Discovery, expression, and characterization of biosynthetic gene clusters encoding antimicrobial secondary metabolites from a soil metagenome and cultured Actinobacteria

by

Megan Allyse Sandoval-Powers

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
August 6, 2022

Keywords: Biosynthetic Gene Clusters, Heterologous Expression, Metagenomics, Natural Products, Secondary Metabolites

Approved by

Mark R. Liles, Chair, Professor of Biological Sciences
Paul A. Cobine, Co-Chair, Professor of Biological Sciences
Alexey Petrov, Assistant Professor of Biological Sciences
Peter Panizzi, Professor of Drug Discovery and Development
Laura Huber, Assistant Professor of Epidemiology

Abstract

Dating back to the discovery of penicillin in 1928, nature and its diverse microbial life has been the most prolific source of antimicrobial compounds. These bioactive compounds are often derived from microbial secondary metabolites whose biosynthesis is encoded by a group of genes that cluster together in the genome known as biosynthetic gene clusters (BGCs). Advancements in sequencing and bioinformatics has revealed a wealth of uncharacterized BGCs from environmental microorganisms. Accessing these untapped gene clusters is important for the discovery of novel therapeutic compounds for human health, and to advance our understanding on their function, diversity, and evolution.

The purpose of this dissertation was to identify, express, and characterize novel gene clusters encoding antimicrobial secondary metabolites using culture-dependent and –independent techniques. Although different in their methodologies, each chapter is centralized on identifying BGCs from and/or expressing clusters in taxa from the Gram-positive bacterial phylum Actinobacteria. Members of the phylum Actinobacteria, particularly from the genus *Streptomyces*, are widely distributed in soil and marine environments and are the source of most clinically used antibiotics. The established biosynthetic potential of Actinobacteria taxa makes them an ideal target to find new BGCs and to serve as hosts for the expression of gene clusters encoding bioactive secondary metabolites.

In Chapter 1, I discuss the role of Actinobacteria, with particular emphasis on the genus *Streptomyces*, in past and modern strategies to find natural products and offer insights into their advantages and limitations. For Chapter 2, a culture-based approach was taken to isolate a new species of the genus *Streptomyces* from marine sponges collected in the Trondheim fjord of Norway. This new species, *Streptomyces poriferorum* P01-B04^{T}, was found to harbor many

BGCs encoding secondary metabolites in its genome and expressed antibacterial activity against a drug-resistant *Staphylococcus aureus* pathogen. For Chapter 3, a culture-independent (i.e. metagenomics-based) strategy was used to identify and express BGCs predicted to encode novel chemistry from a soil metagenome in an engineered *Streptomyces coelicolor* heterologous host. A high hit rate of *S. coelicolor* clones were found to inhibit the growth of a multidrug-resistant *Acinetobacter baumannii* pathogen. The expression of a BGC by a particular clone, P17B06, could be enhanced using a dual-inducible expression vector and attempts were made to characterize the antimicrobial secondary metabolites encoded by this gene cluster through a bioactivity-guided fractionation approach. Finally, in Chapter 4, I explored the function and taxonomic origin of another BGC (P12B21 BGC) encoding antimicrobial metabolites derived from the soil metagenomic library. I found that the P12B21 BGC encodes a hybrid biosynthetic pathway with homology to a gene cluster from *Microbacterium hatanonis*, a member of the phylum Actinobacteria. Collectively, this dissertation work demonstrates the value of culture-dependent and metagenomics-based approaches to access and study biologically-relevant gene clusters from microorganisms.

Acknowledgments

There are so many amazing people that I have had the privilege to interact with during my studies. Your support in this journey means more to me than I can ever express, but I would like to take a moment and thank each of you.

First, I wish to thank my advisor, Dr. Mark Liles, for your invaluable advice, constant support, and patience. I am extremely grateful you took me on as a student, and continued to support and challenge me over the years. I am so appreciative of all the opportunities you provided to help me develop as a researcher. Your positive attitude towards science even when faced with obstacles is truly an inspiration and something I hope to take forward throughout my career. You have been the best role model of what an excellent scientist, mentor, collaborator, and person should be.

Thank you to my committee members, Dr. Paul Cobine, Dr. Peter Panizzi, and Dr. Alexey Petrov for their thoughtful guidance throughout my time here at Auburn. I wish to thank Dr. Laura Huber who so generously took time out of her schedule to review this dissertation and serve as my outside reader. I also want to express my gratitude to the amazing collaborators that I have had the privilege to work with over the years including Dr. Stanislava Králová, Dr. Alexander Wentzel, Dr. Wendy Strangman, Dr. David Mead, and so many more. I am very grateful to have had the opportunity to be mentored by each of you.

I would also like to acknowledge the many wonderful labmates of the Liles Lab both past and present. I couldn't ask for a better group of people to be surrounded by. Thank you for all the fun conversations, rants, laughter, and support. Speaking of which… a big thank you to Nancy Capps and Stacey LaFrentz for being the best lab managers. I am very grateful to have worked with two wonderful people like yourselves. I am also forever thankful to the undergraduate

students that helped with this work. I especially have to thank, Dory Fawwal, who always went above and beyond to help me wherever I needed it. I can't thank you enough for the time you devoted to these projects, your positive attitude every step of the way, and for becoming my closest friend. It was truly an honor to serve as your mentor and I am excited to see all the amazing things you accomplish in your future. Thank you to the rest of my friends both near and far for the laughs, support, and being an outlet to distract me from the stress of graduate school: Emily, Jacob, Michael, and Brandon to name a few.

I think it cannot be overstated how important a strong support group is. Thank you to all the members of my immediate and extended family for being my support group. I especially want to thank my parents, Mark and Kim Sandoval, for your endless support, love, and encouragement. You both have always been "in my corner" and having your continued support through this journey meant the world to me. My accomplishments are because you believed in me. Thank you to my brother and sister-in-law, Zac and Joy Sandoval, and my adorable nephew Emerson Sandoval, for your support and keeping me grounded. I am also thankful for the support and encouragement from my in-laws, Kathy and John Powers.

I also have to acknowledge my small, crazy dog Tracer. Seeing your happy "waggles" everytime I came home could instantly make me happy even after a long, hard day in the lab. You kept life fun when I needed it the most.

Finally, I owe my deepest gratitude to my husband, Matthew Powers. I am forever thankful for your love, support, and encouragement. Even when we were both faced with stressful deadlines or hurdles, you were always willing to stop anything that you were doing to help and motivate me. Thank you for believing in me when I didn't believe in myself. I would not have made it this far without you by my side.

Table of Contents

Abstract 2

Acknowledgments 4

List of Tables 8

List of Figures 9

Chapter 1. Literature Review 11

Abstract 11

Introduction 11

Understanding the genus *Streptomyces* 14

Streptomyces' role in past approaches for microbial natural product discovery 21

Modern approaches for natural product discovery from *Streptomyces* and other bacteria 25

Concluding remarks 50

Chapter 2. *Streptomyces poriferorum* sp. nov., a novel marine sponge-derived Actinobacteria species expressing anti-MRSA activity 52

Abstract 52

Introduction 53

Materials and Methods 55

Results and Discussion 62

Conclusion 81

Chapter 2 Supplementary Material 83

Chapter 3. Heterologous expression of metagenome-derived biosynthetic pathways in an engineered strain of *Streptomyces coelicolor* 93

Abstract 93
Introduction 94
Materials and Methods 98
Results 110
Discussion 132
Chapter 3 Supplementary Material 142
Chapter 4. Characterization and taxonomy of a metagenome-derived biosynthetic gene cluster that encodes an antimicrobial secondary metabolite 156
Abstract 156
Introduction 157
Materials and Methods 160
Results 168
Discussion 179
Chapter 4 Supplementary Material 186
Chapter 5. Conclusions 193
References 196

List of Tables

Chapter 2

Table 2.1. Genomic features of *S. poriferorum* sp. nov. and closest related type strains 66

Table 2.2. Genome-based comparisons of novel strains and closely related type strains 68

Table 2.3. Differentiating characteristics of novel strains and closely related type strains 71

Table 2.4. Cellular fatty acid composition of novel strains and closely related type strains 74

Chapter 3

Table 3.1. Antibacterial activities of metagenomic clone extracts from polymeric resins 118

Table 3.2. Zones of inhibition of crude extracts and fractions from clone P17B06 126

List of Figures

Chapter 1

Figure 1.1. Scanning electron micrograph of a *Streptomyces* species 16

Figure 1.2. Timeline of the discovery of different antibiotic classes and their origin 23

Figure 1.3. Graphical representation of the construction and screening of a metagenomic library 39

Chapter 2

Figure 2.1. Maximum-likelihood 16S phylogenetic tree based on 16S rRNA sequences 64

Figure 2.2. Phylogenomic tree based on genome sequences from isolates and type strains 67

Figure 2.3. Biosynthetic gene clusters predicted in the genome of isolates and type strains 76

Figure 2.4. Annotation of a hybrid biosynthetic gene cluster from isolate P01-B04^{T} 77

Chapter 3

Figure 3.1. Scatterplot depicting the relationship between GC% content of each BGC and the GC% content of the expression host 112

Figure 3.2. Inducible expression analysis of metagenomic biosynthetic gene clusters by *S. coelicolor* M1154 and schematic representations of each gene cluster 114

Figure 3.3. Organization of the predicted biosynthetic gene cluster of metagenomic clone P17B06 119

Figure 3.4. Flowchart of the extraction and fractionation of bioactive compounds produced by *S. coelicolor* clone P17B06 121

Figure 3.5. Zones of inhibition against *A. baumannii* and *E. coli* of crude extracts from *S. coelicolor* clone P17B06 grown with and without OTC inducer 122

Figure 3.6. Zones of inhibition of parent fraction 2 against *A. baumannii* and *E. coli* 123

Figure 3.7. OPLS-DA of metabolites analyzed by UPLC-MS-QToF in crude extracts of *S. coelicolor* clone P17B06 and the empty vector clone 128

Figure 3.8. Representative overlay of base peak chromatograms from UPLC-MS analysis of crude extracts and active fraction from clone P17B06 129

Figure 3.9. Base peak chromatograms of four subfractions derived from *S. coelicolor* clone P17B06..... 131

Chapter 4

Figure 4.1. Annotation of the genes encoded in metagenomic clone P12B21 169

Figure 4.2. Genetic organization and annotation of the P12B21 insert sequence and schematic representation of deletion constructs 171

Figure 4.3. Antibacterial activity of *S. coelicolor* exconjugants harboring modified and unmodified P12B21 gene clusters 173

Figure 4.4. MAUVE alignment of biosynthetic gene clusters of *S. coelicolor* clone P12B21 and *Microbacterium* species..... 175

Figure 4.5. Agarose gel electrophoresis of PCR products from colony PCR screening of *S. coelicolor* clone P12B21..... 178

Chapter 1
Literature Review
The genus *Streptomyces* and its role in microbial natural product discovery

1.1. Abstract

Microbial natural products are a historically rich resource for therapeutic compounds such as antibiotics. More than two-thirds of the antibiotics deemed clinically useful originated from the bacterial genus *Streptomyces*, most of which were discovered in the 1950s to 1970s. Despite this early success, pharmaceutical companies lost interest in pursuing *Streptomyces* and other microorganisms after the 1990s for new drug leads largely due to technical challenges in screenings, isolation, and characterization. Meanwhile, new antibiotics have been increasingly needed to combat rising rates of drug-resistant infections. Recent advancements in culture-dependent and –independent techniques have revealed that the majority of the biosynthetic potential from microorganisms remains underexploited and the techniques to access this potential are constantly improving. This review begins with an introduction on the importance of microbial natural products and an overview on the development and production of bioactive metabolites by the genus *Streptomyces*. Subsequently, classical and modern strategies to discover microbial natural products and the role that *Streptomyces* has in these approaches is discussed.

1.2. Introduction

1.2.1. The Need for New Antibiotics

Although antibiotic resistance is not a new phenomenon, there has been a rapid increase in the emergence of resistant bacteria since the early 2000's (Alanis, 2005). More than 23,000 deaths per year in the United States are caused by drug-resistant infections alone (Jackson et al., 2018; Michael et al., 2014). Recent reports predict that 2.4 million people in North America,

Europe, and Australia will die in the next 30 years from antibiotic-resistant infections and health care-related costs in the U.S. could be as high as $3.5 billion per year if current patterns persist (Baltz, 2019; Walsh, 2004).

The reason for the significant increase in drug-resistant infections is multifactorial; however, the lack of new antibiotics in clinical development over the last 30 years is a significant contributor (Dutescu & Hillier, 2021; Talbot et al., 2019). Since 2017, 11 antibiotics have been approved by the FDA and only two of these represent a novel class of antibiotics (World Health Organization, 2020). The remaining approved antibiotics are derivatives of existing antibiotic classes making them less likely to stand the test of time against the evolution of resistance mechanisms by pathogenic microorganisms. Even more worrisome is that few of these new antibiotics target the World Health Organization's (WHO's) priority pathogens. A recent review by the WHO analyzed 43 antibiotics currently in clinical development and established that 26 antibiotics target some of the priority pathogens, but only two are active against the most concerning multidrug-resistant Gram-negative pathogens like *Acinetobacter baumannii* and *Pseudomonas aeruginosa* (World Health Organization, 2020). The WHO concluded in this report that "the clinical pipeline and recently approved antibiotics are insufficient to tackle the challenge of increasing emergence and spread of antimicrobial resistance" (World Health Organization, 2020).

To help address this public health crisis, agencies such as the WHO, the World Bank, and national governments have called for a global commitment for research and development on new antibiotics (Dutescu & Hillier, 2021; World Health Organization, 2020). Historically, the antibiotic development pipeline was led by pharmaceutical firms, but in the last 30 years, 15 of the 18 largest pharmaceutical companies have abandoned antibiotic research and development

due to scientific, regulatory, and economic challenges (Dutescu & Hillier, 2021; Talbot et al., 2019). Instead, academic labs and smaller companies have taken the mantle for antibiotic discovery (Hutchings et al., 2019) and these researchers are revisiting the most historically important source of new drugs – microbial natural products.

1.2.2. Microbial Natural Products as a Source of Antibiotics

Although the term "natural product" is broadly defined as any naturally occurring substance, natural products are synonymous with secondary metabolites (chemical compounds) synthesized by living organisms that are not directly required for the growth, development, or reproduction of the organism (Dias et al., 2012). The production of secondary metabolites by an organism have been shown to be critical in self-defense, nutrient scavenging, and virulence making them attractive candidates for therapeutic compounds (Hanson, 2003). Natural products are well-known sources of secondary metabolites with complex chemical structures and diverse functionalities, and the most well-characterized classes include polyketides and nonribosomal peptides (Williams, 2013).

The discovery of penicillin in 1928 from the fungus *Penicillium notatum* (Fleming, 1929) sparked the beginning of the golden age of microbial natural product discovery and revolutionized modern medicine (Shen, 2015). Microorganisms were the source of nearly 90% of all antibiotics widely prescribed today including tetracyclines, cephalosporins, aminoglycosides, lipopeptides, glycopeptides, and macrocyclic compounds (Newman & Cragg, 2012). Beyond antibiotics, microorganisms were found to be an important source of therapeutic drugs with antifungal (amphotericin), antiparasitic (ivermectin), antitumor (doxorubicin), and antichloesterolemic (lovastatin) activities (Newman & Cragg, 2012). Thus, microbial natural

products became recognized for their immense diversity in structure, function, and biosynthesis which contributed to a dramatic reduction in worldwide mortality (Nakashima et al., 2018).

Within the extensive diversity of the kingdom *Bacteria*, more than half of all antibiotics currently in clinical use were derived from natural products produced by members of the phylum Actinobacteria, most notably the genus *Streptomyces* (Procópio et al., 2012; Rashad et al., 2015). Actinobacteria represent one of the largest phyla among *Bacteria* with over 400 genera (Salam et al., 2020). This diverse phylum is characterized by a high guanine plus cytosine content in their genome (> 55%), a ranging genome size from 0.93 Mb (*Tropheryma whipplei*) to 12.7 Mb (*Streptomyces rapamycinicus*) (Shivlata & Tulasi, 2015), and the ability to undergo complex morphological differentiation (Barka et al., 2016). Although the discovery of drugs from this taxa significantly declined since the 1970s, new advancements in culture-dependent and –independent technologies have revealed that we have only scratched the surface of the biosynthetic potential of cultured and uncultured microorganisms (Devine et al., 2017). These findings have revitalized interest in microbial natural product discovery and research efforts are demonstrating that prolific producers like *Streptomyces* still have a major role in drug discovery.

Herein, the biology and development of *Streptomyces* is briefly summarized to provide context to their remarkable ability to produce diverse, bioactive secondary metabolites. Thereafter, this review discusses the classical and modern approaches to discover microbial natural products with emphasis on how *Streptomyces* has and continues to play an important role in microbial drug discovery.

1.3. Understanding the genus *Streptomyces*

Streptomyces represents the best studied genus among the vast phylogenetic diversity of the phylum Actinobacteria, largely due to their proficient production of bioactive natural

products (Seipke et al., 2012). In fact, 75% of clinically available antibiotics are derived from *Streptomyces* (Kemung et al., 2018) and new secondary metabolites with biological activities continue to be discovered from this genus. This Gram-positive, filamentous genus was first proposed by Waksman and Henrici in 1943 (Waksman & Henrici, 1943) and has since become the largest genus of the phylum with more than 900 validly described species (http://www.bacterio.net, 25 June 2021, date last accessed). Although the discovery of this genus is relatively recent, a study by McDonald and Currie (2017) suggests that streptomycetes are ancient, originating at the same time period as land plants nearly 380 million years ago. Most *Streptomyces* spp. have been isolated from soils where they are known to play a significant role in organic matter turnover (Barka et al., 2016). Research within the last decade has shown that these ecologically and medically important bacteria are widely distributed terrestrially and can also be found in marine environments and host-associated microbiomes (Chevrette, Carlos-Shanley, et al., 2019).

Perhaps one of the most defining features of streptomycetes is their complex developmental life cycle which is more similar to that of filamentous fungi than other bacteria (Hopwood, 2007; Lapaz et al., 2019). Early observations of these bacteria found that they formed a mat of branching cells or "hyphae", but the size of the cells were smaller than other fungal species leading Waksman and Henrici to name them *Streptomyces* meaning "twisted fungus" (Waksman, 1953). Secondary metabolite production in these organisms is linked closely to their life cycle as they undergo differentiation from spores to aerial hyphae (de Jesus Sousa & Olivares, 2016). The most well understood mechanisms of development and secondary metabolism is in the model organism *Streptomyces coelicolor*.

1.3.1. *Streptomyces* Development and Physiology

The life cycle of *S. coelicolor* and most streptomycetes typically begins when environmental conditions are favorable leading to the germination of spores which grow outward to form vegetative hyphae (de Jesus Sousa & Olivares, 2016; Manteca & Yagüe, 2019). Unlike most other unicellular bacteria, cells of *Streptomyces* and other filamentous Actinobacteria do not divide in the middle (Dyson, 2009). Instead, the vegetative hyphae grows by tip extension, and branching at the cell pole is thought to be regulated by the protein complex DivIVA (de Jesus Sousa & Olivares, 2016; Dyson, 2009). Streptomycetes are an interesting example of a multicellular bacterium since the cellular division process of the vegetative hyphae results in multinucleoid compartments (Willemse et al., 2011). As cell division continues, a vegetative mycelium forms that functions to acquire nutrients from the surrounding environment (Lapaz et al., 2019). When environmental conditions become adverse, such as when nutrient availability becomes depleted, the vegetative mycelium undergoes programmed cell death and serves as a substrate to form aerial hyphae/mycelium (Figure 1.1) (de Jesus Sousa & Olivares, 2016; Flärdh & Buttner, 2009). At this stage in development, there are two populations of cells: the substrate mycelium that is attached to the surface and the aerial mycelium/hyphae that extend outward (Flärdh et al., 2012).

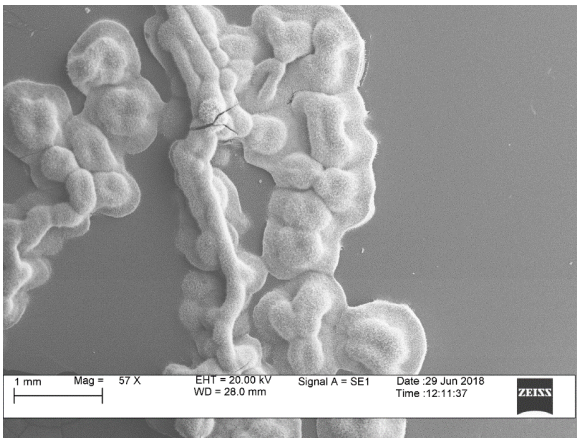

Figure 1.1. Scanning electron micrograph (SEM) of a *Streptomyces* species showing growth of both vegetative and aerial mycelium.

Aerial hyphae of *S. coelicolor* have a characteristic fluffy appearance, often pigmented, and are covered with hydrophobic proteins (Lapaz et al., 2019). This sheath of hydrophobic proteins typically consists of rodlins, chaplins, and SapB that allow the aerial hyphae to break through the surface tension of the medium and grow out from the vegetative environment (Flärdh & Buttner, 2009). Rodlins and chaplins are predicted to work together to form the sheath, but it is not known if or how the surfactant peptide SapB interacts with the other components (Claessen et al., 2003; Elliot et al., 1998). Rodlins are not universally conserved among all streptomycetes (Claessen et al., 2004). For example, *Streptomyces avermitilis* lacks the rodlin genes but can still develop normal aerial hyphae likely due to a larger role from the chaplins and SapB proteins (Claessen et al., 2004). In addition to the role of chaplins and SapB in initiating differentiation, a group of genes called the *bld* genes (*bldA* through *bldN*) encode regulatory proteins that work through a signaling pathway that leads to aerial development (Claessen et al., 2006). Mutants of *S. coelicolor* that lack the *bld* genes do not have the typical fuzzy features of the wild type and appear "bald" (Elliot et al., 1998). Interestingly, these mutants typically fail to produce antibiotics suggesting that this stage of morphological differentiation is linked closely to their ability to produce bioactive, secondary metabolites (Bibb, 2005).

Similar to vegetative hyphae, aerial hyphae grow by tip extension until sufficient biomass is reached and the apical cells then differentiate into a spore chain of up to 100 uninucleoid compartments (Manteca & Yagüe, 2019). In the final stages of the life cycle, multiple rounds of cell division lead to septation by FtsZ and spores are produced that can disperse to new environments (Flärdh, 2003). The process from aerial hyphae to spores is regulated by the *whi* genes (*whi*A through *whi*J) whose name designation stems from the white morphology typical of

whi mutants in *S. coelicolor* that are not capable of forming the common gray-pigmented spores (Chater, 2016; McCormick & Flärdh, 2012).

It is worth noting that although significant progress has been made to understand the three core developmental stages (vegetative hyphae, aerial hyphae, and spores) described here, the complex processes governing regulation and control are still not fully understood. In particular, how regulatory proteins interact to lead to aerial hyphae formation is unclear but the signaling *bld* cascade has been proposed to aid in the development (Nodwell et al., 1999). Moreover, recent research suggests that some *Streptomyces* species are capable of an additional developmental stage, exploratory cells, which when triggered by co-culture with fungi, can form non-branching vegetative hyphae that can transverse abiotic surfaces (S. E. Jones et al., 2017).

1.3.2. Biosynthesis of secondary metabolites in *Streptomyces*

The last two phases of the *Streptomyces* life cycle, aerial hyphae formation and differentiation into spores, are closely linked to the production of secondary metabolites in these organisms (de Jesus Sousa & Olivares, 2016). Microbial secondary metabolites, also referred to as specialized metabolites, are not directly involved in the primary metabolic processes of the producing organism but often provide advantages including nutrient acquisition, defense mechanisms, signaling agents and communication (Craney et al., 2013). In *Streptomyces*, as nutrients become limiting and the vegetative mycelium undergoes programmed cell death (PCD), the onset of aerial hyphae begins and it is during this phase that secondary metabolites are most often produced (Barka et al., 2016). The colony is particularly vulnerable during this time and it is thought that these secondary metabolites, which often have antimicrobial properties, may help defend the colony and their resources from competing microorganisms (Barka et al., 2016).

In support of this prediction, the accumulation of *N*-acetylglucosamine (GlcNAc) after PCD of the vegetative mycelium has been shown to trigger the onset of secondary metabolism. Supplementation with GlcNAc can accelerate the onset of antibiotic production and aerial hyphae development under poor nutrient conditions (Barka et al., 2016; Rigali et al., 2008). Production of secondary metabolites has also been proposed to help remove the dying cell population during PCD in order to provide nutrients for the developing cells. Such an example has been shown in *S. coelicolor*, where the production of prodiginines was triggered after the first round of cell death and occurred in the dying zone of the mycelial network (Tenconi et al., 2018). More research is needed to understand the environmental and morphological cues that trigger *Streptomyces* antibiotic production which could in turn provide more insight towards activating these pathways. For a more exhaustive review of the relationship between stress and antibiotic production in *Streptomyces* see Rigali et al. (2008).

