Pseudonocardia*: What we know about its biological properties, abilities and current application in biotechnology. *J Appl Microbiol* **132**, 890-906 (2022).

316 Tyagi, A., Singh, B., Billekallu Thammegowda, N. K. & Singh, N. K. Shotgun metagenomics offers novel insights into taxonomic compositions, metabolic pathways and antibiotic resistance genes in fish gut microbiome. *Arch Microbiol* **201**, 295-303 (2019).

317 Sharpton, T. J., Stagaman, K., Sieler, M. J., Jr., Arnold, H. K. & Davis, E. W., 2nd. Phylogenetic integration reveals the zebrafish core microbiome and its sensitivity to environmental exposures. *Toxics* **9** (2021).

318 Stephens, W. Z. *et al.* The composition of the zebrafish intestinal microbial community varies across development. *ISME J* **10**, 644-654 (2016).

319 Bolyen, E. *et al.* Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* **37**, 852-857 (2019).

320 Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**, 581-583 (2016).

321 Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**, 3059-3066 (2002).

322 Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490 (2010).

323 Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**, D590-596 (2013).

324 Jami, M., Ghanbari, M., Kneifel, W. & Domig, K. J. Phylogenetic diversity and biological activity of culturable Actinobacteria isolated from freshwater fish gut microbiota. *Microbiol Res* **175**, 6-15 (2015).

325 Maciejewska, M. *et al. Streptomyces lunaelactis* sp. nov., a novel ferroverdin A-producing *Streptomyces* species isolated from a moonmilk speleothem. *Antonie Van Leeuwenhoek* **107**, 519-531 (2015).

326 Elsayed, S. S. *et al.* Atypical spirotetronate polyketides identified in the underexplored genus *Streptacidiphilus*. *J Org Chem* **85**, 10648-10657 (2020).

327 Abdel-Mageed, W. M. *et al.* Dermacozines, a new phenazine family from deep-sea dermacocci isolated from a Mariana Trench sediment. *Org Biomol Chem* **8**, 2352-2362 (2010).

328 Coelho, L. P. *et al.* Towards the biogeography of prokaryotic genes. *Nature* **601**, 252-256 (2022).

329 Sharrar, A. M. *et al.* Bacterial secondary metabolite biosynthetic potential in soil varies with phylum, depth, and vegetation type. *mBio* **11**, e00416-20 (2020).

330 Martiny, J. B. *et al.* Microbial biogeography: putting microorganisms on the map. *Nat Rev Microbiol* **4**, 102-112 (2006).

331 D'Onofrio, A. *et al.* Siderophores from neighboring organisms promote the growth of uncultured bacteria. *Chem Biol* **17**, 254-264 (2010).

332 Nichols, D. *et al.* Use of ichip for high-throughput in situ cultivation of “uncultivable” microbial species. *Appl Environ Microbiol* **76**, 2445-2450 (2010).

333 Gu, S. *et al.* Competition for iron drives phytopathogen control by natural rhizosphere microbiomes. *Nat Microbiol* **5**, 1002-1010 (2020).

334 Gu, S. *et al.* Siderophore-mediated interactions determine the disease suppressiveness of microbial consortia. *mSystems* **5**, e00811-19 (2020).

335 Radzki, W. *et al.* Bacterial siderophores efficiently provide iron to iron-starved tomato plants in hydroponics culture. *Antonie Van Leeuwenhoek* **104**, 321-330 (2013).

336 Lin, L. & Tan, R. X. Cross-kingdom actions of phytohormones: a functional scaffold exploration. *Chem Rev* **111**, 2734-2760 (2011).

Curriculum vitae

Doris Antonia van Bergeijk was born in Hoofddorp, The Netherlands, on the 24th of January, 1992. She attended secondary school at the Haarlemmermeerlyceum in Hoofddorp, during which she followed three years of bilingual education (English-Dutch). After graduating in 2010, she conducted voluntary work in the Philippines and China, and worked and travelled in New Zealand.

In 2011, she continued her studies with the Bachelor Pharmacy at Utrecht University and discovered her passion for microbiology and research. After obtaining her BSc in July 2014, she started the Research Master Drug Innovation at Utrecht University. During her first internship in the pharmaco-epidemiology group of prof. dr. Anke-Hilse Maitland-van der Zee, she collected clinical data for the Children Anticoagulation and Pharmacogenetics Study and extracted DNA from saliva samples for genotyping under supervision of dr. Hedy Maagdenberg. During her second internship, Doris studied the effect of salt exposure on macrophages in zebrafish larvae under supervision of dr. Neeraj Dhaun and dr. Alicja Czopek at the University of Edinburgh. Eager to gain more experience in the lab, she completed an extracurricular internship in the group of dr. ing. Erik Hendriksen, working on the development of a new *in vitro* model to study mast cell-neuron interactions in the brain.

In February 2017, Doris started her PhD project at the Institute of Biology (IBL) at Leiden University, as part of the metagenomics PhD program of the Netherlands Centre for One Health (NCOH), connecting to the NCOH research theme *tackling antimicrobial resistance*. Under the joint supervision of prof. dr. Gilles van Wezel and prof. dr. Herman Spaink, she worked on different ecological approaches to access the full biosynthetic potential of Actinobacteria. During her PhD, Doris organised several workshops, seminars, and symposia as board member of Young NCOH and as part of the spotlight committee of the IBL. Doris will continue her scientific career as a postdoctoral researcher in the laboratory of prof. dr. Jeroen Raes at KU Leuven, Belgium, to work on intestinal bacteria and their role in health and disease.

List of publications

van Bergeijk, D.A., Elsayed, S.S., Du, C., Nuñez Santiago, I., Roseboom, A.M., Zhang, L., Carrión V.J., Spaink, H.P., van Wezel, G.P. The ubiquitous catechol moiety elicits siderophore and angucycline production in *Streptomyces. Commun Chem* **5**, 14 (2022).

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)

Hendriksen E., **van Bergeijk D.A.,** Oosting R.S. Redegeld F.A. Mast cells in neuroinflammation and brain disorders. *Neurosci Biobehav Rev.* **79**, 119-133. (2017)