In *Streptomyces* and other microorganisms, the biosynthesis of secondary metabolites, like antibiotics, is typically governed by the expression of genes that cluster together in the genomes of the producing organism termed “biosynthetic gene clusters” or BGCs (Bibb, 2005). BGCs can vary in size from a few thousand bases to over 100 kb and often contain and tightly control all the genes responsible for the biosynthesis, secretion, and regulation of a given metabolite (R. Chen et al., 2019). Genes encoding activators or regulators are usually present within the BGC and are capable of receiving an environmental cue and subsequently triggering expression of downstream genes (G. Liu et al., 2013). The specific environmental cues are difficult to identify, and most demonstrations of antibiotic production from BGC expression are linked to pH fluctuations, nutrient depletion, metal composition, or dissolved oxygen levels (Rigali et al., 2018).

Once an environmental signal is received by *Streptomyces* cells, a transcriptional response is activated that is often controlled in *cis*-regulatory elements targeted by transcription factors (Y. Lee, Lee, et al., 2020). This system is regarded as a "lock-and-key" mechanism where specific activators can switch on the expression of a BGC, or repressors can lock the transcriptional response (Rigali et al., 2018). The regulators involved in secondary metabolite production are immensely complex and very few regulatory networks are fully understood. The first molecular analyses of BGCs revealed that the clusters can encode "cluster-situated regulators" which control the transcription of neighboring genes within a BGC and can dramatically affect the production levels of secondary metabolites like antibiotics (J. Wei, He, et al., 2018). Since then, some pathway-specific regulators encoded by a BGC have been found to affect global transcription patterns by controlling expression in other clusters such as the regulation of tylosin in *Streptomyces fradiae* (Rigali et al., 2018; Stratigopoulos & Cundliffe, 2002).

Among the published literature on the regulation of BGCs, known antibiotic producers like *S. coelicolor* are often utilized to investigate the role of the phosphorylated guanosine nucleotide (p)ppGpp in triggering natural product biosynthesis (Bibb, 2005). Under conditions of nitrogen limitation in *S. coelicolor*, the ribosome-associated ppGpp synthetase (RelA) was required for BGC expression (Chakraburtty & Bibb, 1997). When a modified *relA* gene was used to induce ppGpp synthesis in *S. coelicolor*, transcription of the regulatory gene for the antibiotic actinorhodin occurred (Bibb, 2005; Hesketh et al., 2001). Deletion of *relA* gene has also been shown to suppress antibiotic production which further emphasizes its role in BGC expression (Chakraburtty & Bibb, 1997). However, the exact mechanism that is responsible for activation of ppGpp remains poorly understood (Barka et al., 2016).

Following expression of a BGC, enzymatic complexes encoded by the gene cluster work to build a functional secondary metabolite by incorporating precursors from primary metabolism. Modification enzymes encoded by the BGC are responsible for adding chemical groups (sugars, hydroxyls, etc.) to the chemical scaffold to render the metabolite functional (Wohlleben et al., 2016; Zotchev, 2014). The resulting secondary metabolite can be harmful to the host, so BGCs typically encode for mechanisms to avoid self-toxicity either by active efflux of the metabolite, by a specific transporter, or by the use of resistance proteins that inactivate the metabolite inside the cell (Ziemert et al., 2016). The specific mechanisms for biosynthesis and efflux can vary depending on the type of secondary metabolite produced and the host organism. Different enzyme complexes encoded by BGCs can use different substrates and modify final products via unique chemical reactions leading to significant diversity in metabolites produced by biosynthetic pathways (Seyedsayamdost, 2019). *Streptomyces* spp. can harbor more than 20 BGCs on average in their genomes, thus making them a valuable source for secondary metabolites with wide structural diversity and interesting bioactivities (Nguyen et al., 2020). The latter half of this review aims to discuss the importance of the genus *Streptomyces* in past and future trajectories of natural product discovery.

1.4. Streptomyces' role in past approaches for microbial natural product discovery

The "accidental" discovery of penicillin from *P. notatum* in 1928 (Fleming, 1929) marked the starting point for natural product isolation from microbial sources (Shen, 2015). Following this epiphany, Selman Waksman and colleagues at Rutgers led the next major discoveries in the 1940s. Waksman is credited with being the first to systematically isolate microorganisms and screen for novel natural products (Milshteyn et al., 2014). His work, along with his graduate students, led to the isolation of several important compounds all derived from

Streptomyces spp. including actinomycin from *Streptomyces antibioticus* in 1940 (Waksman & Woodruff, 1940), streptothricin from *Streptomyces lavendulae* in 1942 (Waksman & Woodruff, 1942), and streptomycin from *Streptomyces griseus* in 1944 (Schatz et al., 1944). These discoveries piqued the interest of pharmaceutical companies and most of the early years of natural product research during the antibiotic "Golden Age" from the 1940s to 1970s centered on the isolation of actinomycetes, particularly *Streptomyces* and fungi (L. Katz & Baltz, 2016).

The first step in these traditional "top-down" approaches was the acquisition of microorganisms, typically collected from soil samples by various research and pharmaceutical groups scattered throughout different countries (Luo et al., 2014). It is believed that every country had at least one or more soil samples collected during this time for screening (L. Katz & Baltz, 2016). Pure-culture colonies were then grown in shake flask fermentations and crude extracts from the microbial producers were subsequently screened for antibacterial or antifungal activity without knowing the drug target. These phenotypic screenings were usually performed by disc diffusion assays where a small sample of extract was spotted on a filter paper disc and placed on agar plates inoculated with a test organism to screen for growth inhibition (L. Katz & Baltz, 2016). After active extracts or "hits" were identified, pure compounds were isolated for structural characterization by chromatography and analyzed by nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS) (Wohlleben et al., 2016).

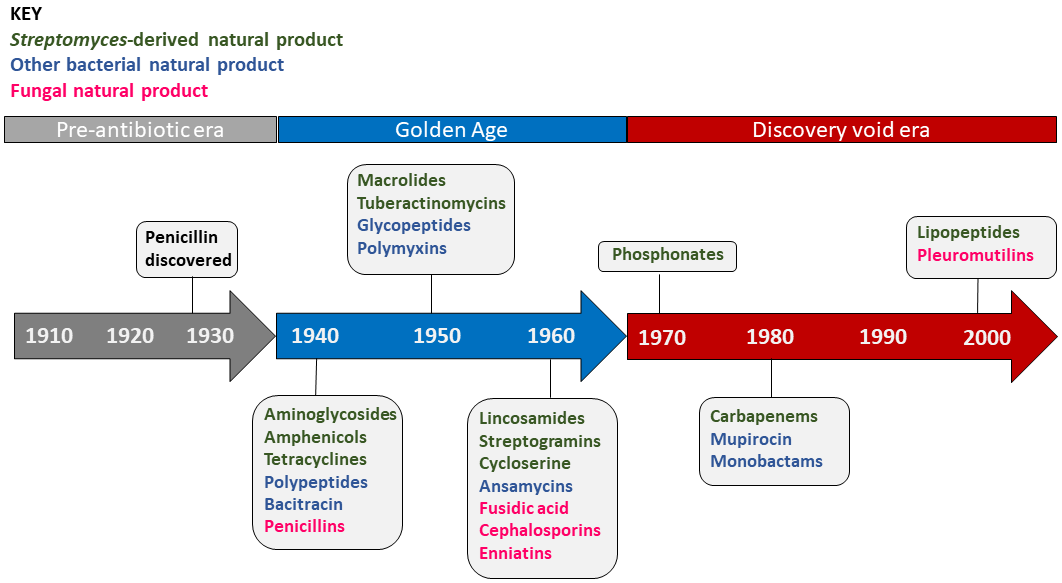

Figure 1.2. Timeline of the discovery of different antibiotic classes and their origin. Green indicates an antibiotic class derived from *Streptomyces*, orange indicates an origin from other bacteria, and purple indicates an origin from fungi. Synthetic compounds and their derivatives are not shown in this figure. This figure was adapted from Hutchings et al. (2019).

In their infancy, the straightforward approach of bioactivity-guided screenings was very successful having yielded numerous antibiotic compounds and their derivatives that are still in use today. A famous example was the discovery of avermectin from *Streptomyces avermitilis* originally isolated in the 1970s by Professor Satoshi Ōmura after collecting a soil sample from a golf course in Tokyo (L. Katz & Baltz, 2016). Pharmaceutical companies were actively leading the global search for antibiotic-producing microbes and their combined efforts led to the discovery of half of the drugs commonly used today (Gould, 2016). The number of medically important antibiotic compound classes discovered from *Streptomyces* isolates during the "Golden Age" surpassed those originating from other bacteria and included the macrolides, aminoglycosides, amphenicols, tetracyclines, tuberactinomycins, lincosamides, streptogramins, and cycloserine (Figure 1.2) (Hutchings et al., 2019). Despite this success, there was a significant

decline in the number and rate of new natural products discovered after the 1960s largely due to the constraints of classical screening methods.

1.4.1. Shortcomings of classical methods to discover microbial natural products

After the 1960s, one of the major challenges that faced classical antibiotic screening approaches was the high rate of rediscovery of known compounds (Hover et al., 2018). It is estimated that over 10 million Actinobacteria isolates were screened prior to 2000, and those species that were highly abundant in soil samples and easy to culture were likely screened multiple times by different researchers (J. W.-H. Li & Vederas, 2009). As a result, many bioactive compounds became repeatedly discovered to the point where it was assumed that most of the accessible chemical scaffolds for antibiotics from the environment had already been extracted.

Other limiting factors were that classical bioactivity-guided screenings were laborious, low-throughput, and often not sensitive enough to detect low abundance compounds (Luo et al., 2014). Growing microorganisms in shake flask fermentations took time and space that became less enticing as the rate of compound rediscovery increased. Potentially unique compounds that were in low abundance or hidden by more abundant known compounds would often go undetected due to the lack of sensitivity of classical bioassays and MS instruments (Luo et al., 2014). To help address these pitfalls, new screening approaches were adopted in the late 1980s in hopes of improving natural product discovery rates.

Target-based screening became a common method used after the “Golden Age” in the later 1970s to 1980s. Contrary to phenotype screenings, target-based screening sought to identify compounds that acted directly on a known target such as an enzyme or receptor that were important in the pathology of interest (Wohlleben et al., 2016). Although this approach did

identify some beneficial drugs like lovastatin from fungi and clavulanic acid from *Streptomyces clavuligerus*, very few natural products were discovered and developed into drugs compared to classical screening approaches. Only one novel antibacterial natural product, daptomycin from *Streptomyces rosesporus*, was discovered in the period of 1970 to 2000 by this approach (L. Katz & Baltz, 2016).

Combinatorial chemistry was another approach adopted in the 1990s where massive compound libraries were created synthetically based on existing chemical scaffolds (Dias et al., 2012). These libraries could be screened in a high-throughput manner, which was an improvement to traditional bioassays that were often limited, by screening small numbers of extracts at a time. Millions of compounds were screened between 1990 to 2000; however, very few compounds were discovered from this approach and no new drugs were approved (L. Katz & Baltz, 2016). The "hit-rate" for identifying new drugs from classical screenings was near 0.3%, and although the percentage is seemingly small, the rate was much higher than the < 0.001% hit rate for screenings of synthetic compound libraries (J. W.-H. Li & Vederas, 2009). By the end of the 20th century, many pharmaceutical companies abandoned the search for new drugs from microorganisms. This disinterest led to a significant decline in the number of new antibiotics on the market, yet the rate of antimicrobial resistance was rising (Genilloud, 2017).

1.5. Modern approaches for natural product discovery from *Streptomyces* and other bacteria

A renaissance in natural product discovery was sparked after the 2000s with the advent of genomics, powerful bioinformatics and analytical methods, and advances in genetic manipulation (Carini, 2019). A wealth of genetic data from environmental microorganisms became available and the computational tools to mine genomes for important biosynthetic

pathways were now readily accessible (Zerikly & Challis, 2009). In what became known as the "great plate count anomaly" by Staley and Konopka (1985), scientists realized that more than 99% of microorganisms could not be easily isolated and cultured in the laboratory using classical methods (Carini, 2019). This finding indicated that classical methods were missing a large percentage of microbial diversity in nature and leaving behind a potential trove of natural products.

The next major influence in the revitalization of microbial natural product discovery came from genome sequencing of *S. coelicolor* by Bentley et al. (2002). This revealed more than 20 BGCs encoding secondary metabolites in the genome of *S. coelicolor*, which was 4x as many as previously assumed to be comprehensive of the natural product potential of this organism (Amara et al., 2018). High-throughput genome sequencing has since revealed that most secondary metabolite BGCs are "silent" or "cryptic" under pure culture laboratory conditions (Kim et al., 2021). For example, more than 11,000 natural product BGCs were identified from a study of 830 actinomycete genomes and the majority of these have yet to be experimentally validated (Doroghazi et al., 2014). These findings strongly indicated that there are many natural products waiting to be discovered from nature, but the challenge is accessing them.

New approaches have been employed in the post-genomic era to expand natural product discovery efforts while seeking to avoid high rediscovery rates that plagued the prior decade (Stefani et al., 2015). These approaches can broadly be classified into two categories: 1) culture-dependent approaches which largely focus on isolating novel microorganisms from a wider spectrum of habitats and coercing these novel isolates to produce new compounds, and 2) culture-independent or "metagenomic" approaches that rely on the expression of DNA that originated from uncultured microorganisms in a culturable host (M. Katz et al., 2016). Though

different in their methodologies, some of the most successful applications of these approaches have involved *Streptomyces.*

1.5.1. Culture-dependent strategies

The repeated rediscovery of known microorganisms and compounds after the 1960s meant that classical culture-based methods were no longer sufficient to discover new natural products. Early attempts to improve the cultivability of novel microorganisms adopted the OSMAC (One Strain MAny Compounds) approach where easily accessible cultivation parameters were changed to promote *in vitro* growth of an organism and affect their secondary metabolism (Kim et al., 2021). By changing culture conditions (media composition, temperature, pH, osmotic stress, etc.), a single strain could be coerced to activate cryptic BGCs and produce new compounds (Stewart, 2012). The OSMAC approach has been successfully employed to stimulate production of novel compounds; for example, Q. Wu et al. (2018) altered the media type of a marine-derived *Streptomyces* sp. YB104 and stimulated production of a bioactive compound inthomycin B. Unfortunately, the empirical nature of the OSMAC approach requires extensive laboratory experiments and still fails to address the overarching problem of slow rates of novel compound discovery.

More promising culture-dependent approaches in the post-genomics era have sought to combine the classical "top down" strategy of characterizing the biology and chemistry of an organism with the new "bottom up" methods of using genetic information to evaluate the biosynthetic potential of microorganisms (Schneider, 2021). Advances in bioinformatics has enabled the use of genome mining to evaluate the biosynthetic potential of diverse *Streptomyces* spp. and prioritize the most promising strains (N. Lee, Hwang, et al., 2020). Innovations in cultivation strategies, such as the maintenance of co-cultures, have been used successfully to

isolate novel *Streptomyces* and their metabolites by more closely simulating their natural environment (Lodhi et al., 2018). Additionally, researchers are now looking beyond soils and terrestrial environments for new *Streptomyces* and are instead searching in underexplored habitats and interspecies symbiotic associations, thus instilling exciting integrative, coevolutionary, and ecological aspects to the field of natural product discovery. The following sections of this review describe a few of the most common culture-dependent approaches but are not an exhaustive compendium of all the strategies currently being employed.

1.5.1.1. Genome mining for biosynthetic gene clusters

Next-generation sequencing has expanded the field of natural product discovery and fostered a tremendous increase in the number of available microbial genome sequences that can be mined for secondary metabolite BGCs (Milshteyn et al., 2014). For example, as of 2021, a total of 4,734 *Streptomyces* genomes were deposited in the NCBI database (https://www.ncbi.nlm.nih.gov/genome/ accessed on 7 November 2021). This wealth of genomic sequencing data has driven the development of bioinformatics tools capable of *in silico* predicting and annotating BGCs from assembled genomes or metagenomes (Khater et al., 2016). Some of the most popular tools include antiSMASH (Blin et al., 2021), PRISM (Skinnider et al., 2017), CLUSEAN (Weber et al., 2009), RiPPMiner (Agrawal et al., 2017), and ClustScan (Starcevic et al., 2008). Among all of these, antiSMASH is the most well-known, largely due to its user friendly interface and ability to predict up to 71 different categories of BGCs (Blin et al., 2021). AntiSMASH relies on identifying highly conserved sequences within a BGC using a sequence alignment-based profile in a Hidden Markov Model (HMM) (N. Lee, Hwang, et al., 2020). Gene clusters predicted from this tool, and others like PRISM, are then compared to a database of known clusters such as the MIBiG database (Kautsar et al., 2020). Moreover,

antiSMASH is capable of predicting chemical scaffolds from a biosynthetic pathway (Milshteyn et al., 2014).

Genome mining is often a first step towards isolating a new natural product in the modern era. A typical workflow consists of identifying a gene cluster of interest using the previously mentioned tools and subsequently targeting the gene cluster for expression through the use of advanced molecular biology techniques. This approach has been used to discover numerous natural products. By a genome mining approach, a RiPP gene cluster was predicted from the genome of *Streptomyces curacoi* and led to the discovery of a new cytotoxic peptide called curacozole (Kaweewan et al., 2019). Genome mining of *Streptomyces* sp. Tü 6314 identified a cryptic type II polyketide synthase (PKS) gene cluster that was subsequently expressed in *S. coelicolor* and led to the production of three new anti-HIV compounds named streptoketides (Qian et al., 2019). Genome sequencing and mining by antiSMASH of *Streptomyces* sp. Tü 4128 identified a putative bagremycin BGC which enabled the isolation of the antibacterial compounds bagremycin A and B (Ye et al., 2019). For a more detailed list of natural products discovered by genome mining of *Streptomyces*, the reader is pointed to N. Lee, Hwang, et al. (2020).

There are several advantages to genome mining approaches including more efficient identification of biosynthetic pathways, ease of use, cheap costs, and the ability to predict chemical structures of natural products (Albarano et al., 2020). However, genome mining is still in its infancy and there are major bottlenecks that need to be addressed. The most accurate BGC prediction relies on complete genome assemblies with high quality sequences since genes in a BGC are often scattered throughout contigs and can contain repetitive sequences (N. Lee, Hwang, et al., 2020). This emphasizes the need for improvements to long-read technologies that

will allow researchers to obtain complete, high-quality genome sequences. Additionally, rule-based tools like antiSMASH are limited to predicting similar gene clusters to known pathways (Zerikly & Challis, 2009). ClusterFinder (Cimermancic et al., 2014) and DeepBGC (Hannigan et al., 2019) are two machine learning packages that have been developed to predict unknown BGCs but these tools still suffer from high false-positive rates (N. Lee, Hwang, et al., 2020). Thus, new bioinformatic tools are needed that are capable of more accurately detecting and annotating completely novel BGCs.

1.5.1.2. Exploring "underexplored" environments for new species

Historically, soil ecosystems have been the primary focus for isolation of antibiotic-producing microorganisms. Therefore, one strategy to improve the hit rate of natural product discovery has been to sample from "underexplored" or less sampled ecosystems including extreme environments, marine ecosystems, and caves (Sivalingam et al., 2019). In the modern era of next-generation sequencing, this approach often combines classical methods to isolate novel microorganisms followed by genome sequencing and mining using the technologies described in the previous section. By looking in less explored environments, many novel microorganisms including *Streptomyces* have been isolated and found to produce new compounds with diverse bioactivities (Kemung et al., 2018).

Among all of the underexplored habitats, marine ecosystems represent one of the richest habitats for microbial diversity (Subramani & Aalbersberg, 2013). *Streptomyces* and other actinomycetes are one of the prominent taxa within these marine ecosystems where they can form stable populations and respond quickly to environmental heterogeneity (Subramani & Aalbersberg, 2012). Similar to terrestrially-derived species, the marine-derived microorganisms are capable of producing unique secondary metabolites to compete for defense and survival

(Petersen et al., 2020). As such, marine environments have been the source of many new compounds that are either already on the pharmaceutical market or in development. In fact, since the 1960s more than 20,000 natural products have been identified from marine environments and is estimated to only be a small portion of the chemical diversity that exists (Hu et al., 2011). Some examples of new natural products isolated from marine *Streptomyces* include the antibacterial tirandamycins from *Streptomyces* sp. 307-9 (Carlson et al., 2009), the antibacterial meroterpenoids from *Streptomyces* sp. CNH-189 (Kaysser et al., 2012), and the recently discovered zhaoshumycins A and B from *Streptomyces* sp. ITBB-ZKa6 with cytotoxic activity (Guo et al., 2021). An extensive review of marine drugs from *Streptomyces* and other microorganisms is provided by Subramani and Aalbersberg (2012).

Marine environments are not the only underexploited source for new drugs. Antimicrobial metabolite-producing *Streptomyces* have been recently isolated from extreme habitats including deserts, arctic regions, volcanoes, and high-altitude mountains (Sivalingam et al., 2019). For example, *Streptomyces* sp. strain C34 was isolated from the Atacama Desert which led to the discovery of the antibacterial chaxalactins (Rateb, Houssen, Harrison, et al., 2011) and chaxamycins (Rateb, Houssen, Arnold, et al., 2011). The antibacterial and cytotoxic peptides called ohymyungsamycins were isolated from *Streptomyces* sp. SNJ042 sampled from Korean volcanic rock (Um et al., 2013), and the antibacterial compound 2-amino-3-dodecanol was found to be produced by the arctic-derived *Streptomyces avidinii* SB9 (Ivanova et al., 2010). Recent studies have found caves to be another source of novel bacteria. A study by Hamm et al. (2017) identified 25 *Streptomyces* species, 15 of which were novel, from five cave systems. Interesting from an ecological perspective, these isolates inhibited the fungus *Pseudogymnoascus destructions* that causes white-nose syndrome in bats.

As evident by the examples provided above, there is tremendous potential to find novel bacteria that produce unique natural products from marine ecosystems and other habitats. Given the immense genomic and metabolic capabilities of *Streptomyces*, finding novel species of this taxa from diverse environments should continue to be a priority. Innovations in cultivation strategies will aid in isolating more rare species and advancements in genome mining technologies will improve the ability to study their full biosynthetic potential in order to find new drug leads.

1.5.1.3. *Streptomyces* symbiotic interactions

Streptomyces are now recognized as symbiotic partners with eukaryotic hosts including humans, invertebrates, and plants (Seipke et al., 2012). In fact, it is thought that the chemical diversity of *Streptomyces* has evolved as a result of these interactions within *Streptomyces* and other symbiote species (Manteca & Yagüe, 2019). The secondary metabolites produced by these organisms can influence their interactions within a microbial community, host, and environment. These interactions can range from commensalism to antagonism to mutualism. There are many well-documented examples of mutualistic symbiotic associations where *Streptomyces* provide a host and/or the niche of a host with chemical defense through the production of bioactive metabolites. In return, the bacteria is provided with shelter and nutrition (Chevrette, Carlson, et al., 2019; Seipke et al., 2012).

Invertebrates such as fungus-growing ants, beewolves, termites, beetles, and marine sponges are the most studied co-evolutionary systems for natural product-producing bacteria which are typically from the genus *Streptomyces*. A new antimicrobial polyene called cyphomycin with activity against multidrug resistant fungal pathogens was isolated from a *Streptomyces* species associated with the fungus-growing ant *Cyphomyrmex* and is thought to

ward off the fungus *Escovopsis* that can infect ant farms (Chevrette, Carlson, et al., 2019). The polyketide formicamycins with activity against MRSA and VRE were found to be produced by a new *Streptomyces* species, *Streptomyces formicae*, living in association with African plant-ants (Qin et al., 2017). Four new antibacterial derivatives of anthraquinones, termstrin A, B, C, and D, were isolated from the termite-associated *Streptomyces* sp. BYF63 (L. Zhang, Song, et al., 2020). Two *Streptomyces* symbionts were isolated from *Dendroctonus frontalis* southern pine beetles and found to produce the polyketide mycangimycin (Oh et al., 2009) and macrolactams frontalamides A and B (Blodgett et al., 2010) that inhibit fungi that are pathogenic to the insect. These compounds were also demonstrated to have anti-malarial activity and antifungal activity, respectively (Manteca & Yagüe, 2019). Insects have also been shown to use bioactive metabolites produced by *Streptomyces* to protect their brood against infection. One such example is the beewolf that incorporates *Streptomyces* in their brood cell wall and these symbionts produce diverse antimicrobial compounds including the antifungal piericidin and the antibacterial compound streptochlorin (Kroiss et al., 2010; Van der Meij et al., 2017).

As mentioned previously, *Streptomyces* are prominent in marine habitats and this is not limited to solely sediments and seawater. Interestingly, antimicrobial-producing *Streptomyces* have been isolated from other non-terrestrial invertebrates, most notably marine sponges. The marine invertebrates often harbor the symbionts in their mesohyl matrix where they are thought to protect the host from infection and predation through the production of secondary metabolites (Abdelmohsen et al., 2014; Nair et al., 2011). Several examples of new natural products have been found from sponge-associated *Streptomyces* species including mayamycin from *Streptomyces* sp. strain HB202 (Schneemann et al., 2010), tetromycins 1-4 isolated from *Streptomyces axinellae* PoI001 (Abdelmohsen et al., 2014; Pimentel-Elardo et al., 2011),

lobophorin C and D from *Streptomyces carnosus* strain AZS17 (R.-B. Wei et al., 2011), and urauchimycin A and B from *Streptomyces* sp. Ni-80 (Imamura et al., 1993).

These symbiotic interactions with *Streptomyces* and invertebrates are thought to be far more widespread than characterized to date (Seipke et al., 2012). Notably, such examples may underscore the potential for *Streptomyces* symbiosis and secondary metabolite production to affect eukaryotic fitness and, thus, play a role in the evolution of animal life histories (McFall-Ngai et al., 2013). It is clear from the examples described thus far that soil *Streptomyces* are no longer the only hotspot for natural product isolation. Future efforts should invest in exploring new eukaryotic microbiomes for novel *Streptomyces* species, particularly from diverse insects and marine sponges. By "looking" beyond soil habitats, these efforts will help reduce the likelihood of compound rediscovery and may also yield more fruitful results from clinical screenings since these compounds may be less toxic to animal and human hosts (Chevrette, Carlson, et al., 2019).

1.5.1.4. Coculture of *Streptomyces*

Beyond searching in new environments and hosts for novel bacteria, there have been advancements in cultivation techniques that aim to modulate secondary metabolism and promote antibiotic production. One such strategy is called cocultivation which is a technique where two or more organisms are cultured together. This approach seeks to replicate the natural environment of bacteria that is often dynamic with many inter- and intraspecies interactions (Kim et al., 2021). Recent studies have shown that this technique can result in a competitive environment, thereby leading to induction of unexpressed pathways due to interspecies crosstalk or chemical defense mechanisms (Pettit, 2009). One organism might stimulate secondary metabolism in another organism by the secretion of chemical signals or physical interactions, or

one organism can induce biosynthesis of signaling molecules which stimulate production of cryptic natural products in another organism (Scherlach & Hertweck, 2009).

During *Streptomyces*-*Streptomyces* coculture, interspecies interactions are often mediated by small signaling molecules, such as butyrolactones and siderophores, that can trigger secondary metabolism (Kim et al., 2021). Traxler et al. (2013) showed that siderophores made by different *Streptomyces* spp. and *Amycolatopsis* sp. AA4 induced production of the antibiotics γ-actinorhodin, prodiginine, and at least 12 different desferrioxamines. Although *Streptomyces*-*Streptomyces* interactions seem to be the most widely studied, some of the most successful applications of coculture have actually come from the combination of *Streptomyces* with other distantly related bacteria.

When two Actinobacteria taxa, *Rhodococcus fascians* and *Streptomyces padanus*, were cultured together, new aminoglycoside antibiotics called rhodostreptomycins were produced. These rhodostreptomycin antibiotics were not detected in pure cultures of either bacterium indicating that the mixed fermentation was required for production (Kurosawa et al., 2008). Several new antibiotics, granaticin, granatomycin D, and dihydrogranaticin B, were isolated after co-culture of a *Streptomyces* sp. strain PTY08712 with three human pathogens – *Bacillus subtilis*, *Staphylococcus aureus*, and *Pseudomonas aeruginosa* (Sung et al., 2017). Another example is that of the marine *Streptomyces cinnabarinus* which produced the diterpene lobocompactol after co-culture with *Alteromonas* sp. belonging to the *Proteobacteria* phylum (Cho & Kim, 2012). Coculture of a marine *Streptomyces* sp. with *Bacillus* sp. led to the discovery of a new metabolite, dentigerumycin E, with antiproliferative and antimetastatic properties (Shin et al., 2018). Thus, *Streptomyces* coculture is a promising culture-based strategy to trigger production of cryptic secondary metabolites.

1.5.1.5. Limitations of culture-dependent approaches

Advancements to culture-dependent approaches in the post-genomic era have improved our ability to find new bioactive molecules from environmental microorganisms. Despite these improvements the majority of microorganisms still remain uncultured (Charlop-Powers et al., 2014). Culture-based strategies favor the recovery of organisms that are best able to thrive under the *in vitro* conditions to which they are exposed (Vartoukian et al., 2010). For this reason, environmental microbiologists estimate that only a small fraction of microbial species can be cultured by these strategies (Bodor et al., 2020; Hofer, 2018).

Replicating all of the necessary conditions *in vitro* may simply be an impossible task for accessing the true diversity of *Streptomyces* species and the wealth of biosynthetic potential they hold. There are endless combinations of growth parameters that might be needed to support isolation of a novel microorganism and some of those same conditions might inhibit other unidentified organisms. For example, isolation strategies to culture novel bacteria from marine environments are still lacking and more effort is needed to develop enrichment strategies for rare bacteria (Subramani & Aalbersberg, 2012). Even with improvements to environmental mimicry in a laboratory setting, the time involved in isolating a pure culture can be extensive. In cases of symbiotic associations, obtaining pure cultures relies on removing the microorganism from its host, which may also remove the environmental cues provided by the host to stimulate natural product production.

Even if a novel microorganism is successfully isolated, its full natural product potential may be missed due to lack of expression or failure to detect BGCs by genome mining tools. The most limiting factor of all is that many of the culture-dependent methods developed in the last decade are still in their infancy and we have yet to fully understand their scope. As such, culture-

independent or “metagenomic” approaches have gained interest as a potential avenue to access the natural product diversity from uncultured microorganisms (M. Katz et al., 2016). Some of the most successful culture-independent approaches to date utilize *Streptomyces* as a host organism for metagenome-derived BGCs.

1.5.2. Culture-independent or “metagenomic” methods

Culture-independent or “metagenomic” methods are an approach to study the biosynthetic potential of uncultured microorganisms by cloning and analyzing microbial DNA extracted directly from the environment (M. Katz et al., 2016). The term “metagenomics” was first used by Handelsman et al. (1998) and was based on the genetic advancements that preceded its development. After Carl Woese first used 16S ribosomal RNA (16S rRNA) as a phylogenetic marker in the 1970s (Woese & Fox, 1977), Pace and colleagues expanded on this theory and proposed to clone DNA containing 16S rRNA genes directly from environmental samples (Pace, 1985). The first application of this cloning was applied by T. M. Schmidt et al. (1991). The next major turning point came from the first construction of a metagenomic library in 1995 (Healy et al., 1995), and ultimately the construction of multiple libraries from bacteria in seawater by Stein et al. (1996). Together these breakthroughs defined the field and demonstrated that DNA could be captured and studied from uncultured microorganisms.

In the modern post-genomic era, shotgun metagenomic sequencing and powerful bioinformatics has made it possible to sequence phylogenetically diverse microbial assemblages that are present in soils and marine ecosystems (Santana-Pereira et al., 2020). Combined with *in silico* tools, DNA from environmental organisms could be sequenced and mined for phylogenetic markers or natural product gene clusters (M. Katz et al., 2016). This meant that natural product discovery was no longer limited by the dependency of culturing the source

organism. Instead, DNA containing secondary metabolite BGCs could be captured, cloned, and heterologously expressed in a culturable host (i.e. *Streptomyces* spp.) to screen for bioactivities of interest and new metabolites (Handelsman, 2004).

Metagenomics-related approaches for natural product discovery can be largely classified into two categories – function-based and sequence-based (M. Katz et al., 2016). Both strategies involve metagenomic library construction by isolating and cloning environmental DNA (eDNA) but differ in the approaches used to identify phenotypes of interest from each metagenomic clone. Function-based metagenomics seeks to identify individual clones for a desired activity through metabolic or bioactivity screens, whereas sequence-based metagenomics aims to first identify clones containing genes of interest such as secondary metabolite BGCs (Alam et al., 2021). Modern metagenomic approaches have combined these two approaches taking a more targeted route by first using sequence-based approaches to identify clones containing specific genes, then targeting those clones for heterologous expression and screening by function-based methods to generate new compounds (Figure 1.3) (Santana-Pereira et al., 2020). The following sections discuss these culture-independent strategies and the application of *Streptomyces* as an ideal heterologous host for metagenome-derived pathways.

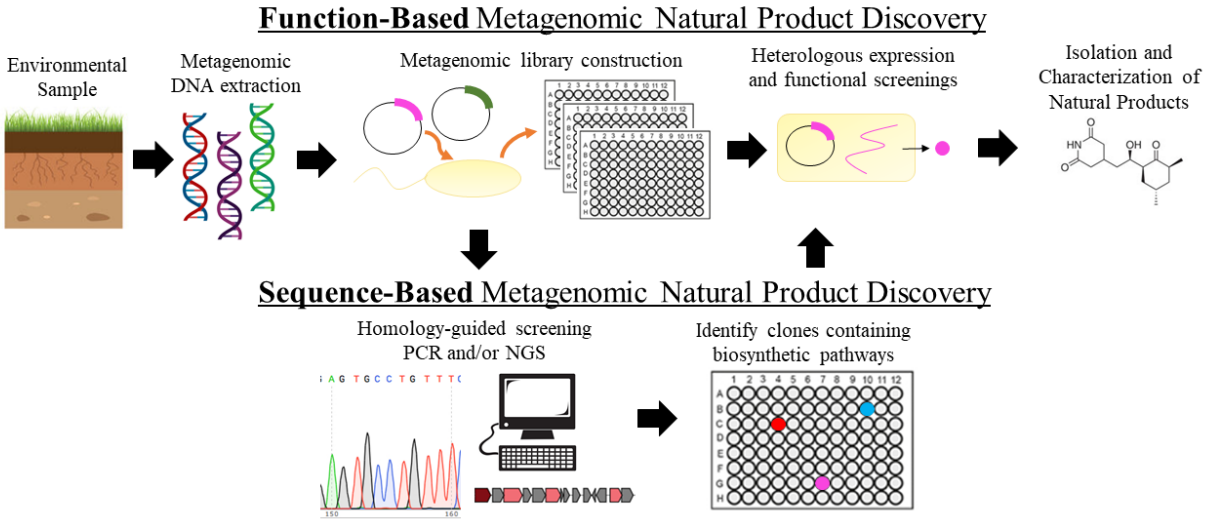

Figure 1.3. Graphical representation of the construction and screening of a metagenomic library for natural products using a function- and sequence-based approach. This figure was adapted from Handelsman (2004).

1.5.2.1. Function-based metagenomics

Classical function-based metagenomics followed a simple and random approach where eDNA was cloned into cosmid vectors in *E. coli* and clones were directly screened for an easily observable phenotype linked to natural product expression (e.g. color, antibiosis, chromatographic pattern) (M. Katz et al., 2016). Several novel bioactive compounds and clones have been discovered by this approach including turbomycin A and B (Gillespie et al., 2002), six antibacterial clones from a cosmid library (Iqbal et al., 2014), two novel lactonases (Schipper et al., 2009), and an uncharacterized antimicrobial protein from a soil library (Biver & Vandenbol, 2013). However, the overall success rate of this classical approach was lower than desired mostly due to limitations in DNA isolation, cloning, and poor expression (Alam et al., 2021; M. Katz et al., 2016). These limitations still exist, but efforts have been made to lessen their detrimental effects.

The ability to find new natural products from culture-independent methods is dependent on the successful capture and cloning of complete gene clusters. Early metagenomics used a direct DNA isolation method where cells within an environmental sample were lysed *in situ* prior to DNA extraction and purification (Alam et al., 2021). This method was quicker but often resulted in smaller DNA fragments that were less likely to contain full-length gene clusters (Kakirde et al., 2010). Indirect DNA isolation methods have since been developed to obtain larger fragments and increase the likelihood of extracting complete BGCs (Wenfang Wang et al., 2021). For example, Liles et al. (2008) used an indirect approach to obtain high molecular weight DNA greater than 1 Mb by separating microbial cells from soil, embedding and lysing the cells in an agarose plug to protect from mechanical stress, and treating with formamide to remove environmental inhibitors.

The next refinement to classical function-based metagenomics came from the development of new cloning systems that could maintain longer DNA inserts. Traditionally, eDNA was cloned into cosmid or fosmid vectors but these were limited to capturing ~40 kb (Kakirde et al., 2010). Many natural product gene clusters can exceed 100 kb which meant that the cosmid/fosmid vectors were less likely to carry complete biosynthetic pathways in a single clone (Trindade et al., 2015). Commercial artificial chromosomes were developed to support cloning of larger eDNA, and some of the most popular vectors include bacterial artificial chromosome (BAC) vectors (up to 300 kb inserts), P1 artificial chromosome (PAC) vectors (up to 100 kb inserts), and yeast artificial chromosome (YAC) vectors (up to 2 Mb inserts) (Wenfang Wang et al., 2021). In particular, BAC vectors are the most frequently used for metagenomic library construction. The earliest BAC vectors, such as pBeloBAC11, were only capable of single-copy replication in *E. coli*. Other BAC vectors are now more widely used for

metagenomic libraries including pSBAC, pSMART-BAC-S, and pStreptoBAC which contain the inducible *oriV* to increase copy number and *oriT* for conjugative transfer into other hosts (Wenfang Wang et al., 2021). These vectors have made it possible to recover larger gene clusters from metagenomes to facilitate functional screenings (Nasrin et al., 2018; Santana-Pereira et al., 2020).

Despite the advancements to DNA extraction and cloning systems, classical function-based metagenomics were still limited by the choice of bacterial hosts for expression. In fact, this step is one of the most significant challenges in metagenomics-based drug discovery since poor expression by a host can preclude compound detection and purification (Chistoserdova, 2009). From a broad viewpoint, problems in heterologous expression are linked to inadequate transcription and translation of DNA and biochemical incompatibilities between a host and a native producer (M. Katz et al., 2016). More specifically, codon bias, lack of substrates, failure to recognize promoters, improper protein folding and secretion, and missing regulatory elements can each limit the probability of expressing metagenome-derived biosynthetic pathways (Craig et al., 2010; Rebets et al., 2017). Classical function-based metagenomics predominantly used *E. coli* as the sole host due to its high cloning efficiency (Kakirde et al., 2010). *E. coli* hosts can be successful for expression of Gram-negative derived enzymes, but were less promising for heterologous expression of BGCs (Alam et al., 2021). It is estimated that *E. coli* can readily express up to 40% of environmental genes, but this value can be as low as 7% for the expression of high-GC% DNA (Craig et al., 2010; Gabor et al., 2004).

Improvements to *E. coli* heterologous hosts were made in an attempt to address these limitations such as modifications to promoters and ribosome binding sites, the introduction of heterologous sigma factors, and chaperone co-expression (Craig et al., 2010; M. Katz et al.,

2016; Rebets et al., 2017). Unfortunately, these modifications were often unsuccessful suggesting that the majority of biosynthetic potential of a metagenomic library may be missed by using *E. coli* as a sole heterologous host (Rebets et al., 2017). In recent years, alternative hosts have been evaluated for their potential as heterologous hosts. The development of transmissible BAC and other vectors has enabled the use of different Gram-positive and Gram-negative hosts including *Pseudomonas putida*, *Agrobacterium tumefaciens*, *Burkholderia graminins*, *Caulobacter vibrioides*, and *Streptomyces* species (Alam et al., 2021; Craig et al., 2010). Among all of the heterologous hosts, engineered strains of *Streptomyces* spp. are increasingly popular in modern functional metagenomics.

1.5.2.1.1. *Streptomyces* as a host for heterologous expression

Several key factors have contributed to the use of *Streptomyces* as a chassis for the production of natural products from metagenomes. *Streptomyces* are known producers of diverse secondary metabolites and capable of making some of the intracellular precursors and cofactors needed to form natural products (Charles et al., 2017). Their high GC codon usage is another rationale for their selection as hosts since this can increase the likelihood of expressing genes from other organisms with high GC% contents, many of which are abundant in soils and marine environments (Charles et al., 2017; Rebets et al., 2017). Many cloning vectors can be readily shuttled from *E. coli* to *Streptomyces* hosts by conjugation and support site-specific integration in the ϕC31 *attB* site of the host that allows for stable expression of exogenous DNA without antibiotic selection (Bierman et al., 1992). An additional strength of Gram-positive hosts, like *Streptomyces*, is their innate, extracellular secretion capacity that can support proper folding and function of secreted proteins (Rebets et al., 2017). Numerous studies have demonstrated that *Streptomyces* can be used to express gene clusters that are difficult to express or silent in other

bacteria which has provided further support for their use as a heterologous host in metagenomic studies (Rebets et al., 2017). Due to these advantages, several *Streptomyces* strains have been engineered for heterologous expression with most belonging to the species *S. coelicolor*, *S. lividans*, and *S. albus*.

S. coelicolor A3(2) is the most extensively studied strain with respect to the biosynthesis and regulation of secondary metabolites (Charles et al., 2017). This wild-type strain has been used to express large gene clusters such as the 106 kb salinomycin gene cluster (Yin et al., 2015). However, *S. coelicolor* A3(2) natively produces several antibiotics including coelimycin (Cpk), undecylprodigiosin (Red), actinorhodin (Act), and the lipopeptide calcium-dependent antibiotic (CDA) (Charles et al., 2017). These endogenous gene clusters can interfere with the expression of exogenous DNA due to competitive sinks in precursor and intermediate availability and can complicate mass spectrometry analysis (Gomez-Escribano & Bibb, 2011).

Derivatives of *S. coelicolor* A3(2) that lack endogenous gene clusters have been developed to mitigate these issues. Deletions of Cpk, Red, Act, and CDA gene clusters from wild-type *S. coelicolor* resulted in a strain called *S. coelicolor* M1146 that produced higher yields of antibiotics from cloned gene clusters and possessed a simpler metabolite profile (Gomez-Escribano & Bibb, 2011). In this study, another derivative was constructed (*S. coelicolor* M1154) that further increased antibiotic production compared to the parent strain through the inclusion of two point mutations in *rpoB* and *rpsL*, encoding the RNA polymerase β-subunit and the ribosomal protein S12, respectively. Both *S. coelicolor* M1154 and M1146 have expressed many different biosynthetic gene clusters and led to the production of bioactive compounds such as vancoresmycin (Kepplinger et al., 2017), cacibiocins A and B (Zettler et al., 2014), kocurin

(Linares-Otoya et al., 2017), anthracimycin (Alt & Wilkinson, 2015), gougerotin (Niu et al., 2013), and desotamides A and B (Song et al., 2014).

Despite the success of *S. coelicolor* hosts, their use has been limited to the expression of clusters from cultured microorganisms and not from metagenome-derived BGCs. Reasons for this are multifactorial but are largely due to the slower growth of *S. coelicolor* compared to other *Streptomyces* hosts and its strict restriction-modification system that can make it less likely to accept foreign DNA (McMahon et al., 2012). There are techniques that can circumvent the methyl-specific restriction of *S. coelicolor* through the use of intergeneric conjugation with a methylation-deficient *E. coli* strain and helper plasmids (A. C. Jones et al., 2013). Even with this approach, conjugation efficiency can be lower using *S. coelicolor* (Ahmed et al., 2020). However, a recent study by Sandoval-Powers et al. (2020; unpublished doctoral dissertation) used *S. coelicolor* to express soil metagenome-derived PKS and NRPS gene clusters and this led to the identification of over 20 bioactive clones. This finding suggests that *S. coelicolor* should not be overlooked as a host for function-based metagenomics.

Beyond *S. coelicolor* hosts, *S. lividans* and *S. albus* are routinely used for the expression of gene clusters from cultured microorganisms and metagenomes. In a similar manner to *S. coelicolor*, genetically engineered strains of *S. lividans* and *S. albus* were developed with deletions of endogenous BGCs providing a "cleaner" host background (Nepal & Wang, 2019). *S. lividans* TK24 is widely used and is capable of accepting methylated DNA, has low protease activity, and possesses a streptomycin-resistant mutation shown to improve production of secondary metabolites (Ahmed et al., 2020). Other hosts have been generated such as *S. lividans* K4-114 and K4-155 with deletion of the Act cluster (Ziermann & Betlach, 1999), the deletion of both Act and Red to generate *S. lividans* ΔactΔred (Martinez et al., 2004), *S. lividans* SBT5 with

integration of global regulatory genes and deletion of negative regulatory genes (Peng et al., 2018), and more recently the engineering of *S. lividans* ΔYA9 by the removal of 11 endogenous gene clusters and the inclusion of additional *attB* sites (Ahmed et al., 2020). Examples of heterologous expression of metagenome-derived BGCs in some of these *S. lividans* strains include the terragines A-E and the siderophoric nocardamine (G.-Y.-S. Wang et al., 2000).

To date, *S. albus* derivatives are the most successful heterologous hosts for functional metagenomics. *S. albus* J1074 has been engineered to lack *Sal*I restriction and modification (Baltz, 2010), and this parent strain has been used to generate other derivatives including *S. albus* Del14 with deletions in 15 endogenous BGCs (Myronovskyi et al., 2018) and *S. albus* S4 with five mutagenized gene clusters (Fazal et al., 2020). The application of these *S. albus* hosts for function-based metagenomics have led to the discovery of a KB-346-5 derivative with antibacterial activity (Feng et al., 2011), the antibiotic landomycin E (Feng et al., 2011), the antibacterial malacidins (Hover et al., 2018), iron-chelating myxochelin A (Bitok et al., 2017), utahmycins A and B with anti-MRSA activity (J. D. Bauer et al., 2010), and metathramycin A with anticancer activity (Stevenson et al., 2021).

The introduction of *Streptomyces* as heterologous hosts for function-based metagenomics has significantly improved our ability to find novel compounds from a wider diversity of environmental microorganisms. Scientists are constantly building on existing *Streptomyces* hosts in order to more efficiently express novel gene clusters from metagenomes and cultured microbes. The field of modern function-based metagenomics is becoming an exciting tool to identify novel natural products and will be accelerated by advancements in sequence-based metagenomics.

1.5.2.2. Sequence-based metagenomics

Since metagenomic libraries can consist of thousands of clones, classical function-based methods were limited by the ability to screen for one type of function at a time and often required high-throughput screening methods (Robinson et al., 2021). Sequence-based metagenomics offers a way to overcome some of these challenges by first identifying clones carrying gene clusters of interest (M. Katz et al., 2016; Trindade et al., 2015). This approach has the advantage of providing more information about each metagenomic clone including the types of BGCs present in a library, their predicted molecular products, and the phylogenetic origin of the eDNA. This information can help target the most promising clones in terms of BGC potential and guide the choice of heterologous host to use in function-based screenings (Trindade et al., 2015).

Initial implementations of sequence-based metagenomics used a PCR approach where degenerate primers were designed targeting a conserved region of a metabolic gene or ribosomal RNA (M. Katz et al., 2016). Specifically for natural product discovery, degenerate primers were designed to target biosynthetic domains of polyketide synthase (PKS) and nonribosomal peptide synthetase (NRPS) clusters (Trindade et al., 2015). Biosynthesis of the molecular products of PKS and NRPS clusters is controlled by multimodular enzymes and each module contains a core set of domains needed for the activation and condensation of an amino or carboxylic subunit (Nikolouli & Mossialos, 2012). The ketosynthase (KS) domain of PKS clusters and the adenylation (AD) domain of NRPSs responsible for these reactions contain highly conserved motifs making them an ideal target for PCR amplification (Cacho et al., 2015; Charlop-Powers et al., 2016). The first study exploring a soil metagenome used this strict PCR approach where primers were designed targeting conserved regions of the KS domain and chain-length-factor genes common to type II PKSs. PCR products containing the type II PKS genes were then

amplified by PCR, cloned, and yielded hybrid type II PKS gene cassettes that could be targeted for expression (Handelsman et al., 1998).

More recent approaches have taken the PCR method one step further. PCR amplicons from clones containing the desired motifs are sequenced resulting in sequencing reads often referred to as natural product sequence tags (NPSTs) (M. Katz et al., 2016). NPSTs can then be analyzed using bioinformatic tools such as eSNaPD (Reddy et al., 2014), NaPDoS (Ziemert et al., 2012), and antiSMASH for BGC predictions. Clones containing NPSTs of interest can be selected from the metagenomic library and used for expression studies. One such example is that of C. Wu et al. (2019) who designed degenerate primers for the AD domain of NRPS clusters, sequenced amplicons to generate AD domain NPSTs, and identified a BGC that could be expressed in *S. albus* to produce the novel antibiotics called cadasides. Another important example is that of the sequence-guided screening of soil metagenomes for BGCs encoding calcium-binding motifs by Hover et al. (2018). In this study, the authors analyzed NPSTs by the eSNaPD tool which led to the discovery of BGCs encoding calcium-dependent antibiotics that could be expressed in *S. albus* to produce the novel malacidins.

These examples have demonstrated the application of combined PCR and bioinformatics methods to discover natural products in metagenomic libraries. However, these approaches are not without limitations. PCR approaches can be biased towards identifying sequences most similar to known pathways and may miss highly divergent sequences capable of encoding novel metabolites (Santana-Pereira et al., 2020). As such, purely *in silico* approaches have been developed to sequence and mine metagenomic libraries without the need for PCR (Trindade et al., 2015). One such approach, shotgun metagenomic sequencing, has made it possible to directly

sequence eDNA but often results in short contigs making it difficult to assemble and predict longer gene clusters (Meleshko et al., 2019).

Alternative approaches have recently been developed to overcome these limitations. For example, Santana-Pereira et al. (2020) used a combination of next-generation sequencing and bioinformatics to screen a metagenomic library for clones predicted to carry a PKS or NRPS cluster. In this study, 151 clones carrying these clusters were identified from the library; notably, this was a far greater rate of identification compared to a PCR approach. Moreover, many of the predicted BGCs appeared to be full-length and unique demonstrating the power of this approach to assess the biosynthetic potential of soil metagenomes.

Regardless of the approach taken, the ultimate goal for natural product strategies from a culture-independent perspective is to identify, express, and purify novel metabolites from uncultured microorganisms. Sequence-based metagenomics has already demonstrated its value in this goal by enabling faster dereplication of gene clusters in a metagenomic library (Robinson et al., 2021; Trindade et al., 2015).

1.5.2.3. Limitations of culture-independent approaches

Despite the advantages of culture-independent methods, there has yet to be a commercial success from these approaches (Baltz, 2019). Several limitations have impeded the success of metagenomics in drug discovery. First, even with the advancements made in engineering "optimized" heterologous hosts, poor expression of BGCs is still a major bottleneck in metagenomics-based drug discovery (Pereira, 2019). Modern techniques to enhance expression of "model BGCs" with known architecture often manipulate the pathway of interest by overexpressing or disrupting regulators, modifying transcription elements, or manipulating the host (Huo et al., 2019; Pereira, 2019). However, these techniques are time-consuming and can be

impossible to apply to function-based metagenomic screenings where the diversity of gene clusters are vast, their pathway organization is often unknown prior to expression, and their metabolic products may not be easily detected or quantified. Thus, better understanding ways to manipulate and ideally enhance heterologous expression of metagenomic BGCs in a high-throughput manner are needed to accelerate metagenomics-based drug discovery. There are now well-described inducible expression systems in streptomycetes, such as the P_{nitA}-NitR system (Herai et al., 2004) and the P_{otr} system (Weishan Wang, Yang, et al., 2016), which have been shown to increase heterologous production by *Streptomyces* hosts greater than 10-fold (Herai et al., 2004; Matsumoto et al., 2016; Weishan Wang, Yang, et al., 2016). These techniques have only been applied to expressing gene clusters from cultured bacteria, but may be critical to improve the expression of novel metabolites from metagenomes.

Another limitation of these approaches is the selection of a heterologous host since one host may not possess all the genetic and biochemical elements needed to express all BGCs in a metagenome (Craig et al., 2010). For this reason, it is widely assumed that the phylogenetic distance between the origin of the gene cluster and the host organism plays a significant role for effective heterologous expression (Gadd & Sariaslani, 2016; J. J. Zhang et al., 2017). Despite the theoretical appeal, this assumption has yet to be thoroughly tested in metagenomic studies (J. J. Zhang et al., 2017). Since proper heterologous expression of metagenomic DNA is crucial to identify active clones (J. J. Zhang et al., 2017), employing multiple heterologous hosts in metagenomic library screenings may improve the detection frequency (Craig et al., 2010). Optimized *Streptomyces* heterologous hosts are a great starting point, but hosts from other bacteria within Actinobacteria or other phyla may need to be developed beyond the established *E. coli* strains.

Limitations in the detection capabilities of sequence-based screenings like PCR and BGC-prediction tools are another major challenge in a metagenomic approach. Sequence-based strategies still heavily rely on homology screenings to known gene clusters from a few classes (Santana-Pereira et al., 2020; Zerikly & Challis, 2009). Over time, this inherent bias may lead back to the problem of compound rediscovery since more divergent gene clusters may not be predicted with the current tools. Failure to detect these unique gene clusters may cause us to miss novel bioactive compounds with modes of action that are not yet impacted by drug-resistance mechanisms of pathogenic microorganisms. Continued developments in bioinformatics and other technologies are needed to predict a larger diversity of gene clusters. These advancements can be complemented by the deposition of more metagenomic sequence data in publicly available databases (Trindade et al., 2015).

Due to these limitations, the most efficient pipeline for culture-independent drug discovery remains unclear at this time. These bottlenecks must be addressed in order for metagenomics to play a larger role in antibiotic discovery. Advancements in molecular biology, bioinformatics, and sequencing technologies will help tackle these limitations and improve our ability to explore the chemical diversity of uncultured microorganisms.

1.6. Concluding Remarks

The application of the aforementioned approaches has demonstrated that *Streptomyces* continues to be an important source of novel natural products with diverse bioactivities. Modern culture-dependent techniques have uncovered novel *Streptomyces* spp. from unique habitats and symbioses and led to the discovery of many new bioactive compounds. Moreover, culture-independent or “metagenomic” approaches have opened the door to exploring the chemical

diversity of uncultured microorganisms and it is clear that *Streptomyces* has an important role in these approaches as heterologous hosts.

Despite their success, modern methods for natural product discovery are still hindered by the inability to detect highly novel gene clusters by bioinformatic tools, inadequate gene expression by native and heterologous hosts, and incomplete understanding of natural product biosynthesis. Advancements in molecular biology, metabolite detection, and bioinformatics are constantly expanding and will undoubtedly help address these limitations. Furthermore, integrative approaches that combine culture-independent approaches to access novel gene clusters with the culture-dependent tools to express the cultures in the laboratory will provide an opportunity to find new compounds from multiple avenues. It is through these innovations that we will continue to unlock the potential of nature's pharmacists and tackle the antimicrobial resistance crisis.

Chapter 2

Discovery and characterization of a novel marine *Streptomyces* species expressing anti-MRSA activity

Manuscript published in *Systematic and Applied Microbiology* (Sandoval-Powers et al., 2021)

2.1. Abstract

Marine sponges represent a rich source of uncharacterized microbial diversity, and many are host to microorganisms that produce biologically active specialized metabolites. Here, a polyphasic approach was used to characterize two Actinobacteria strains, P01-B04^T and P01-F02, that were isolated from the marine sponges *Geodia barretti* (Bowerbank, 1858) and *Antho dichotoma* (Esper, 1794), respectively. Phylogenetic analysis based on 16S rRNA gene sequences indicated that strains P01-B04^T and P01-F02 are closely related to *Streptomyces beijiangensis* DSM 41794^T, *Streptomyces laculatispora* NRRL B-24909^T, and *Streptomyces brevispora* NRRL B-24910^T. The two strains showed nearly identical 16S rRNA gene sequences (99.93%), and the average nucleotide identity (ANI) and digital DNA-DNA hybridization (dDDH) relatedness values were 99.96% and 99.6%, respectively, suggesting that these strains are affiliated with the same species. Chemotaxonomic and culture characteristics of both strains were consistent with the genus *Streptomyces*, while phenotypic properties, genome-based comparisons, and phylogenomic analyses distinguished strains P01-B04^T and P01-F02 from their closest phylogenetic relatives. *In silico* analysis predicted that the 8.9 Mb genome of P01-B04^T contains at least 41 biosynthetic gene clusters (BGCs) encoding secondary metabolites, indicating that this strain could express diverse bioactive metabolites; in support of this prediction, this strain expressed antibacterial activity against Gram-positive bacteria including a clinical isolate of methicillin-resistant *Staphylococcus aureus* (MRSA) EAMC30. Based on these

results, the marine sponge-associated isolates represent a novel species of the genus *Streptomyces*, for which the name *Streptomyces poriferorum* sp. nov. is proposed, with P01-B04^T (= DSM 111306^T = CCM 9048^T) as the type strain.

2.2. Introduction

Members of the genus *Streptomyces*, first proposed by Waksman and Henrici in 1943 (Waksman & Henrici, 1943), are well-known producers of numerous bioactive metabolites including antibacterial (de Lima Procópio et al., 2012), antifungal (Barka et al., 2016), and anticancer compounds (Olano et al., 2009). These Gram-positive filamentous bacteria are affiliated with the phylum Actinobacteria, one of the most diverse phyla within the domain *Bacteria* (Arocha-Garza et al., 2017; Goodfellow, 2012). *Streptomyces* spp. possess a complex life cycle wherein vegetative mycelia differentiate into reproductive aerial mycelia prior to spore formation (Yagüe et al., 2013). *Streptomyces* spp. can be further distinguished by physiological and genetic characteristics including high DNA %G+C content (69-78%) (M. H. Farris et al., 2011), LL-diaminopimelic acid (L-DAP) in their peptidoglycan (Anderson & Wellington, 2001; J. Y. Lee, Lee, et al., 2005), and the amounts of saturated iso- and anteiso-branched fatty acids (Sujarit et al., 2016). As such, polyphasic taxonomic studies have been critical for the classification of new *Streptomyces* species by comparing the phenotypic and genotypic characteristics of a presumably novel isolate to closely related type strains of previously described species (Anderson & Wellington, 2001; Vandamme et al., 1996). These studies have led to the isolation of more than 900 validly described species within the genus *Streptomyces* (http://www.bacterio.net, 25 June 2021, date last accessed) (Parte, 2014), making it the most well-described genus of the phylum Actinobacteria.

Many *Streptomyces* spp. and other members of the phylum Actinobacteria have been isolated from terrestrial soils (Seipke et al., 2012); however, research within the last decade has shown that these microorganisms also are widely distributed in marine environments (Abdelmohsen et al., 2014; Bredholt et al., 2008; Dharmaraj, 2010). In particular, marine sponges have gained attention as a hotspot for the discovery of novel *Streptomyces* spp. with uncharacterized biosynthetic potential (Kennedy et al., 2009; Nair et al., 2011). These marine invertebrates in the phylum *Porifera* typically harbor their microbial symbionts in their mesohyl matrix (Abdelmohsen et al., 2014; Nair et al., 2011), where they provide nutrition to the host (Abdelmohsen et al., 2014; Nair et al., 2011; Tianero et al., 2019) and produce potent secondary metabolites that may protect the host from infection and predation (Abdelmohsen et al., 2014; Nair et al., 2011). The discovery of novel sponge-associated *Streptomyces* taxa has led to the characterization of new natural products with diverse bioactivities, such as the tetromycins 1-4 isolated from *Streptomyces axinellae* PoI001^T (Abdelmohsen et al., 2014; Pimentel-Elardo et al., 2011), urauchimycin A and B from *Streptomyces* sp. Ni-80 (Abdelmohsen et al., 2014; Imamura et al., 1993), and lobophorin C and D from *Streptomyces carnosus* strain AZS17 (Abdelmohsen et al., 2014; R.-B. Wei et al., 2011). Thus, the isolation and characterization of novel *Streptomycetes* from marine sponges is important for identifying potentially novel therapeutic compounds.

In the present study, a polyphasic taxonomic approach was conducted to characterize two novel Actinobacteria strains, designated P01-B04^T and P01-F02, isolated from different marine sponges collected in the Trondheim fjord of Norway. We also investigated the antimicrobial activity and biosynthetic potential of these isolates and conclude that these strains could be an important source of novel antimicrobial metabolites. We report that the isolates P01-B04^T and

P01-F02 represent a novel species of the genus *Streptomyces*, for which the name *Streptomyces poriferorum* sp. nov. is proposed.

2.3. Materials and Methods

2.3.1. Isolation and culture conditions.

Strains P01-B04^T and P01-F02 were isolated from the marine sponges *Geodia barretti* (Bowerbank, 1858) and *Antho dichotoma* (Esper, 1794), respectively, which were collected in September 2005 at an underwater wall close to Tautra island located in the Trondheim fjord of Norway (GPS: 63° 36' 53 N 10° 31' 22 E). The sampling of, strain isolation, and whole-genome sequencing of the isolates from these two sponge species was part of an extensive campaign by SINTEF and the Norwegian University of Science and Technology (NTNU) to explore the biosynthetic potential and phylogenetic diversity of marine Actinobacteria and has been described earlier (Guerrero-Garzón et al., 2020; Ian et al., 2014). In brief, sponges were transferred to the laboratory in containers with sterile seawater. Sponge samples were ground and plated on different selective seawater containing solid Actinobacteria isolation Media (IM) containing different selective compounds. P01-B04^T was isolated from IM4 agar with cycloheximide (50 µg/ml), nystatin (75 µg/ml), and nalidixic acid (30 µg/ml), while P01-F02 was isolated from IM18 agar with cycloheximide (50 µg/ml), nystatin (75 µg/ml), nalidixic acid (30 µg/ml), and trimethoprim (20 µg/ml). Pure cultures of the isolates were maintained on International *Streptomyces* Project (ISP) 2 agar (Shirling & Gottlieb, 1966) at room temperature and preserved as a suspension of mycelia in glycerol (20% v/v) at -80 °C.

Three of the most closely related phylogenetic relatives to strains P01-B04^T and P01-F02 based on 16S rRNA gene sequence comparative analysis were selected and maintained under the same conditions with the isolates of interest. These reference strains included *Streptomyces*

beijiangensis DSM 41794^T obtained from DSMZ (German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany), and *Streptomyces laculatispora* NRRL B-24909^T and *Streptomyces brevispora* NRRL B-24910^T obtained from the USDA Agricultural Research Service Culture Collection (Northern Regional Research Laboratory, Peoria, IL). Reference strains were tested in parallel with strains P01-B04^T and P01-F02 for comparison purposes.

2.3.2. Morphology.

Morphological characterization of strains P01-B04^T, P01-F02, and reference strains were determined following growth on various media, including ISP media (Shirling & Gottlieb, 1966), nutrient agar, modified Bennett's agar (K. L. Jones, 1949), and Czapek's Dox agar (Atlas, 2010) after incubation at 30 °C for 1 – 3 weeks. The color of aerial and substrate mycelium and soluble pigments were determined using the ISCC-NBS color charts (Kelly & Judd, 1965). Production of melanin pigments was observed on ISP 6 medium. Morphology of mycelia and spores was observed by scanning electron microscopy (SEM) by first growing the strain on ISP 2 for 7 days at 30 °C. A small square was cut from the agar and dried on an aluminum stub. The sample dried on the stub was sputter coated with a thin layer of gold within a vacuum and a Zeiss (Germany) digital SEM was used to image the samples at 72X, 500X, and 1000X.

2.3.3. Physiology and biochemical properties.

The effects of different temperatures (3 – 35 °C, at intervals of 5 °C) and pH range (3 – 14, at intervals of 0.5 pH unit) on growth of strains cultured on ISP 2 were determined according to Shirling and Gottlieb (1966). NaCl tolerance was examined on ISP 2 agar supplemented with 0 – 10% (w/v) NaCl (at intervals of 0.5% w/v) at 30 °C for 2 – 4 weeks. Catalase and oxidase activity of the novel isolates and reference strains were performed following the methods by Smibert et al. (1994). Urease activity was determined on Christensen's medium (Christensen,

1946). Nitrate reduction was evaluated following the method described by Lanyi (1988) with growth of the strains in ISP 2 media. Hydrolysis of xanthin, hypoxanthin, casein, and tyrosine were determined using the *Nocardia* hydrolysis kit (Hardy Diagnostics) following the manufacturer's protocol. Hydrolysis of DNA was determined on DNase agar (BD Biosciences). Hydrolysis of Tween 80, starch, gelatin and esculin was tested as described by Collins (1967). The utilization of cellulose was determined with a strip of Whatman paper No.1 added to each culture in ISP 2 broth and subsequently observed for four weeks. Enzyme profiles were determined using API ZYM kits (bioMérieux, France) following the manufacturer's instructions, and the utilization of various carbon sources was determined according to the methods described by Shirling and Gottlieb (1966) using ISP 9 medium containing 1% (w/v) of each carbon source. Antibiotic resistance profiles were determined by the Kirby-Bauer method (A. Bauer, 1966) using antibiotic discs (Oxoid) including ampicillin (10 µg), chloramphenicol (30 µg), gentamicin (10 µg), kanamycin (30 µg), neomycin (5 µg), streptomycin (10 µg), tetracycline (30 µg), penicillin G (10 µg), and clindamycin (2 µg). Growth inhibition zones surrounding the antibiotic discs were assessed periodically for up to 3 weeks of incubation at 30 °C, and susceptibility was determined following the Clinical and Laboratory Standards Institute (CLSI) breakpoints (CLSI, 2015).

2.3.4. Chemotaxonomic analyses.

Fatty acids were analyzed after cultivating each isolate in TSB prepared by dispensing 20 ml of media to 150 ml Erlenmeyer flasks and sterilized by autoclaving (30 min/ 121 °C). All strains were cultivated in parallel for 10 days at 28 °C on a platform shaker at 150 rpm. Extraction of fatty acid methyl esters was performed following the standard protocol by Sasser (1990) and analyzed using an Agilent 7890B gas chromatograph according to the standard

protocol of the Sherlock MIDI Identification System (MIDI Sherlock version 6.2, MIDI database RTSBA 6.21). Analysis of respiratory quinones, whole-cell sugars, and polar lipids were carried out by the Identification Service of DSMZ (Braunschweig, Germany). Menaquinones were extracted according to Tindall (1990a); (Tindall, 1990b) and analyzed by HPLC-DAD-MS/MS. Hydrolysates of the whole-cell sugars and diaminopimelic acid were analyzed according to the procedures of Schumann (2011). The analysis of polar lipids was performed following the methods of Minnikin et al. (1984).

2.3.5. Genome sequencing, assembly, and annotation.

Genomic sequences of the isolates were determined in three separate Illumina sequencing runs and used to generate consensus sequences determined as follows: Biomass for genome sequencing was produced in TSB medium supplemented with 50% artificial sea water (1X sea water: 0.425 M NaCl, 0.009 M KCl, 0.0093 M $CaCl_2$, 0.0255 M $MgSO_4$, 0.023 M $MgCl_2$, 0.002 M $NaHCO_3$, pH adjusted to 7.8) at 30 °C. The biomass was collected by centrifugation and sent for sequencing to BaseClear BV (Leiden, The Netherlands), where the extraction of DNA, sequencing, and post-sequencing data processing of P01-B04^T and P01-F02 were carried out. Paired-end sequence reads were generated using the Illumina HiSeq2500 system. FASTQ sequence files were generated using the Illumina Casava pipeline version 1.8.3. Initial quality assessment was based on data passing the Illumina Chastity filtering. Subsequently, reads containing adapters and/or PhiX control signal were removed using BaseClear's in-house filtering protocol. The second quality assessment was based on the remaining reads using the FASTQC quality control tool version 0.10.0. The quality of the FASTQ sequences was enhanced by trimming off low-quality bases using the “Trim sequences” option of the CLC Genomics Workbench version 8.0 (Qiagen, Aarhus, Denmark). This resulted in a total of 2,658,557 and

2,488,558 reads for strain P01-F02 and strain P01-B04^T, yielding 630 Mb and 595 Mb sequencing data with the average quality score (Phred) obtained of 37.03 and 36.97, respectively. For genome assembly and scaffolding, the quality-filtered sequence reads were assembled *de novo* into contig sequences using CLC Genomics Workbench version 8.0. Mis-assemblies and nucleotide disagreement between the Illumina data and the contig sequences were corrected with Pilon (Walker et al., 2014) version 1.11. The contigs were linked and placed into scaffolds or super-contigs. The orientation, order, and distance between the contigs was estimated using the insert size between the paired-end and/or mate-pair reads. The analysis was performed using the SSPACE Premium scaffolder version 2.3 (Boetzer et al., 2011). The gapped regions within the scaffolds were (partially) closed in an automated manner using GapFiller version 1.10 (Boetzer & Pirovano, 2012), taking advantage of the insert size between the paired-end and/or mate-pair reads. The obtained draft genomes were subsequently used for phylogenetic analyses and genome annotations. *De novo* assemblies of P01-B04^T and P01-F02 contain 695 and 653 scaffolds, respectively, and have been deposited at NCBI Genbank under accession numbers JAELVH000000000 and JAELVG000000000, respectively.

Genomes from the type strains *S. laculatispora* NRRL B-24909^T and *S. beijiangensis* DSM 41794^T were sequenced in the same manner described above as these were not publicly available. Raw sequencing reads of the type strains were trimmed and assembled *de novo* using CLC Genomics Workbench version 9.0.1 (Qiagen, Aarhus, Denmark). Assemblies were uploaded to the KBase web server (Arkin et al., 2018), and CheckM version 1.0.18 (Parks et al., 2015) was used to assess genome assemblies for contamination and completeness. The genome of the reference strain *S. brevispora* NRRL B-24910^T was not available in public databases; therefore, the genome sequence from the equivalent DSMZ type strain, *S. brevispora* DSM

42059^T, was used. The draft genome assembly of *S. beijiangensis* DSM 41794^T is available at GenBank/ENA/DDBJ under the accession number JAFLRJ000000000 and the version described in this paper is version JAFLRJ010000000. The draft assembly of *S. laculatispora* NRRL B-24909^T was also deposited at GenBank/ENA/DDBJ under the accession number JAFMPZ000000000. The version described in this paper is version JAFMPZ010000000.

A prediction of BGCs in each genome was performed using antiSMASH version 5.0 (Blin et al., 2019) with antiSMASH beta version 6 enabled and default parameters. A heatmap depicting the number of each BGC type across the query genomes was generated using the RStudio packages "ggplot2" (Wickham, 2016), "reshape2" (Wickham, 2007a), and "tidyverse" (Wickham et al., 2019). Annotations of ORFS and domains from an individual BGC, Cluster 1, as predicted by antiSMASH v5.0 were visualized using the R package "gggenes" (Wilkins & Kurtz, 2019) and "ggplot2". Assembled genome sequences from the novel strains, P01-B04^T and P01-F02, were further analyzed by the Rapid Annotation using Subsystem Technology (RAST) server version 2.0 (Aziz et al., 2008). Functional annotations of the genomes were performed using RAST's KEGG Automatic Annotation Server (KAAS) (Moriya et al., 2007) and viewed using the KEGG Mapper. The potential distribution of *S. poriferorum* across other sponge-associated metagenomes was determined by querying all available sequences from phylum *Porifera* sponges in the JGI IMG/M database (I.-M. A. Chen et al., 2021). Using the BLAST tool, the 16S rRNA gene sequences of each *S. poriferorum* isolate were queried against the sponge metagenome sequences to determine percent identities.

2.3.6. 16S rRNA phylogenetic analysis.

The 16S rRNA gene sequence of strain P01-B04^T (1,528 bp) and P01-F02 (1,528 bp) were extracted from the draft genome sequences and submitted to the EzBioCloud server (Yoon

et al., 2017) for initial identification. The 16S rRNA gene sequences of the novel isolates and the related type strains were aligned using ClustalW in MEGA version 10.0.5 software (Kumar et al., 2018). Phylogenetic trees were reconstructed using the neighbor-joining (NJ) (Saitou & Nei, 1987), maximum-likelihood (ML) (Felsenstein, 1981), and maximum-parsimony (MP) (Kluge & Farris, 1969) methods. The topologies of the resulting trees were evaluated with 1000 iterations for bootstrap support (Felsenstein, 1985). Root positions of the phylogenetic trees were estimated using the 16S rRNA gene sequence of *Kitasatospora setae* KM-6054^T (GenBank accession number NR_112082.2). The 16S rRNA gene sequences of the novel isolates were deposited in GenBank under the accession numbers MW583039 for strain P01-B04^T and MW583038 for strain P01-F02.

2.3.7. Whole genome comparisons and phylogenomic analysis.

Calculations of the average nucleotide identity (ANI) values were determined using the OrthoANI software package (I. Lee et al., 2016). DNA-DNA hybridization (DDH) of strains P01-B04^T, P01-F02, and the reference strains was estimated using the Genome to Genome Distance Calculator (GGDC) version 2.1 (Meier-Kolthoff et al., 2014) (*in silico* DDH). Phylogenomic analysis was performed by uploading the genome sequences of isolates P01-B04^T, P01-F02, and the reference strains to the type strain genome server (Meier-Kolthoff & Göker, 2019). These genome sequences were compared against all of the type strain genomes available in the TYGS database via the MASH algorithm (Ondov et al., 2016) and the closest related type strains were selected for pairwise comparisons using the Genome BLAST Distance Phylogeny (GBDP) method (Meier-Kolthoff et al., 2013). Intergenomic distances were then used to construct a balanced minimum evolution tree with branch support via FASTME 2.1.6.1 (Lefort

et al., 2015). Branch support was inferred from 100 pseudo-bootstrap replications and the tree was rooted at the midpoint (J. S. Farris, 1972).

2.3.8. Antimicrobial activity assays.

Target organisms for antimicrobial activity assays included *Micrococcus luteus* ATCC 10240, methicillin-sensitive *Staphylococcus aureus* (MSSA) Xen29 (Park et al., 2012), a clinical MRSA isolate from the East Alabama Medical Center (EAMC) designated as *S. aureus* EAMC30, *Escherichia coli* BL21 (Liria & Kilikian, 1997), and *Candida albicans* ATCC 90028. Additionally, both isolates were screened for the ability to inhibit the plant pathogens *Clavibacter michiganensis* subsp. *michiganensis* 89C-4 and *Curtobacterium flaccumfaciens* subsp. *flaccumfaciens* CV3. The antibacterial and antifungal activity of P01-B04^T and P01-F02 was evaluated using a spot-on-lawn assay with cell-free supernatants. The isolates were cultured in 250 mL shake flasks with a stainless steel coiled spring containing 100 mL of various media, which included: yeast extract-malt extract (YEME) (Shepherd et al., 2010), marine fermentation (MF) broth (Costa et al., 2017), glucose yeast extract (GYE) broth (Koo et al., 1999), ISP 2 broth, and TSB. Cultures were incubated for 35 days at 25 °C with 200 rpm shaking. One milliliter was separated from the cultures every seven days, and cells were pelleted by centrifugation at 4,000 x *g*. Supernatants were filtered through a 0.2 µm microporous membrane, and 10 µl of cell-free supernatant was spotted on the surface of agar plates containing the target organisms. A 10 µl spot of each media type was used as a control. After spots were dried in a clean bench, plates were incubated at the appropriate temperature for the target organism for 24 to 48 hr. The diameters of the zones of inhibition were subsequently measured.

2.4. Results and Discussion

2.4.1. 16S rRNA phylogenetic analysis.

The complete 16S rRNA gene sequences of strain P01-B04^T (1,528 bp; GenBank accession number MW583039) and strain P01-F02 (1,528 bp; GenBank accession number MW583038) were extracted from their genome sequences and compared to sequences of type strains using the EzBioCloud server. The phylogenetic analysis based on 16S rRNA gene sequences confirmed that strains P01-B04^T and P01-F02 are most closely related to members of the genus *Streptomyces*. Strain P01-B04^T showed the highest 16S rRNA gene sequence similarity with soil isolates *S. beijiangensis* DSM 41794^T (99.03%), *S. brevispora* NRRL B-24910^T (99.02%), and *S. laculatispora* NRRL B-24909^T (98.95%). Similarly, the 16S rRNA sequence of strain P01-F02 was most similar to the same type strains with minor differences in the percent similarity: 99.09% similar to *S. brevispora* NRRL B-24910^T, 99.02% similar to *S. laculatispora* NRRL B-24909^T, and 98.96% similar to *S. beijiangensis* DSM 41794^T. Thus, these three *Streptomyces* type strains were selected as the closest phylogenetic relatives for comparative analyses.

A maximum likelihood phylogenetic analysis of the 16S rRNA gene sequences demonstrated that strains P01-B04^T and P01-F02 formed a monophyletic clade with *S. beijiangensis* DSM 41794^T (= NBRC 100044^T), *S. rectiviolaceus* NRRL B-16374^T, and *S. tauricus* JCM 4837^T (Fig. 2.1).

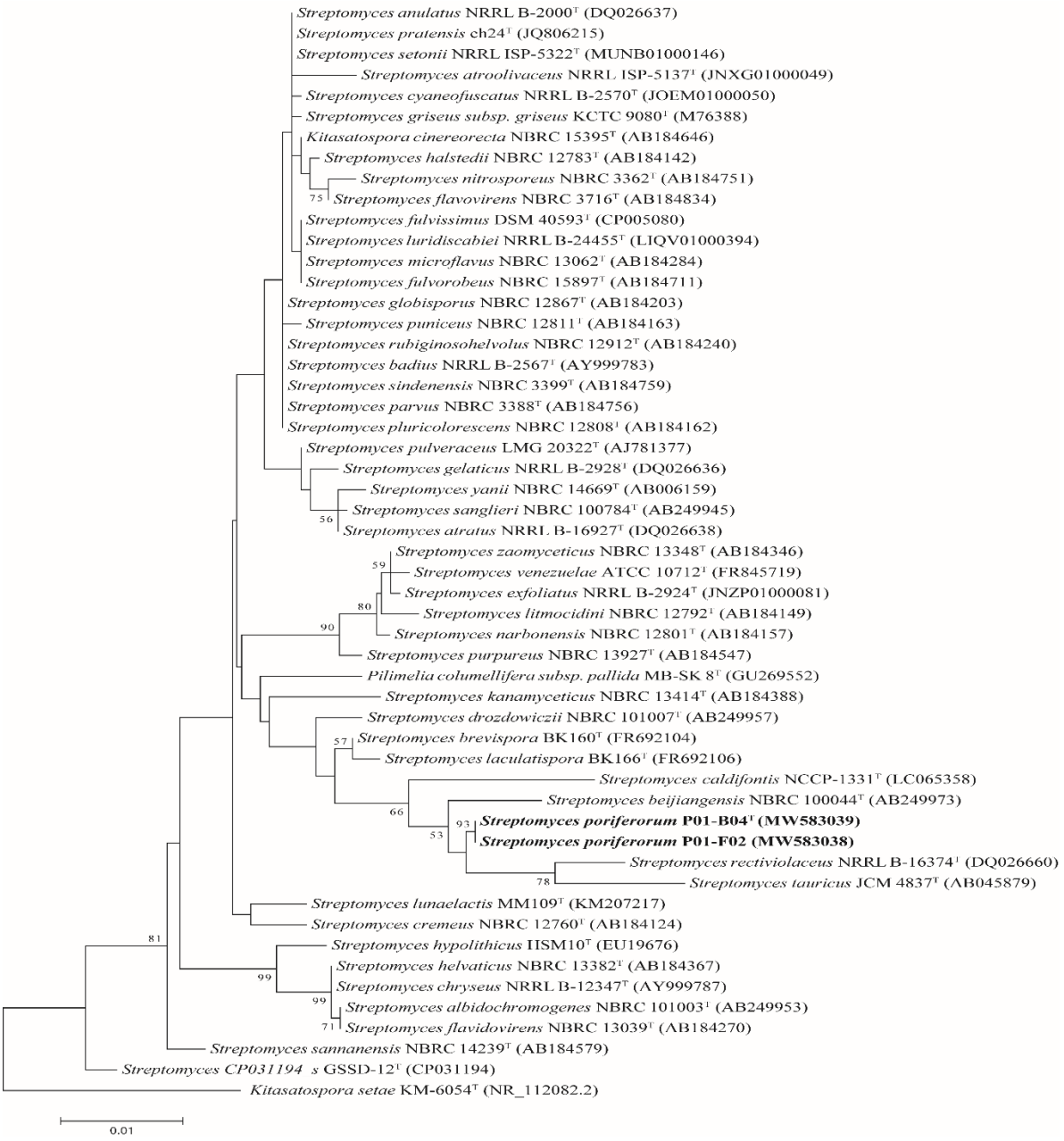

Figure 2.1. Maximum likelihood (ML) 16S phylogenetic tree based on 16S rRNA gene sequences depicting the relationship between strains P01-B04^T, P01-F02, and closely related taxa of the genus *Streptomyces*. *Kitasatospora setae* strain KM 6054^T (NR_112082.2) was used as an outgroup. Numbers at branch nodes indicates bootstrap values based on 1000 replications and only values >50% are shown. Bar, 0.01 substitutions per nucleotide position.

This clustering was also observed from neighbor joining (Supplementary Figure S2.1) and maximum parsimony (Supplementary Figure S2.2) trees and both strains subclustered with

S. brevispora NRRL B-24910^T (= BK160^T) and *S. laculatispora* NRRL B-24909^T (= BK166^T) in each of the three phylogenetic analyses. Additionally, a comparison of 16S rRNA identity between the two isolates was performed using a pairwise BLAST analysis, revealing that P01-B04^T and P01-F02 had nearly identical 16S rRNA gene sequences (99.93%). This finding suggests that the two strains belong to the same novel species, and further genome-based comparisons described below supported this conclusion.

2.4.2. Genome features and genotypic analysis.

Sequencing of the strain P01-B04^T genome resulted in a draft genome size of approximately 8.9 Mb (695 scaffolds) with a DNA G + C content of 70.7 mol%, which is consistent with *Streptomyces* taxa (Wright & Bibb, 1992). The genome sequence of strain P01-B04^T was deposited at GenBank under the accession number JAELVH000000000. The N50 contig length was 26,939, and a total of 8,386 protein-coding genes (CDSs) and 63 RNAs were predicted by RAST annotation. In comparison, the genome size of strain P01-F02 was slightly larger (9.0 Mbp) with 653 scaffolds and a DNA G + C content of 70.8% mol. The N50 contig length was 29,528 and 8,500 CDSs and 67 RNAs were annotated by RAST. The genome sequence of strain P01-F02 was deposited at GenBank under the accession number JAELVG000000000. The general genome characteristics of both isolates and their closest related type strains are listed in Table 2.1.

Table 2.1. Genomic features of *S. poriferorum* sp. nov. strains P01-B04^T and P01-F02 and the closest related *Streptomyces* type strains.

Feature	***Streptomyces poriferorum* sp. nov.**		***S. laculatispora* NRRL B-24909^T**	***S. beijiangensis* DSM 41794^T**	***S. brevispora* DSM 42059^T**
	P01-B04^T	**P01-F02**			
NCBI accession NO.	JAELVH000000000	JAELVG000000000	JAFMPZ000000000	JAFLRJ000000000	VIWW00000000
Genome size (bp)	8,913,815	9,034,630	7,316,623	8,403,055	7,540,224
Mean genome coverage (x)	59	64	120	76	464
Scaffolds	695	654	663	1,769	3
Contigs N50 (bp)	26,939	29,528	14,724	21,718	6,430,118
Contigs L50 (bp)	99	91	158	113	1
DNA G + C content (mol %)	70.7	70.8	70.6	69.6	70.3
Protein coding genes (CDS)	8,444	8,500	6,802	8,634	6752
RNAs	63	67	61	73	98

Taxa from the genus *Streptomyces* are known to possess high sequence similarity in the 16S rRNA gene and often found to be much higher than the <97% cutoff (Stackebrandt & Goebel, 1994) used to delimit species (Labeda, 2011). Therefore, more robust analyses such as whole genome comparisons and phylogenomic analysis are recommended to provide greater taxonomic resolution when comparing potentially novel *Streptomyces* species to other closely related species (Labeda et al., 2014). Here, a phylogenomic tree was generated which revealed that strains P01-B04^T and P01-F02 form a distinct clade with high bootstrap support (100%) (Fig. 2.2). Moreover, the tree further supported that the novel isolates were most closely related to the type strains *S. laculatispora* NRRL B-24909^T and *S. brevispora* DSM 42059^T as these strains formed a monophyletic group with P01-B04^T and P01-F02. The remaining *Streptomyces* type strains evaluated in the analysis, including *S. beijiangensis* DSM 41794^T, were more distantly related to strains P01-B04^T and P01-F02.

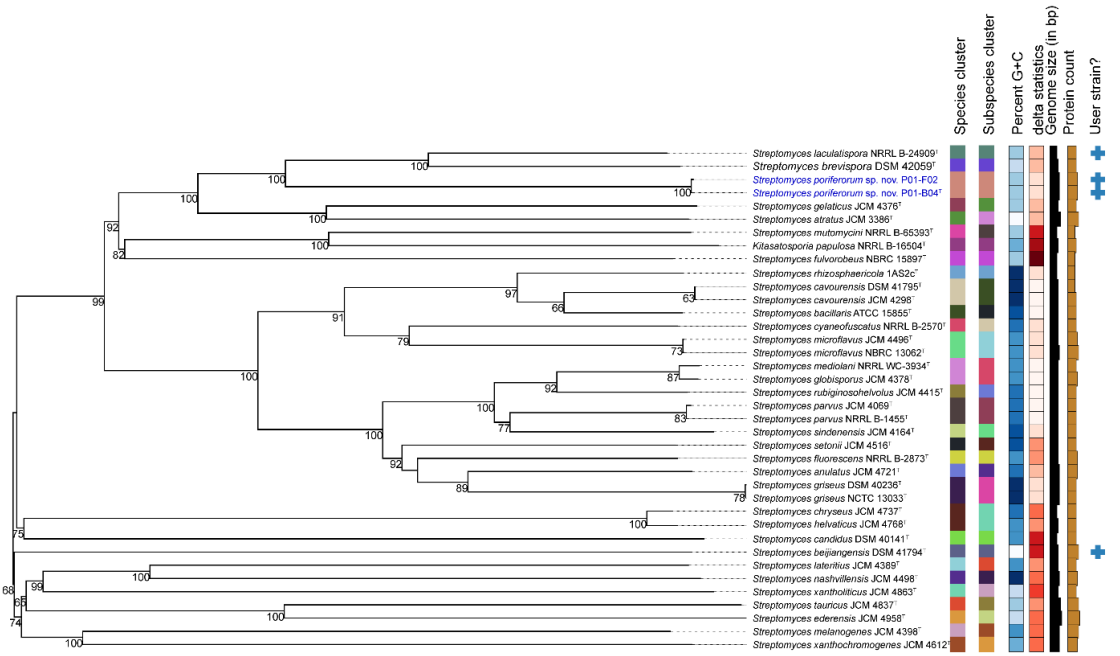

Figure 2.2. Phylogenomic tree based on genome sequences from strains P01-B04^T, P01-F02 and the closest related type strains in the TYGS database (Meier-Kolthoff & Göker, 2019). The tree was inferred with FastME 2.1.6.1 (Lefort et al., 2015) from GBDP distances calculated from genome sequences. The branch lengths are scaled in terms of GBDP distance formula d_5 (Meier-Kolthoff et al., 2013). The numbers above branches are GBDP pseudo-bootstrap support values >60% from 100 replications, with an average branch support of 82.6%. The tree was rooted at the midpoint (J. S. Farris, 1972). The position of the strains of interest are shown in blue.

ANI values determined by the orthoANI software supported that strains P01-B04^T and P01-F02 represent a novel species. ANI values between strain P01-B04^T and the reference strains *S. beijiangensis* DSM 41794^T, *S. laculatispora* NRRL B-24909^T, and *S. brevispora* DSM 42059^T were 79.79%, 88.28%, and 88.32%, respectively (Supplementary Figure S2.3). Comparison of strain P01-F02 to *S. beijiangensis* DSM 41794^T, *S. laculatispora* NRRL B-24909^T, and *S. brevispora* DSM 42059^T revealed ANI similarities of 79.81%, 88.29%, and 88.30%, respectively. This finding aligned with 16S rRNA phylogenetic analyses, which indicated that strains P01-B04^T and P01-F02 are most closely related to *S. beijiangensis* DSM 41794^T. Overall, ANI values between the novel isolates and the reference strains were far below the ANI species threshold value of ≥95% (Richter & Rosselló-Móra, 2009) (Table 2.2)

supporting their distinction from their closest relatives. The genomes of strains P01-B04^T and P01-F02 were also compared between each other and showed a high similarity of 99.96%, further suggesting that the two isolates represent a single species.

Table 2.2. Genome-based comparisons showing the relationship of strains P01-B04^T and P01-F02 with three phylogenetically closely related *Streptomyces* species. Average nucleotide identity (ANI) was calculated using OrthoANI (I. Lee et al., 2016), and DNA-DNA hybridization (DDH) values were determined by the GGDC calculator v2.1 (Meier-Kolthoff et al., 2014).

Comparison	Strain	*S. poriferorum* sp. nov.		*S. beijiangensis*	*S. laculatispora*	*S. brevispora*
		P01-B04^T	P01-F02	DSM 41794^T	NRRL B-24909^T	NRRL B-24910^T
ANI (orthoANI) (%)	P01-B04^T	-	99.96	79.79	88.28	88.32
	P01-F02	99.96	-	79.81	88.29	88.30
	DSM 41794^T	79.79	79.81	-	79.81	80.15
	NRRL B-24909^T	88.28	88.29	79.81	-	92.51
	NRRL B-24910^T	88.32	88.30	80.15	92.51	-
DDH (GGDC) (%)	P01-B04^T	-	99.60	23.50	36.30	35.90
	P01-F02	99.60	-	23.50	36.30	35.80
	DSM 41794^T	23.50	23.50	-	23.70	23.80
	NRRL B-24909^T	36.30	36.30	23.70	-	49.70
	NRRL B-24910^T	35.90	35.80	23.80	49.70	-

Similar to the ANI results, DDH values calculated by the GGDC algorithm revealed a high degree of similarity (99.6%) between the genomes of strain P01-B04^T and strain P01-F02 (Table 2.2). The closest related species determined by DDH values was *S. laculatispora* NRRL B-24909^T with 36.3% similarity to strains P01-B04^T and P01-F02, a value far below the cut-off point of 70% recommended for species delineation (Auch et al., 2010). DDH values between P01-B04^T and the other reference strains *S. brevispora* DSM 42059^T and *S. beijiangensis* DSM 41794^T were 35.9% and 23.5%, respectively. These values were observed to be nearly identical between P01-F02, *S. brevispora* DSM 24059^T, and *S. beijiangensis* DSM 41794^T. Taken

together, the ANI and DDH comparisons support the conclusion that strains P01-B04^T and P01-F02 belong to the same novel species of the genus *Streptomyces*.

2.4.3. Morphological and physiological characteristics.

The observed cultural and morphological features of strains P01-B04^T and P01-F02 aligned with the characteristics typical of the genus *Streptomyces*. Colonies of both strains were filamentous and growth was observed on all tested media types at 30 °C for 7 – 14 days. The colors of the aerial and substrate mycelium varied depending on the media type, and differences in colony color were observed between the two strains in relation to their closest phylogenetic neighbors and to each other (Supplementary Table S2.1). Aerial mycelia observed when strain P01-B04^T was grown on ISP 2 were pale yellow whereas the aerial mycelia produced by strain P01-F02 when grown under the same conditions appeared grey-white in color. No melanin production was observed during growth on ISP 6 or ISP 7 by either strain. Well-developed aerial hyphae (Supplementary Figure S2.4) were observed using SEM after incubating strains P01-B04^T and P01-F02 for 5 days on ISP 2, and the hyphae later differentiated into spherical spores after the isolates were cultured for an additional three days on ISP 2.

In addition to morphological and cultural characteristics, strains P01-B04^T and P01-F02 were investigated for phenotypic and biochemical features that could distinguish the novel isolates from other *Streptomyces* taxa (Table 2.3). The novel isolates and reference strains were psychro-tolerant, with the ability to grow at temperatures as low as 3 °C; however, the growth rate was much slower compared to the optimum temperature of 25 – 30 °C. In addition, strain P01-B04^T grew at pH 4.5 – 9.5 (optimum pH 7.0), which was distinguishable from *S. brevispora* NRRL B-24910^T and *S. laculatispora* NRRL B-24909^T that tolerated a pH range of 5 – 12. In comparison, strain P01-F02 grew at pH 5 – 10. Strain P01-F02 exhibited the highest NaCl

tolerance in ISP 2 broth (up to 9% w/v NaCl). This was significantly higher than the NaCl tolerance demonstrated by strain P01-B04^T (up to 5% w/v NaCl), implying that NaCl tolerance is a strain-specific characteristic within *S. poriferorum*, and also higher than among reference strains – *S. beijiangensis* DSM 41794^T (up to 1.2% w/v NaCl), *S. brevispora* NRRL B-24910^T and *S. laculatispora* NRRL B-24909^T (up to 5% w/v NaCl). Isolate P01-B04^T tolerated the same NaCl concentrations as *S. brevispora* NRRL B-24910^T and *S. laculatispora* NRRL B-24909^T, but differed from *S. beijiangensis* DSM 41794^T.

Phenotypic traits of the isolates including carbon source utilization, hydrolysis tests, enzyme activity, and resistance to antibiotics were assessed in parallel to the type strains. Several phenotypic characteristics including the hydrolysis of L-tyrosine, resistance to kanamycin and neomycin, weakly positive enzyme activity of α-chymotrypsin, and the utilization of L-arabinose were found to separate the isolates P01-B04^T and P01-F02 from one or more of their closest relatives (Table 2.3). Oxidase activity, nitrate reduction, hydrolysis of Tween 80, utilization of D-lactose, L-rhamnose, raffinose, and β-galactosidase activity among other features were also found to be unique to P01-B04^T when compared to strain P01-F02 (Supplementary Table S2.2). All phenotypic features observed for both isolates are shown in Supplementary Table S2.2.

Table 2.3. Summary of characteristics that differentiate strains P01-B04^T and P01-F02 from type strains of the closest related species of the genus *Streptomyces*.

Characteristics	*S. poriferorum* sp.nov P01-B04^T	P01-F02	*S. beijiangensis* DSM 41794^T	*S. laculatispora* NRRL B-24909^T	*S. brevispora* NRRL B-24910^T
Morphology (on ISP 2):					
Color of aerial mycelium	Pale yellow	Grey white	Yellow	White	Beige
Color of substrate mycelium	Pale yellow	Green	Yellow	Pale yellow	Dark purple brown
Urease production	+		-	+	-
Hydrolysis of:					
L-tyrosine	+		-	+	+
Growth with/at:					
Temperature range (°C)	3-30	3-30	3-30	3-30	3-30
pH range	4.5-9.5	5.0-10.0	4.5-9.5	5.0-12.0	5.0-12.0
Tolerance on NaCl (%)	5.0	9.0	1.2	5.0	5.0
Carbon source utilization 1% (w/v):					
Salicin	-		+	+/-	+/-
D-sorbitol	-		-	+	-
L-arabinose	+		+	+	-
Enzyme activity (API ZYM):					
α-chymotrypsin	(+)		-	-	-
β-galactosidase	v		+	+	+
β-glucosidase	(+)		+	+	-
Cystine arylamidase	v		+	-	(+)
Lipase C14	v		-	-	-
Leucine arylamidase	(+)		+	+	+
N-Acetyl-β-glucosaminidase	v		+	+	+
Trypsin	v		-	-	-
Valine arylamidase	v		+	-	+
Resistance to antibiotics:					
Kanamycin (30 µg)	-		+	+	+/-
Neomycin (5 µg)	-		+	-	-

Note: +, positive; (+), weakly positive; -, negative; v, variable result observed between strain P01-B04^T and P01-F02; +/-, phenotypic result observed in this study differed from result in original type strain species descriptions by W.-J. Li et al. (2002) for *S. beijiangensi*s DSM 41794^T and Zucchi et al. (2012) for the equivalent type strains *S. brevispora* BK160^T and *S. laculatispor*a BK166^T.

Additionally, when applicable, the observed phenotypic properties of the type strains in this study were compared to the original species descriptions by W.-J. Li et al. (2002) for *S. beijiangensis* DSM 41794^T and Zucchi et al. (2012) for the equivalent type strains *S. brevispora*

BK160^T and *S. laculatispora* BK166^T. Most characteristics in this study were consistent with the original type strain descriptions including positive hydrolysis of starch, utilization of D-glucose, D-xylose, glycerol, negative utilization of sucrose, and susceptibility to gentamicin and tetracycline for all three strains. Differences were also noted; for example, *S. beijiangensis* DSM 41794^T and *S. laculatispora* BK166^T were observed to be positive for D-mannose and cellulose utilization in this study, but were negative in their original species descriptions. Macroscopic morphology on ISP 3 of *S. brevispora* BK160^T was identical to the species description of W.-J. Li et al. (2002), but differed in this study for *S. beijiangensis* DSM 41794^T and *S. laculatispora* BK166^T compared to Zucchi et al. (2012). These differences in phenotypic traits are not unexpected since the conditions used for cultivation and analysis of the type strains may differ between the studies; however, any variation in the phenotypic traits listed in Table 2.3 were noted.

2.4.4. Chemotaxonomic properties.

The results from chemotaxonomic analysis of strains P01-B04^T and P01-F02 supported their classification as members of the genus *Streptomyces*. The diagnostic diamino acid of the peptidoglycan was LL-diaminopimelic acid, and the major whole-cell sugars were glucose and ribose with minor amounts of mannose. The polar lipid profiles for both strains consisted of the major compound diphosphatidylglycerol, moderate amounts of phosphatidylethanolamine, phosphoaminolipid, two unidentified phospholipids (PL1-2), one unidentified phosphoglycolipid, and minor traces of one unidentified glycolipid, three unidentified aminolipids (AL1-3) and two unidentified phospholipids (PL3-4) (Supplementary Figure S2.5). The menaquinones detected were nearly identical for the two strains, P01-B04^T and P01-F02,

and consisted of MK-9(H_6) (61.8%, 61.7%), MK-9(H_4) (7.0%, 6.8%), and MK-9(H_8) (31.2%, 31.4%).

Fatty acid analysis revealed that five predominant fatty acids (>5%) were common to both, P01-B04^T and P01-F02, namely anteiso-$C_{15:0}$ (31.0%, 27.7%), iso-$C_{16:0}$ (25.6%, 16.9%), anteiso-$C_{17:0}$ (12.5%, 19.2%), $C_{16:0}$ (7.1%, 8.6%), and iso-$C_{15:0}$ (5.3%, 6.0%). Although the overall fatty acid profile was similar between the two isolates, significant differences in the amounts of fatty acid present were observed for iso-$C_{14:0}$, iso-$C_{16:0}$, and anteiso-$C_{17:0}$. The fatty acid profiles of both strains were also similar to that of the closest related reference strains, further confirming their affiliation with the genus *Streptomyces*. The most notable difference between the fatty acid content of the novel marine isolates to the reference strains was observed for anteiso-$C_{15:0}$, which was significantly higher for *S. beijiangensis* DSM 41794^T (44.8%), *S. laculatispora* NRRL B-24909^T (37.6%), and *S. brevispora* NRRL B-24910^T (42.8%) compared to P01-B04^T (31.0%) and P01-F02 (27.7%). A full comparison of the fatty acid profiles of strains P01-B04^T, P01-F02, and type strains of the closest relatives is presented in Table 2.4.

Table 2.4. Cellular fatty acid composition of strains P01-B04^T and P01-F02 and their closest phylogenetic *Streptomyces* neighbors.

Fatty acid (%)	*S. poriferorum* sp. nov. P01-B04^T	P01-F02	*S. beijiangensis* DSM 41794^T	*S. laculatispora* NRRL B-24909^T	*S. brevispora* NRRL B-24910^T
$C_{16:0}$	7.1	8.6	7.8	7.8	7.4
$C_{17:0}$	1	1.9	TR	TR	TR
iso-$C_{14:0}$	7.2	2.9	1.8	6.3	5
iso-$C_{15:0}$	5.3	6	3.2	4.4	5.3
iso-$C_{16:0}$	25.6	16.9	8.7	20.5	18.6
iso-$C_{17:0}$	3.9	3.6	3.1	2.8	2.3
anteiso-$C_{15:0}$	31	27.7	44.8	37.6	42.8
anteiso-$C_{17:0}$	12.5	19.1	21.5	12.8	12
isoH-$C_{16:1}$	1	2.3	TR	1	TR
SF 3*	1.3	2.1	1.1	1.7	1.4
SF 9$^{\Delta}$	1	1.4	TR	1	TR
anteiso-$C_{17:1}$ *ω9c*	1	3.2	2	1.4	1
$C_{17:1}$ *ω8c*	TR	1	TR	TR	TR

Note: Values of less than 1% are reported as TR – traces; *Summed Feature 3 $C_{16:1}$ *ω7c*/ $C_{16:1}$ *ω6c*; $^{\Delta}$ Summed Feature 9 $C_{16:0}$ 10-methyl/ $C_{17:1}$ iso *ω9c*. All data were obtained in this study under identical growth conditions for each strain in tryptic soy broth for 10 days at 28 °C with shaking at 150 rpm.

2.4.5. *In silico* genome annotations and biosynthetic gene cluster analysis.

Genome assemblies of the novel isolates and reference strains were annotated using RAST and antiSMASH web servers. RAST annotation predicted that the majority of annotated genes across 323 subsystems from strain P01-B04^T were involved in amino acid and derivatives metabolism (20.0%), carbohydrate metabolism (15.8%), protein metabolism (10.2%), and cofactor, vitamin, prosthetic group, and pigment metabolism (8.2%) (Supplementary Figure S2.6). Genes linked to secondary metabolism made up 0.3% of the genome of strain P01-B04^T. Annotations predicted by RAST for strain P01-F02 (Supplementary Figure S2.6) were highly similar to those for strain P01-B04^T, further supporting their designation as one species.

Genome mining for potential secondary metabolite associated BGCs was performed using antiSMASH version 5.0 and to reduce redundancy in our analysis the data was manually curated to ensure that clusters counted were not split between multiple contigs. This led to the

identification of 41 gene clusters for strain P01-B04^T and 42 BGCs associated with strain P01-F02. The majority of these putative BGCs from both strains were classified as nonribosomal peptide synthetases (NRPSs) and type I polyketide synthases (T1PKSs), which represent biosynthetic pathways dominant in the genomes of many Actinobacteria and known to synthesize diverse, pharmaceutically active compounds (Komaki et al., 2018). Other commonly predicted BGCs for both strains included butyrolactones, terpenes, and PKS-NRPS hybrids. Many of the predicted gene clusters were conserved between the two isolates, aligning with the high degree of genome similarity that was observed from whole genome comparisons.

The genomes of the reference strains were also analyzed for BGCs using antiSMASH to compare the biosynthetic potential of the novel isolates to their closest relatives. Overall, 30 clusters were predicted from *S. beijiangensis* DSM 41794^T, a total of 22 BGCs from *S. brevispora* DSM 42059^T, and 32 clusters from *S. laculatispora* NRRL B-24909^T. Although the number of BGCs from the reference strains were lower in comparison to strains P01-B04^T and P01-F02, this may be due to the smaller genome sizes of the reference strains with respect to the novel isolates. Similar to the novel isolates, NRPSs were the dominant classes of predicted BGCs for *S. laculatispora* NRRL B-24909^T (8 clusters) and *S. brevispora* DSM 42059^T (5 clusters) (Fig. 2.3). For *S. beijiagensis* DSM 41794^T, terpenes (4 clusters), RiPP-like clusters (4 clusters), and butyrolactones (4 clusters) represented the majority of predicted BGCs (Fig. 2.3).

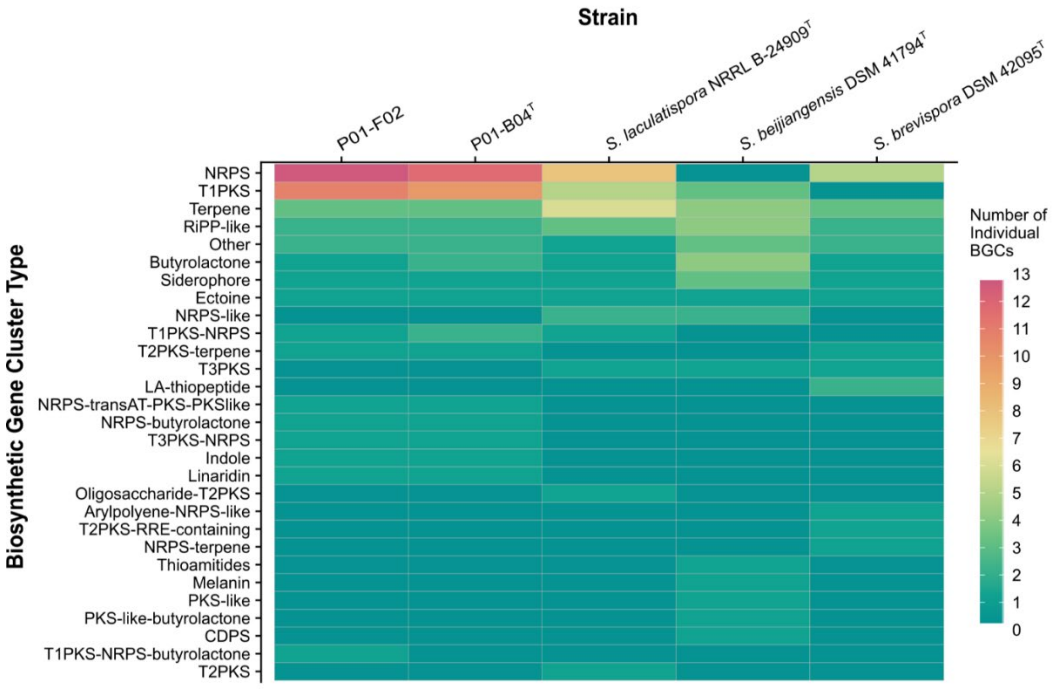

Figure 2.3. Number and type of biosynthetic gene clusters (BGCs) predicted by antiSMASH (Blin et al., 2019) in the genomes of the novel isolates and reference strains. Clusters are arranged top to bottom, beginning with the greatest number of BGC types in the top left. Strains are shown left to right by the highest number of BGCs. The most abundant BGCs were NRPS clusters followed by Type I PKSs for strains P01-B04^T and P01-F02.
Note: BGCs annotated as redox cofactors, NAPAA, RRE-containing, and hgle-KS were grouped into "Other".

In-depth analysis of the BGC annotations revealed that 14 gene clusters from P01-B04^T and 12 BGCs from P01-F02 showed no similarity to any characterized gene clusters deposited in the MIBiG database (Kautsar et al., 2020). Among the clusters that matched to a known BGC, only 4 predicted gene clusters from strain P01-B04^T and 8 putative BGCs from P01-F02 were >50% similar to a known cluster. Interestingly, a core structure was predicted from 19 BGCs for strain P01-B04^T including six that did not match any known BGC. In comparison, a core scaffold was predicted for 16 BGCs from the genome of strain P01-F02 with three being orphan BGCs. Among the gene clusters with a predicted core structure, the largest predicted multi-domain BGC from both isolates spanned over 80 Kb and was classified as a NRPS-transAT-Type I PKS

hybrid with three predicted peptide products from three candidate clusters (Fig. 2.4). This hybrid BGC, named Cluster 1, showed low similarity (11%) to a borrelidin T1PKS cluster (BGC_0000031) (Olano et al., 2004) derived from *Streptomyces parvulus* Tü4055 that encodes a macrolide with antibacterial (Berger, 1949; Singh et al., 1985), antiviral (Dickinson et al., 1965), and anti-angiogenesis activity (Wakabayashi et al., 1997). The overall low degree of similarity to known clusters predicted from the genomes of strains P01-B04^{T} and P01-F02, including Cluster 1, and the identification of orphan gene clusters suggests that these strains may produce several novel bioactive metabolites.

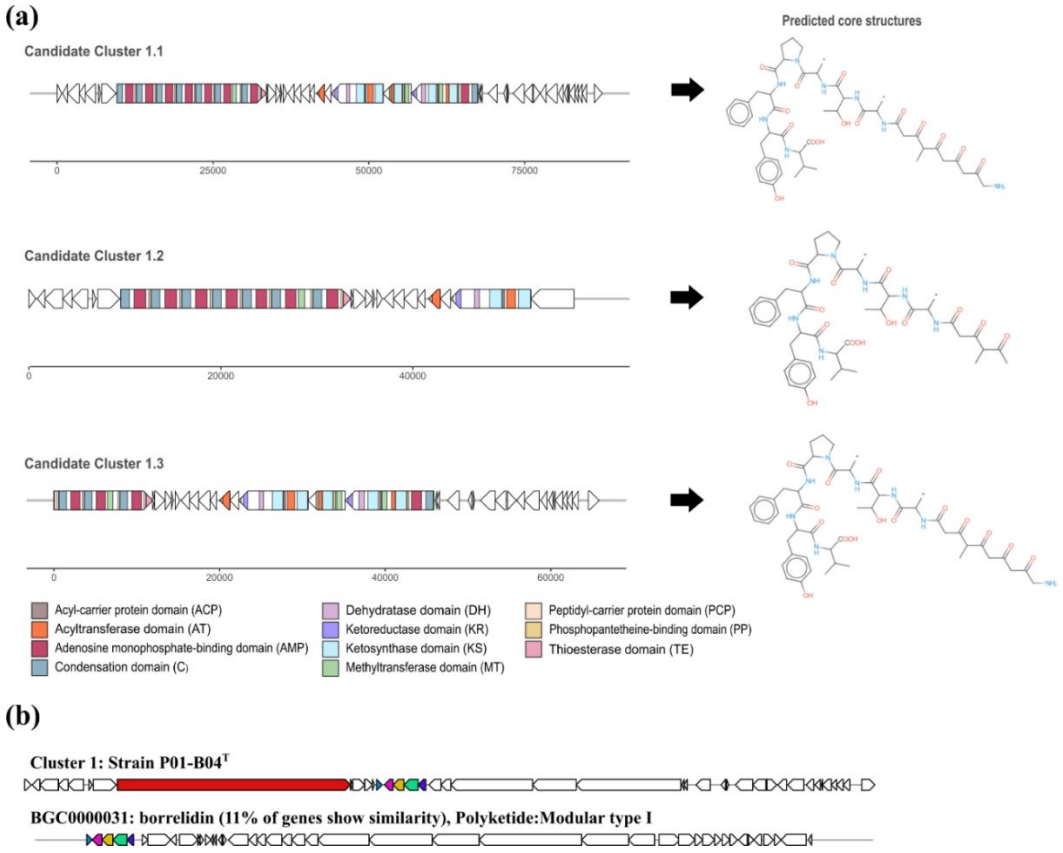

Figure 2.4. A polyketide synthase (PKS)-nonribosomal peptide synthetase (NRPS) hybrid BGC, Cluster 1, predicted by antiSMASH (Blin et al., 2019) from the genome of strain P01-B04^{T} (shown below) and P01-F02. **(a)** antiSMASH annotation of the three candidate clusters from Cluster 1 and their predicted core structures. ORFs are depicted as white arrows and the BGC domains are color coded in each ORF. **(b)** Similarity of Cluster 1 from strain P01-B04^{T} to a Type I PKS borrelidin cluster (BGC_0000031) from *Streptomyces parvulus* Tü4055 (Olano et al., 2004).

2.4.6. Antibacterial activity of strains P01-B04^T and P01-F02.

To further elucidate the biosynthetic potential of strains P01-B04^T and P01-F02, antimicrobial activity assays were performed against target organisms that included human and plant pathogens. The results from these assays indicated that supernatants from strain P01-B04^T were capable of inhibiting the growth of several Gram-positive bacteria, including multidrug-resistant *S. aureus* strains when cultured for 7 – 35 days. Bioactivity was not observed from either of the *S. poriferorum* strains against *E. coli* BL21 or *C. albicans* ATCC 90028. The highest degree of antibacterial activity (largest zone of inhibition) against *S. aureus* strains and *M. luteus* ATCC 10240 by isolate P01-B04^T was observed when the strain was cultured in YEME broth. In contrast, supernatants from strain P01-F02 did not inhibit the growth of *S. aureus* or *M. luteus* when cultured under the same conditions as strain P01-B04^T. This finding suggests that strain P01-B04^T may be expressing BGC(s) that were not shared with P01-F02 leading to the production of unique metabolites that inhibit growth of MRSA and/or *M. luteus*, or that the media and growth conditions used were not adequate to support expression of the BGCs and subsequent synthesis of antimicrobial metabolites by strain P01-F02. Additionally, only extracellular activity was measured in these bioassays, therefore, this could also suggest that secretion of the antimicrobial compound(s) may be more active from strain P01-B04^T.

Test strains for antimicrobial activity assays were expanded to include two plant pathogen strains, *Clavibacter michiganensis* subsp. *michiganensis* 89C-4 and *Curtobacterium flaccumfaciens* subsp. *flaccumfaciens* CV3, to assess the prospective potential of the novel isolates as biological control agents in agriculture. Interestingly, both, P01-B04^T and P01-F02 showed inhibitory activity against the plant pathogens but only when cultured in YEME broth. These results emphasize that strains P01-B04^T and P01-F02 may be promising candidates for the

production of metabolites with inhibitory activities against human and plant pathogens. The full antimicrobial activity profiles of each strain are shown in Supplementary Table S2.3.

2.4.7. Ecological role and host specificity of the sponge-derived isolates.

The sponge microbiome represents a complex relationship between the host and its microbial inhabitants that often encompasses a range of ecological interactions including mutualism (C. R. Wilkinson, 1983), commensalism (Taylor et al., 2007), and even parasitism (Bavestrello et al., 2000; Webster et al., 2002). Despite the growing interest in understanding these holobiome relationships, unraveling the intricacies of a particular symbiont's role within its host sponge are challenging and often rely on inferences from sequencing data (Webster & Taylor, 2012; Webster & Thomas, 2016). Moreover, the extent of habitat- and host-related variation in sponge symbiont assemblages remains understudied (Cleary et al., 2013). We sought to investigate the potential distribution of *S. poriferorum* across different sponge microbiomes. Using a BLAST comparison, all 245 available sponge metagenomes from the JGI IMG/M database were queried for the presence of related sequences to the 16S rRNA gene sequences of the isolates of interest, P01-B04^{T} and P01-F02. This database included sequences of distantly related sponges collected from different geographic locations including the Mediterranean Sea, Adriatic Sea, and the Great Barrier Reef, but was notably lacking in sequences from psychrophilic sponge metagenomes. The closest sequence match was 87% identical to a 16S rRNA gene recovered from the marine sponge *Aplysina aerophoba* sampled from the Gulf of Piran in the Adriatic Sea (JGI ID:3300027391). The low percent identity to other sponge microbiomes could indicate that these symbionts may be specialists that are only affiliated with specific sponge species, and there has been a paucity of sampling from psychrophilic sponge species. This notion of sponge species-specificity has been supported in other studies; for

example, a comparative analysis by Webster et al. (2013) of microbial communities of 13 taxonomically diverse sponge species from the Great Barrier Reef found that the microbiome shared among sponges was low even within sponges collected from the same locality. Additionally, Taylor et al. (2004) showed that the microbial communities of three sponges, *Cymbastela concentrica*, *Callyspongia* sp., and *Stylinos* sp., varied little within each species, but varied significantly between species.

The bioactivity exhibited by strains P01-B04^T and P01-F02 suggests that an ecological role of these presumed specialists may be to provide their host with chemical defense through the production of bioactive metabolites. It is well established that the sessile lifestyle of sponges has encouraged their dependency on bioactive secondary metabolites that provide protection against predators (Baig et al., 2020; E. Schmidt et al., 2000), pathogens (Hentschel et al., 2001), and fouling agents (Unson et al., 1994). These isolates may provide a similar protective function and in return, the microorganisms may be provided with shelter within the tissues of the host sponge from predators or high light levels (Sara, 1971) and supplied with essential nutrients for growth and metabolism (Taylor et al., 2007; Webster & Taylor, 2012).

In addition to their putative chemical defense roles, we predict that strains P01-B04^T and P01-F02 may play a role in microbially mediated nutrient cycling processes in their host sponges. In support of this prediction, annotations using RAST and KAAS predicted 15 functional ortholog groups involved in nitrogen metabolism present in the genome of strain P01-B04^T and 16 in the genome of strain P01-F02, which included predicted functional roles in ammonia assimilation and nitrate/nitrite reduction. Other studies have provided evidence of nitrogen fixation (C. Wilkinson & Fay, 1979; F. Zhang et al., 2014) and transformation roles (Weigel & Erwin, 2017) of sponge-associated bacteria where the symbionts can act as a source

of bioavailable nitrogen for the host (F. Zhang et al., 2019). Beyond nitrogen fixation, another predicted role of these *Streptomyces* strains in sponge nutrient cycling is phosphate metabolism, as genes predicted to be involved in phosphate and phosphonate metabolism were annotated in the genomes of strains P01-B04^T and P01-F02. A previous study found that Actinobacteria sponge symbionts are capable of solubilizing phosphate compounds, suggesting that they may provide a phosphorous source to their host (Sabarathnam et al., 2010; F. Zhang et al., 2019). However, the most compelling evidence of microbially mediated nutrient cycling in sponges is from the well-characterized example of Cyanobacteria (F. Zhang et al., 2019), and the ecological role that Actinobacteria play in sponge nutrient cycles is not clear. The inferences made in this study regarding the putative distribution and ecological role of *S. poriferorum* are constrained by the low number of sequences from other sponge metagenomes, particularly those from cold-water sponges. Future work will seek to investigate the prevalence and abundance of *S. poriferorum* within their original sponge species (*G. barretti* and *A. dichotoma*) and other sponges from cold-water ecosystems. Understanding the degree of host specificity of the isolates will provide more insight on the co-evolutionary history between host and symbiont, as well as a better understanding of how critical these taxa may be within the sponge microbiome and how that may influence their functional roles**.**

2.5. Conclusion

Two novel *Streptomyces* strains with antibacterial activity, P01-B04^T and P01-F02, were isolated from marine sponges in the Trondheim fjord of Norway. Colony morphology, 16S rRNA gene sequence similarity, and chemotaxonomic properties were comparable to that of validly described *Streptomyces* taxa. Differences in physiological characteristics and low ANI and DDH values clearly distinguished strains P01-B04^T and P01-F02 from their closest

phylogenetic neighbors. Although there were phenotypic and genotypic differences observed between strains P01-B04^T and P01-F02, rigorous phylogenomic analyses indicated that these two strains were highly similar and affiliate with the same species. Genome annotations revealed that the isolates harbor on average 41 BGCs, many of which were predicted to be uniquely present in these isolates. This finding together with the observed antimicrobial activity profiles of the isolates suggest that strains P01-B04^T and P01-F02 are capable of producing bioactive metabolites that may contribute to the chemical ecology of their host sponges, as well as have potential clinical efficacy against multidrug-resistant human pathogens. These antibiotic compounds may have applications in human medicine, veterinary medicine and/or agriculture, and ongoing work to determine the structure of these compounds will allow assessment of their clinical and agrochemical potential. Based on the polyphasic taxonomic analysis performed in this study, we propose that strains P01-B04^T and P01-F02 represent a novel species within the genus *Streptomyces,* for which the name *Streptomyces poriferorum* sp. nov. is proposed, with P01-B04^T (=DSM 111306^T =CCM 9048^T) as the type strain.

2.6. Chapter 2 Supplementary Material

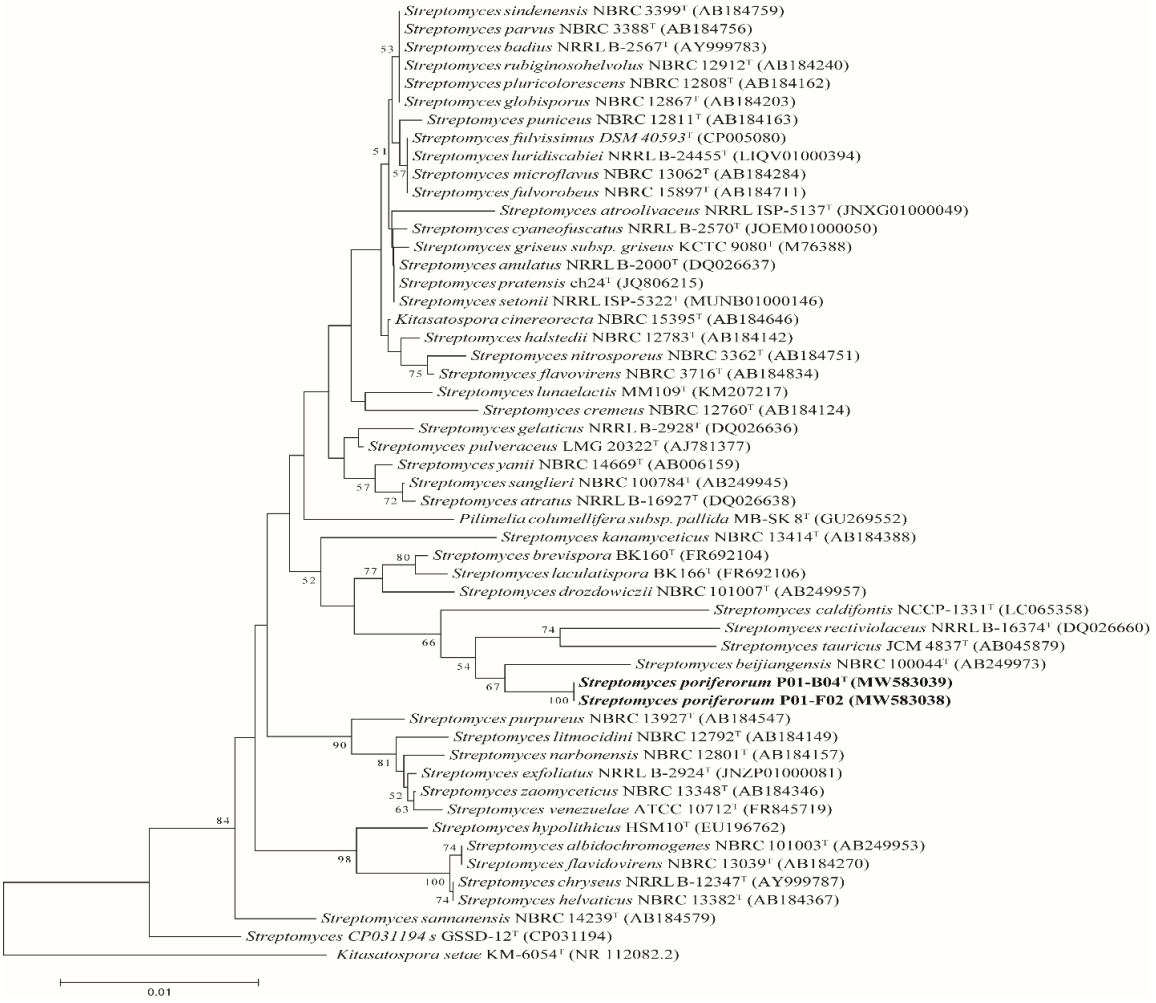

Figure S2.1. Neighbor joining (NJ) phylogenetic tree based on 16S rRNA gene sequences depicting the relationship between strains P01-B04^T, P01-F02, and closely related taxa of the genus *Streptomyces*. *Kitasatospora setae* strain KM 6054^T (NR_112082.2) was used as an outgroup. Numbers at branch nodes indicates bootstrap values based on 1000 replications and only values >50% are shown. Bar, 0.01 substitutions per nucleotide position.

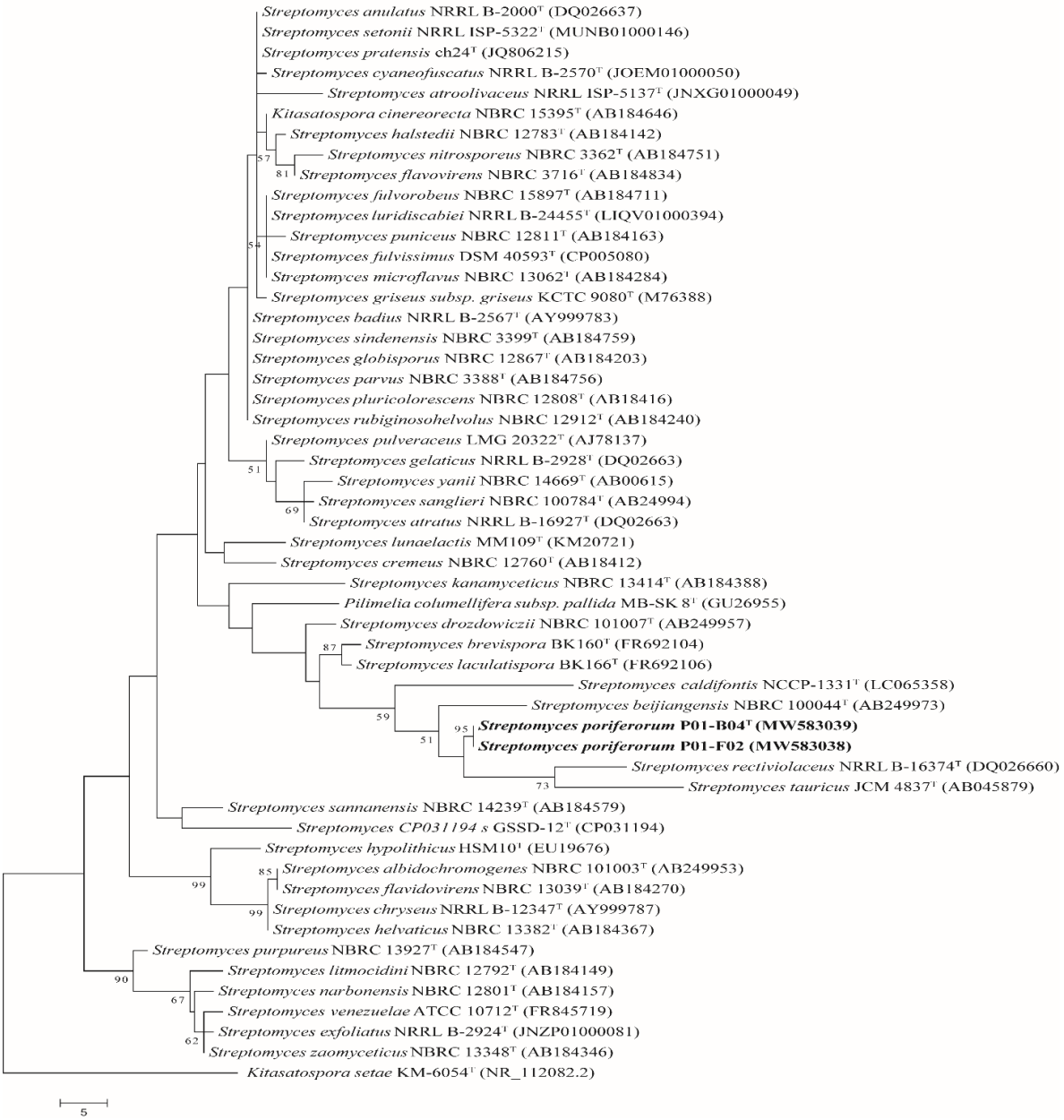

Figure S2.2. Maximum parsimony (MP) 16S phylogenetic tree based on 16S rRNA gene sequences depicting the relationship between strain P01-B04^T, P01-F02, and closely related taxa of the genus *Streptomyces*. *Kitasatospora setae* strain KM 6054^T (NR_112082.2) was used as an outgroup. Numbers at branch nodes indicates bootstrap values based on 1000 replications and only values >50% are shown. Bar, 5 substitutions per nucleotide position.

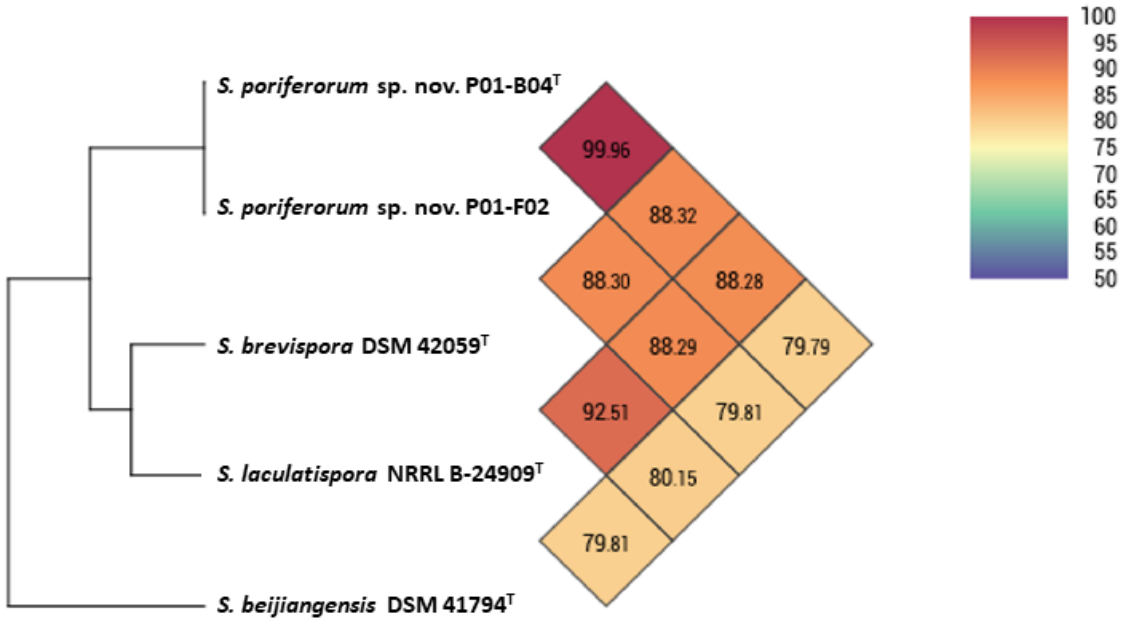

Figure S2.3. Average nucleotide identity (ANI) heat map of ANI values between the closely related strains generated by OrthoANI software (I. Lee et al., 2016).

Table S2.1. Cultural characteristics on various media types of strains. All data were determined during this study.

Medium	*Streptomyces poriferorum* sp. nov. P01-B04^T	P01-F02	*S. beijiangensis* DSM 41794^T	*S. laculatispora* NRRL B-24909^T	*S. brevispora* NRRL B-24910^T
ISP 2					
Aerial mycelium	Pale yellow	Grey white	Yellow	White	Beige
Substrate mycelium	Pale yellow	Green	Yellow	Pale yellow	Dark purple brown
ISP 3					
Aerial mycelium	Grey	White	Pale yellow	White	White
Substrate mycelium	Pale brown	White	Pale yellow	White	White
ISP 4					
Aerial mycelium	Grey	Pale yellow	No growth	Green	White
Substrate mycelium	Dark grey	Greenish orange	No growth	White	Pale yellow
ISP 5					
Aerial mycelium	Greyish white	White	White	White	White
Substrate mycelium	Greyish white	Pale yellow	White	White	White
ISP 6					
Aerial mycelium	Yellow	Pinkish white	Yellow	White	White
Substrate mycelium	Yellow	Yellow	Yellow	Yellow	Yellow
ISP 7					
Aerial mycelium	Grey	Pale yellow	Yellow	Orange	White
Substrate mycelium	Yellow	Green	Yellow	Brown	Yellow
Nutrient agar					
Aerial mycelium	Pale yellow	White	Pale yellow	Pale yellow	Pale yellow
Substrate mycelium	Pale yellow	Pale yellow	Pale yellow	Pale yellow	Pale yellow
Czapek's Dox agar					
Aerial mycelium	Grey	Greyish white	Yellow	Yellowish white	Yellow
Substrate mycelium	Dark grey	Yellowish green	Yellow	Yellowish brown	Yellow
Modified Bennett's agar					
Aerial mycelium	White	White	Pale yellow	White	White
Substrate mycelium	Brown	White	Pale yellow	White	White

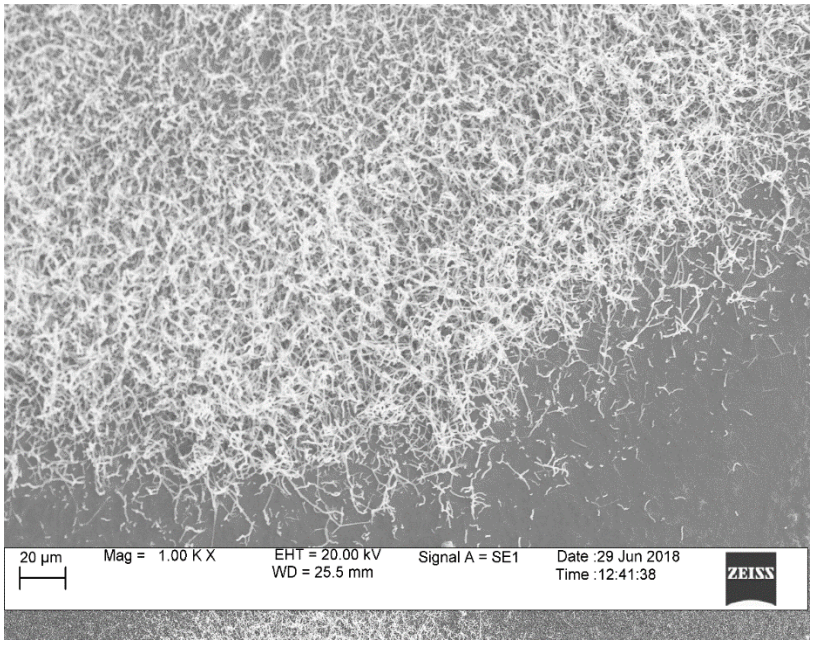

Figure S2.4. Scanning electron micrograph (SEM) of strain P01-B04^{T} showing abundant aerial hyphae following growth on ISP 2 at 30°C for 5 days. Bar, 20 µm.

Table S2.2. Full list of phenotypic characteristics of *S. poriferorum* sp. nov. P01-B04^T.

Characteristics	P01-B04^T	P01-F02
Growth with/at:		
Temperature range (°C)	3-30	3-30
pH range	4.5-9.5	5-10
Tolerance on NaCl (%)	5.0	9.0
Catalase	+	+
Oxidase	+	-
Nitrate reduction	+	-
Urease production	+	+
Melanin production on ISP 6 and ISP 7	-	-
Hydrolysis of:		
DNA	+	+
Gelatin	+	+
Tween 80	+	-
Starch	+	+
Casein	+	+
L-tyrosine	+	+
Esculin	+	+
Hypoxanthin	+	+
Xanthin	+	+
Utilization of cellulose in ISP 2 broth	+	-
Carbon source utilization 1% (w/v):		
D-glucose	+	+
Sucrose	(+)	-
D-fructose	+	+
myo-inositol	-	-
L-rhamnose	+	-
Raffinose	+	-
D-mannitol	-	+
D-xylose	+	+
L-arabinose	+	+
D-ribose	-	+
D-mannose	+	+
Glycerol	+	+
D-sorbitol	-	-
Salicin	-	-
D-galactose	+	+
D-lactose	+	-
Enzyme activity (API ZYM):		
Alkaline phosphatase	(+)	+
Esterase (C 4)	+	+
Esterase lipase (C 8)	+	+
Lipase (C 14)	-	+
Leucine arylamidase	(+)	+
Valine arylamidase	-	+
Cystine arylamidase	-	(+)
Trypsin	-	(+)
α-chymotrypsin	(+)	+
Acid phosphatase	+	(+)
Naphtol-AS-BI-phosphohydrolase	+	(+)
α-galactosidase	-	-
β-galactosidase	+	-
β-glucuronidase	-	-

α-glucosidase	(+)	+
β-glucosidase	(+)	+
N-Acetyl-β-glucosaminidase	-	+
α-mannosidase	-	-
α-fucosidase	-	-
Resistance to antibiotics:		
Ampicillin (10 μg)	-	+
Chloramphenicol (30 μg)	I	-
Gentamicin (10 μg)	-	-
Kanamycin (30 μg)	-	-
Neomycin (5 μg)	-	-
Streptomycin (10 μg)	-	-
Tetracycline (30 μg)	I	-
Penicillin G (10 μg)	-	+
Clindamycin (2 μg)	I	+
Antimicrobial activity against:		
Micrococcus luteus ATCC 10240	+	-
Staphylococcus aureus Xen29	+	-
Staphylococcus aureus MRSA30	+	-
Curtobacterium flaccumfaciens subsp. flaccumfaciens CV3	+	+
Clavibacter michiganensis 89C-4	+	+

Note: +, positive; (+), weakly positive; -, negative; I, intermediate antibiotic resistance.

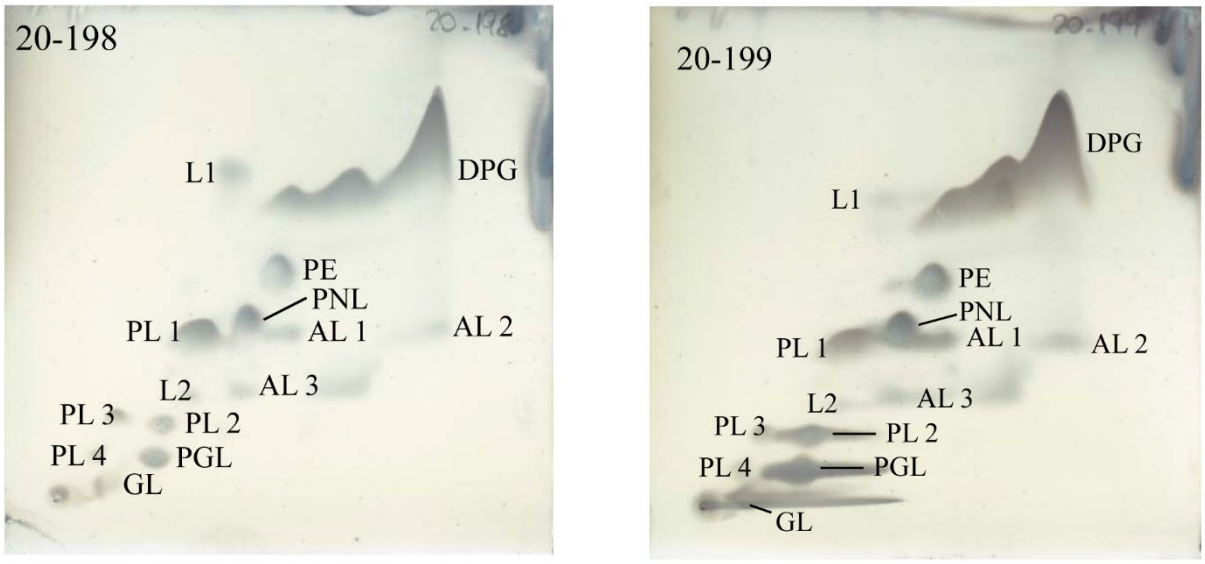

Figure S2.5. Polar lipid profile of (a) strain P01-B04^T and (b) strain P01-F02 after separation by two-dimensional TLC. AL, aminolipid; DPG, diphosphatidylglycerol; GL, glycolipid; L, lipid; PE, phospatidylethanolamine; PGL, phosphoglycolipid; PL, phospholipid; PNL, phosphoaminolipid.

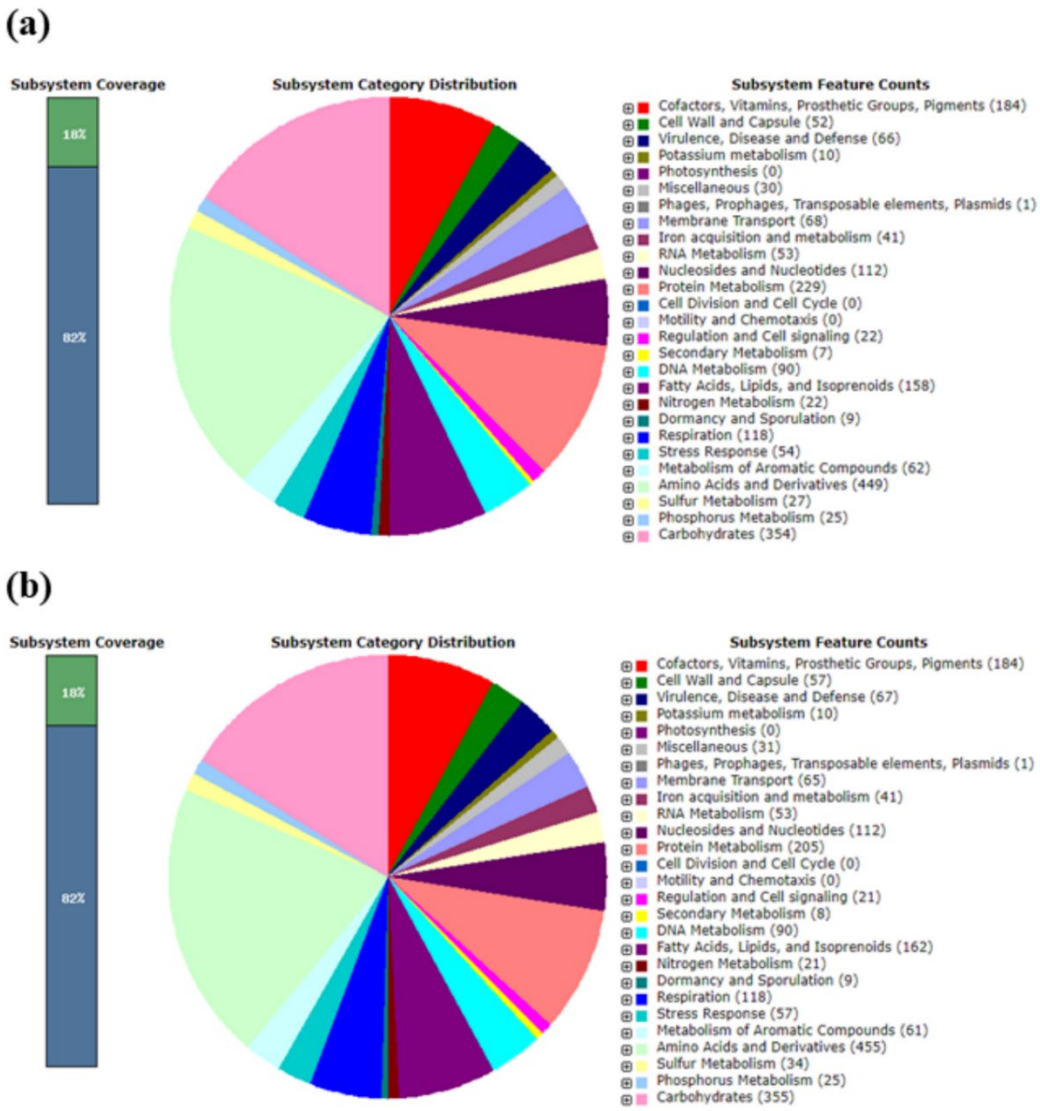

Figure S2.6. Subsystem category distribution of (a) strain P01-B04^{T} and (b) strain P01-F02 based on RAST annotation server (Aziz et al., 2008).

Table S2.3. Antimicrobial activity of strain P01-B04^T and strain P01-F02 when cultured in different media types measured by zones of inhibition against test microorganisms.

Bioassay microorganism	Strain P01-B04^T					Strain P01-F02				
	Distance of inhibition (mm)					Distance of inhibition (mm)				
	GYE	YEME	MF	TSB	ISP 2	GYE	YEME	MF	TSB	ISP 2
Gram-positive bacteria										
Staphylococcus aureus (MSSA) Xen29	-	9	-	-	-	-	-	-	-	-
Staphylococcus aureus (MRSA) EAMC30	-	7	-			-	-	-	-	-
Micrococcus luteus ATCC 10240	3	10	8	4	7	-	-	-	-	-
Clavibacter michiganensis subsp. *michiganensis* 89C-4	-	8	-	-	-	-	8	-	-	-
Curtobacterium flaccumfaciens subsp. *flaccumfaciens* CV3	-	7	-	-	-	-	9	-	-	-
Gram-negative bacteria										
Escherichia coli BL21	-	-	-	-	-	-	-	-	-	-
Fungi										
Candida albicans ATCC 90028	-	-	-	-	-	-	-	-	-	-

Chapter 3

Heterologous expression of metagenome-derived biosynthetic pathways in an engineered strain of *Streptomyces coelicolor*

3.1. Abstract

New antibiotics are needed to treat rapidly emerging multidrug-resistant infections. Soil microorganisms encode a vast diversity of bioactive natural products; however, most environmental microorganisms remain uncultured leaving their biosynthetic potential largely hidden. Significant advancements have been made in recent years to study the natural product potential encoded in the DNA of uncultured microbial communities (metagenomes). Despite these advancements, challenges associated with cloning and expressing biosynthetic gene clusters (BGCs) encoding natural product synthesis pathways have limited metagenomic drug-discovery platforms. To address these challenges and the societal need for new antibiotics, 90 BGCs predicted to encode novel chemistry and derived from a large-insert soil metagenomic library were conjugally transferred to an engineered *Streptomyces coelicolor* host to produce antibiotic compounds. Heterologous expression and functional screening resulted in the identification of 18 clones expressing inhibitory activity against the multidrug-resistant pathogen *Acinetobacter baumannii* and/or methicillin-sensitive *Staphylococcus aureus*, for an antibiosis hit rate of 20%. The antimicrobial activity of four of these heterologously-expressed metagenomic BGCs was enhanced when cloned in a dual-inducible expression vector. Genetic analysis of one clone of interest expressing antimicrobial activity, P17B06, indicated that this clone carries a nonribosomal peptide synthetase (NRPS)-like pathway that is unique from known pathways suggesting that it could encode novel compounds. Bioactivity-guided fractionation and LC-MS analysis identified metabolites produced by clone P17B06 that are enriched in bioactive

extracts and absent from the negative control. These results demonstrate that functional screening of a metagenomic library in an engineered *S. coelicolor* host can improve the low hit rates typically observed for metagenomic screens, thereby expanding our resources for antibiotic discovery from uncultured microorganisms.

3.2. Introduction

Multidrug-resistant (MDR) pathogens are a significant public health threat to modern medicine and are associated with an increase in morbidity, mortality, and economic costs (Alanis, 2005; Tanwar et al., 2014). In the United States alone, more than 2.8 million antibiotic-resistant infections occur each year and over 35,000 deaths annually are caused by these infections (Control & Prevention, 2019). Among drug-resistant microorganisms, carbapenem-resistant *Acinetobacter baumannii* and methicillin-resistant *Staphylococcus aureus* (MRSA) are the cause of many hospital-acquired nosocomial infections and can lead to a multitude of diseases including wound infections, pneumonia, infective endocarditis, and septicemia (Bulens et al., 2018; Kallen et al., 2010). The WHO recently listed *A. baumannii* as a "critical" concern and MRSA as a "high" priority pathogen in a list of 12 bacteria ranked according to critical, high, and medium priority for new antibiotic treatments (Asokan & Vanitha, 2018; Shrivastava et al., 2018). The efficacy of current antibiotic treatments against these MDR bacteria is waning and new antibiotics are needed (Asokan & Vanitha, 2018; Michael et al., 2014).

Natural products, or secondary metabolites, produced by microorganisms have been a historically rich source for antibiotics due to their extensive diversity in structure and function (Dias et al., 2012). The genes encoding these secondary metabolites, referred to as biosynthetic gene clusters (BGCs), are often physically located together in the genome of the producing organism and under tight regulation (Wenfang Wang et al., 2021). Microbially-derived BGCs are

responsible for the synthesis of many important and highly diverse classes of natural products including polyketides and nonribosomal peptides (H. Wang et al., 2014). These particular compounds are synthesized on large enzyme complexes known as polyketide synthases (PKSs) and nonribosomal peptide synthetases (NRPSs) that act in an assembly-line system to build products from carboxylic acid and amino acid monomeric precursors, respectively (Fischbach & Walsh, 2006). In some cases, PKSs and NRPSs can form mixed assembly systems leading to the production of hybrid polyketide-nonribosomal peptide products (Liangcheng Du et al., 2001). More than 23,000 polyketide and nonribosomal peptide natural products have been characterized (Y. Wei, Zhang, et al., 2018), some of which include the widely used antibiotics erythromycin, tetracycline, vancomycin, and daptomycin (Pham et al., 2019).

Despite the historical benefits of these compounds, the pipeline for natural product discovery has steadily declined over the past three decades (Freire-Moran et al., 2011). This shift is largely attributed to the shortcomings of classical methods which relied on culturing microorganisms in a laboratory to screen them for the production of new compounds (L. Katz & Baltz, 2016). However, as few as 1% of environmental microorganisms can be cultured in the laboratory under standard cultivation conditions (Amann et al., 1990). Most BGCs from culturable microorganisms are cryptic or silent under laboratory conditions resulting in insufficient production of the encoded compound (M. Xu et al., 2016). These challenges resulted in high rates of rediscovery of known antibiotic compounds (Hover et al., 2018).

The challenges associated with natural product rediscovery spawned alternative approaches to find and express therapeutic compounds from microorganisms. Culture-independent techniques, i.e. metagenomics, are a promising approach to find new chemistry from uncultured microorganisms by sampling DNA directly from an environment, screening the DNA

for biosynthetic pathways predicted to express novel metabolites, and expressing those pathways in culturable, heterologous hosts (Handelsman, 2004; M. Katz et al., 2016; Pereira, 2019). The application of these metagenomic approaches have proven valuable in accessing new antimicrobial compounds such as turbomycin A and B (Gillespie et al., 2002), the cadasides (C. Wu et al., 2019), malacidins (Hover et al., 2018), and chloramphenicol derivatives (Nasrin et al., 2018). The potential of these approaches is clear, yet the discovery rate of new compounds from a metagenomic source remains low and there are no commercial successes to date (Baltz, 2019).

The ability to heterologously express BGCs is one of the major bottlenecks in metagenomic-based drug discovery and a significant reason for the lack of commercial successes (Chistoserdova, 2009). Heterologous expression of metagenome-derived BGCs often results in little to no metabolite yields due to transcriptional, translational, and biochemical incompatibilities with the host (D Cole Stevens, Hari, et al., 2013). *Escherichia coli* has been the most widely used host for metagenomic studies due to its high cloning efficiency and genetic tractability (Kakirde et al., 2010), but is a poor host for heterologous expression (Iqbal et al., 2016). *E. coli* hosts often lack the genetic and biochemical tools to express biosynthetic pathways encoding complex polyketides and nonribosomal peptides (Murli et al., 2003).

To improve the efficiency of heterologous expression, it is assumed that the best choice of host is one that is more closely related to the phylogenetic origin of the gene cluster since the host likely contains the genetic and molecular features needed for gene recognition and expression (Gadd & Sariaslani, 2016; J. J. Zhang et al., 2017). This approach has been successful in the scheme of expressing BGCs from cultured bacteria (Q. Liu et al., 2016; J. J. Zhang et al., 2017). However, it is highly impractical to match hosts based on phylogenetic

affiliation in functional metagenomic studies where the source may not be known for every BGC and/or may be highly diverse across an entire library.

An alternative approach would be to select a host that is naturally equipped to express and produce a wide range of natural products. *Streptomyces* spp. are excellent candidates to serve as heterologous hosts since these bacteria are known producers of many diverse natural products (Iqbal et al., 2016) and possess many precursors, cofactors, and enzymes needed for secondary metabolite synthesis (N. Lee et al., 2019). For this reason, several *Streptomyces* strains have been engineered for heterologous expression of BGCs including those derived from the species *Streptomyces coelicolor*, *Streptomyces lividans*, and *Streptomyces albus*. These strains have proven valuable in expressing gene clusters derived from cultured bacteria that are cryptic or silent in other bacteria (Rebets et al., 2017).

Most metagenomic studies to date have used engineered strains of *S. lividans* due to its fast growth rate and less active restriction-modification system (McMahon et al., 2012), but have been largely unsuccessful so far (Iqbal et al., 2016). Alternative *Streptomyces* hosts have yet to be rigorously studied for functional screening of metagenomic libraries. However, derivatives of *S. coelicolor*, such as *S. coelicolor* M1154, are an attractive option as this strain has been engineered to promote expression of secondary metabolite BGCs through the deletion of endogenous gene clusters and point mutations in *rpoB* and *rpsL* (Gomez-Escribano & Bibb, 2011). New technologies have also been developed to further promote expression in these *Streptomyces* hosts such as the inducible P_{nitA}-NitR (Herai et al., 2004) and P_{otr} (Weishan Wang, Yang, et al., 2016) systems. Again, these techniques have only been applied to expressing gene clusters from cultured bacteria. The usefulness of these technologies in metagenomic studies

remains unclear, but may expand our ability to express and uncover novel metabolites from metagenomes.

As part of an ongoing effort to discover new antimicrobials from a metagenomic source, our preliminary experiments identified 151 BGCs encoding diverse PKS and NRPS pathways from a large-insert metagenomic library (Santana-Pereira et al., 2020). In the present study, we tested the hypothesis that an engineered strain of *S. coelicolor* could be used as a heterologous host to express these metagenome-derived BGCs and produce antimicrobial metabolites. We predicted that by using *S. coelicolor* M1154 as a host and targeting only clusters containing PKS or NRPS pathways for expression, we would achieve a greater hit rate of antibiotic-producing clones compared to the low hit rates typically observed in functional metagenomic studies (Colin et al., 2015). Through a random functional screening approach, we were able to identify *S. coelicolor* clones that inhibited the growth of *A. baumannii* and *S. aureus* pathogens. We also aimed to improve the expression of antibiotic activity for *S. coelicolor* clones through the use of a dual-inducible expression system in the presence and absence of two inducers: 2.5 μM OTC and/or 0.1% (w/v) ƐCL. Furthermore, we initiated the process of isolating and characterizing the antimicrobial compounds produced by the bioactive *S. coelicolor* clones using bioactivity-guided fractionation and mass spectrometry.

3.3. Materials and Methods

3.3.1. Bacterial strains and plasmids.

Bacterial strains and plasmids used during this study are listed in Supplementary Table S3.1. The metagenomic library was constructed using the pSmartBAC-S vector (Nasrin et al., 2018) and BAC-Optimized Replicator v2.0 *Escherichia coli* DH10B (Lucigen Corporation, Middleton, WI). Functional screening was conducted using the heterologous host *Streptomyces*

coelicolor M1154 which was previously engineered for heterologous expression of BGCs by the removal of four endogenous gene clusters to alleviate precursor competition and mutations in *rpoB* and *rpsL* that pleiotropically upregulate antibiotic expression (Gomez-Escribano & Bibb, 2011). *E. coli* HB101 (Boyer & Roulland-Dussoix, 1969) bearing the helper plasmid pRK2013 (Figurski & Helinski, 1979) was used to facilitate the intergeneric triparental conjugation from *E. coli* to *S. coelicolor* M1154. Inducible expression studies were performed using the pDualP vector (Varigen Biosciences) that contains two inducible promoters flanking the cloning site for heterologous expression in *Streptomyces* (Stankey et al., 2022). Multidrug-resistant *Acinetobacter baumannii* WR3806 (Taitt et al., 2014), methicillin-sensitive *Staphylococcus aureus* Xen29 (Park et al., 2012), and/or antibiotic-susceptible *E. coli* WTZ17006 (Baba et al., 2006) were used as tester strains for the bioassay experiments.

3.3.2. Media and culture conditions.

E. coli strains were cultured in liquid Luria-Bertani (LB) medium or solid agar LB plates containing the appropriate antibiotics at 37°C with shaking at 180 rpm. Chloramphenicol (12.5 µg/ml), apramycin (50 µg/ml), kanamycin (30 µg/ml), and nalidixic acid (25 µg/ml) were supplemented in the medium when necessary. *S. coelicolor* M1154 was grown at 30°C on soya flour-mannitol (SFM) [2% (w/v) soybean flour, 2% (w/v) mannitol, and 2% (w/v) agar] for sporulation. *E. coli – Streptomyces* conjugations were performed on SFM agar medium containing 20 mM $MgCl_2$. *S. coelicolor* clones were cultured on SFM agar with 50 µg/ml apramycin for sporulation, and germination of spores was performed in 2xYT medium [1.6% (w/v) tryptone, 1% (w/v) yeast extract, 0.5% (w/v) NaCl]. Liquid TSB with 50 µg/ml apramycin or 12.5 µg/ml chloramphenicol (depending on the vector) and 3mm glass beads were used for *S. coelicolor* pre-cultures. Modified liquid yeast extract-malt extract (modified YEME) medium

[0.3% (w/v) yeast extract, 0.3% (w/v) malt extract, 0.5% (w/v) peptone, 1% (w/v) glucose] was used for heterologous expression and metabolite production. Strains used for bioassays including *A. baumannii* WR3806, *S. aureus* Xen29, and *E. coli* WTZ17006 were cultured in liquid or solid LB medium and incubated at 37°C with shaking at 180 rpm.

3.3.3. Metagenomic library construction, sequencing, and biosynthetic gene cluster analysis.

A metagenomic library constructed in *E. coli* DH10B from bulk soils collected from the Cullars Rotation (Auburn, AL, United States) was used for this study. Construction of the library, sequencing and assembly, and identification of BGC-containing clones used in the present study has been described in detail previously (Nasrin et al., 2018; Santana-Pereira et al., 2020).

3.3.4. Conjugation into *Streptomyces coelicolor* M1154 expression host.

Previously archived metagenomic clones containing a NRPS and/or PKS gene cluster (*n* = 151) identified in a previous publication (Santana-Pereira et al., 2020) were selected for conjugal transfer to the heterologous host *S. coelicolor* M1154. Intergeneric conjugation between *E. coli* and *S. coelicolor* was conducted as described by Huang et al. (2018) with modifications. Donor cells of individual *E. coli* DH10B clones were prepared by overnight growth at 37°C in 2 ml of LB supplemented with chloramphenicol (12.5 μg/ml). The helper strain *E. coli* HB101/pRK2013 was prepared by overnight growth at 37°C in 2 ml of LB supplemented with kanamycin (30 μg/ml). *E. coli* donors and helpers were diluted 1:100 in fresh LB with appropriate antibiotics and grown to an OD_{600} of 0.4. Cells were then washed twice in ice-cold LB and resuspended in 1 ml of ice-cold LB. Approximately 10^8 spores of *S. coelicolor* M1154 were added to 0.5 ml of 2xYT medium per conjugation, rinsed once in 2xYT medium, and

resuspended in 0.5 ml of 2xYT. Spores were heat-shocked at 59°C for 10 minutes to promote germination. After cooling to room temperature, 100 µl of *E. coli* DH10B clone and 100 µl of *E. coli* HB101/pRK2013 were added to the spore suspension. The mixture was centrifuged at 5,000 x *g* for 5 min, the supernatant was poured off, and the pellet was resuspended in the residual liquid. The mating mixture was plated on MS agar medium containing 20 mM $MgCl_2$ and incubated at 30°C for 24 hr. Plates were overlaid with 1 mL of sterile water containing 0.5 mg of nalidixic acid and 1 mg of chloramphenicol and incubated further at 30°C until exconjugants were visible. Single colonies of exconjugants were picked from the plates and isolated on SFM agar containing 30 µg/ml nalidixic acid and 12.5 µg/ml chloramphenicol. *S. coelicolor* exconjugants were then plated on 12.5 µg/ml chloramphenicol plates and fresh mycelia was picked from the plates and stored in 25% glycerol at -80°C.

3.3.5. Functional screening of metagenomic clones.

Heterologous expression of metagenome-derived BGCs by *S. coelicolor* clones was evaluated by high-throughput 96-well microplate assays. To prepare pre-cultures, streaks of individual *S. coelicolor* clones (including the *S. coelicolor* empty vector control) were inoculated into 5 ml of TSB containing 12.5 µg/ml chloramphenicol and 0.3 g of glass beads (3 mm) and incubated at 30°C, 200 rpm for 3 – 4 days. Pre-cultures were pelleted by centrifugation at 7,700 x *g* for 10 minutes, media was discarded, and cells were washed twice in TSB to remove antibiotic traces. The washed cells were resuspended in 5 ml of modified YEME medium and 100 µl of each seed culture was transferred into 10 ml of modified YEME medium. The cultures were shaken for 7 days at 30°C, 200 rpm. Each *S. coelicolor* clone culture was then harvested by centrifugation at 7,700 x *g* for 10 minutes and the resulting supernatants were filtered through an 0.2-µm filter (VWR).

For microplate assays, bacterial test strains (*A. baumannii* WR3806 or *S. aureus* Xen29) were prepared by inoculating a single colony in 2 ml of LB and shaken overnight at 37°C, 200 rpm. Overnight cultures were diluted 1:100 in fresh LB, grown to an optical density at 600 nm (OD_{600}) of ~ 0.5 – 0.6, and diluted 1:100 in LB. A volume of 100 µl of cell-free supernatants from each *S. coelicolor* clone was added to triplicate wells of sterile 96-well microplates. 100 µl of the diluted log-phase bacterial test strains were added and mixed with the supernatants. Control wells were included with 100 µl of the bacterial test strain mixed with 100 µl of modified YEME medium. The OD_{600} of each well was determined in a microplate reader before incubation (T_0) and following incubation at 37°C for 16 hours with shaking at 200 rpm (T_{16}). Growth inhibition, expressed as a percent inhibition of the test strain wells relative to control wells, was determined using the formula:

$$\text{Inhibition (\%)} = \left(1 - \left(\frac{OD_{600}\text{ test well at } T_{16} - OD_{600}\text{ test well at } T_0}{OD_{600}\text{ control well at } T_{16} - OD_{600}\text{ control well at } T_0}\right)\right) \times 100$$

Percent inhibitions for each triplicate well were averaged per *S. coelicolor* clone and plotted relative to the percent inhibition of the *S. coelicolor* empty vector control.

3.3.6. Subcloning in pDualP and inducible expression studies.

The targeted capture and subcloning of metagenomic BACs by CRISPR-Cas9 restriction and Gibson assembly in the dual-inducible promoter pDualP vector was performed as described by (Stankey et al., 2022). Subcloned BGCs in the pDualP system were reintroduced from recombinant *E. coli* clones into *S. coelicolor* M1154 by the triparental conjugation method detailed above. Exconjugant *S. coelicolor* clones containing the integrated pDualP vector were confirmed by colony PCR using the primers UniversalBACF and UniversalBACR

(Supplementary Table S3.2). Spore stocks of exconjugants were cryopreserved at -20°C in 25% glycerol.

Pre-cultures of *S. coelicolor* pDualP clones and *S. coelicolor* pSmartBAC-S clones were prepared by inoculating spores from glycerol stocks into 5 ml of TSB containing 50 µg/ml apramycin and 0.3 g of 3mm glass beads. Cultures were incubated at 30°C, 180 rpm for 3 – 4 days. Cells were pelleted by centrifugation at 7, 700 x *g* for 10 min, washed twice in TSB, and resuspended in 5 ml of modified YEME medium. A 1% inoculum (100 µl) of washed pre-cultures was added to 10 ml of modified YEME. A total of 15 cultures per *S. coelicolor* clone were prepared (5 conditions to be tested X 3 replicates each). For induction of pDualP promoters, cultures were supplemented with one of the following treatments prior to incubation: 1) no inducers, 2) 0.1% (w/v) Ɛ-caprolactam (ƐCL) inducer, 3) 2.5 µM oxytetracycline dihydrate (OTC) inducer, or 4) 0.1% ƐCL and 2.5 µM OTC inducers. *S. coelicolor* pSmartBAC-S clones were not treated with inducers and served as the control to compare expression from non-inducible native promoters. All cultures were incubated at 30°C for 7 days and shaking at 180 rpm. Cell-free supernatants were harvested from the cultures and assayed for bioactivity against *A. baumannii* WR3806 following the same procedure described above. Wells containing YEME broth with each treatment were mixed with pathogen and included as controls to account for any inhibitory effects of the inducers.

Statistical analyses of the bioactivity data were performed in R (R_Core_Team 2019) using the following packages: ‘lme4’ (Bates et al., 2014), ‘lmerTest’ (Kuznetsova et al., 2017b), ‘emmeans’ (Lenth et al., 2018), ‘betareg’ (Cribari-Neto & Zeileis, 2010). Data wrangling was performed in R with the help of the following packages: ‘tidyverse’ (Wickham, 2017b), ‘dplyr’ (Wickham et al., 2015), ‘reshape2’ (Wickham, 2007b) and a bar graph was produced using the

following packages: 'ggpubr' (Kassambara, 2020), 'cowplot' (Wilke, 2019), and 'RColorBrewer' (Erich, 2014). The effect of each treatment on the percent inhibition of pathogen exhibited by each clone was assessed using a beta regression model (to fit a model with a response distribution between 0 and 1). Pairwise comparisons were analyzed between the percent inhibition associated with each treatment for each clone. Pairwise comparisons were corrected by using the default "Tukey method" in the 'emmeans' package to correct for Type 1 error. Statistically significant differences between treatments were established at $\alpha = 0.05$.

3.3.7. Small-scale cultivation and extraction of secondary metabolites with resins.

Pre-cultures of selected *S. coelicolor* clones were prepared by inoculation of spores from glycerol stocks in 5 ml of TSB containing 50 µg/ml apramycin and 0.3 g of 3mm glass beads and incubation at 30°C for 3 – 4 days. Pre-cultures were pelleted by centrifugation at 7,700 x *g* for 10 minutes, media was discarded, and cells were washed twice in TSB to remove antibiotic traces. The washed cells were resuspended in 5 ml of modified YEME medium.

Four polymeric resins were prepared to be used for extractions: Amberlite® XAD-4, XAD-7, XAD-16 (Sigma-Aldrich), and Diaion® HP-20 (Sigma-Aldrich). Resins were prepared and activated by washing in twice their volume of methanol, stirring for 1 hr at room temperature, removal of the methanol by filtration, and washing 6X in DI water before being vacuum filtered. To prepare production media, 0.5 g of resins were added to 250 ml flasks containing 50 ml of modified YEME medium and autoclaved. 500 µl of washed pre-cultures were transferred into each flask (two flasks per resin type). The appropriate inducer (2.5 µM OTC and/or 0.1% (w/v) ƐCL) for each clone as determined in the inducible expression studies was added prior to incubation. Cultures were incubated for 7 days at 30°C and shaking at 180 rpm. After incubation, cultures containing resin and cell biomass were combined for each clone

and resin type. Supernatants were separated from the resin/biomass by centrifugation at 7,000 x *g* for 15 min. To extract metabolites from the resins, 10 ml of 100% methanol was added to the resin/biomass mixture from each clone, vigorously vortexed for 10 min, and shaken for 1 hr at room temperature. The mixture was filtered through 0.22 µM PTFE filters and the eluted extract was collected and dried. Crude extracts were then dissolved in methanol to 10 mg/ml and evaluated for antibacterial activity against *A. baumannii* WR3806 by disc diffusion assays.

3.3.8. Disc diffusion assay.

Bioactivity testing of crude extracts and fractions obtained from the clones was performed by the standard Kirby-Bauer disc diffusion method (A. Bauer, 1966). Agar plates were prepared with ~ 20 ml of sterile Mueller-Hinton agar. Agar plates were then swabbed with a standardized inoculum (0.5 McFarland, ~10^8 CFU/ml) of *A. baumannii* WR3806 or *E. coli* WTZ17006. Filter paper discs (6 mm) were impregnated with 20 µl total (final concentration: 10 mg/ml) of the test extract or, after fractionation steps, test fractions and placed on the agar surface. The agar plates with loaded discs were left for 30 minutes at room temperature for compound diffusion. Gentamicin (10 µg) antibiotic discs (BD BBL Sensi-Disc; Becton, Dickinson and Company, Sparks, MD) was used as the antibiotic positive control when possible, while filter paper discs loaded with 80% methanol were used as the negative control for extract testing. For assays of fractions, the same solvent mixture used to dissolve the fractions was used for the negative control. The agar plates were incubated at 37°C overnight and the diameters of inhibition growth zones were measured.

3.3.9. Sequence analysis of bioactive clones.

Insert sequences from bioactive clones were sequenced and assembled as previously described (Santana-Pereira et al., 2020). BGCs encoded in the insert sequences were initially

characterized by antiSMASH 5.0 (Blin et al., 2019) and manually inspected by importing the antiSMASH output into the Geneious software suite (Kearse et al., 2012). For specific clones of interest, annotations of predicted open reading frames (ORFs) and genes/domains from candidate BGCs as predicted by antiSMASH version 6.1.0 (Blin et al., 2021) were visualized using the R package "gggenes" (Wilkins & Kurtz, 2019) and "ggplot2" (Wickham, 2016). The function of putative gene products were predicted by comparing the amino acid sequences from each detected ORF against the GenBank database using BLASTp.

GC% content of each clone in this study was collected by Geneious software. To analyze the relationship between the clone GC% content, the host GC% content, and the likelihood of expression, a generalized linear mixed effects model was fit to the data. These statistical analyses were performed in R using the following packages: 'lme4' (Bates et al., 2014), 'blmeco' (Korner-Nievergelt et al., 2015), and 'lmerTest' (Kuznetsova et al., 2017a). Data wrangling was performed in R with the help of the follow packages: 'tidyverse' (Wickham, 2017a), 'dplyr' (Wickham et al., 2015), 'reshape2' (Wickham, 2007a), 'plotrix' (Lemon, 2006), and 'HH' (Heiberger et al., 2015).

The absolute difference between the GC% content of each gene cluster and the GC% content of the host was calculated. This data was used as a fixed effect in the model to analyze its relationship with expression. Expression was included as the dependent variable, coded as a binomial outcome with two possible outcomes: "yes" or "no", corresponding to whether or not there was successful expression for a given BGC. Therefore, this model incorporated an inverse link function for binomial data. This model included a random effect of host identity [either *S. coelicolor* M1154 or *E. coli* BTRA (from previously unpublished data)], to account for nonindependence of data collected from each individual host (i.e., accounts for the influence of

differences between the hosts on the individual data points). Due to the use of the binomial family link function, the model estimate was exponentiated to convert it from log odds to the more approachable odds ratio. Figures were produced using the following packages: 'ggpubr' (Kassambara, 2018), 'cowplot' (Wilke, 2016), and 'RColorBrewer' (Neuwirth, 2014).

3.3.10. Large-scale cultivation of *S. coelicolor* clone P17B06 and metabolite extraction.

The seed cultures of *S. coelicolor* P17B06 (50 ml) and *S. coelicolor* with an empty vector were prepared following the same inoculation procedures as the small-scale cultivation. A 1% inoculum of the mature seed culture was used to inoculate four 2.0-L flasks containing 500 ml of modified YEME broth with and without 2.5 µM OTC and 5 g of activated, sterile XAD-16 resins (1% w/v) and plugged with foam stoppers. Cultures were incubated for 7 days at 30°C and shaking at 180 rpm, after which the resin and cell biomass were collected by vacuum filtration using a Buchner funnel fitted with a Whatman GF/F glass fiber filter. The resin/biomass mixture was eluted 2X with 250 ml of 80% MeOH, vigorously vortexed for 10 min, and rocked for 1 hr on a shaker. The collected filtrate was combined in a round bottom flask and partially dried by rotary evaporation under reduced pressure at 50°C. A small sample (1 ml) was pulled from the extract and dried by SpeedVac to be used for bioactivity assays. Approximately 25 g of activated Diaion HP-20 resin was added to the round bottom flask containing the remaining organic extract and the mixture was dried fully. The extract sample was tested for bioactivity against the tester strains *A. baumannii* WR3806 and *E. coli* WTZ17006 using the disc diffusion assay described above.

3.3.11. Fractionation procedure and isolation of compounds.

The combined crude extract and HP-20 resin mixture from both clone P17B06 and the empty vector were fractionated by loading the sample into an open glass column (22 mm i.d. x 350 mm) and elution with a step gradient of MeOH and H_2O (25:75 – 100:0) to collect four parent fractions (FR1~FR4). Fractions were concentrated by rotary evaporation, resuspended to a concentration of 10 mg/ml, and assayed for growth inhibition of *A. baumannii* WR3806 and *E. coli* WTZ17006 using the disc diffusion method described above. The active fraction FR2 was further fractionated by two approaches – solid phase extraction (SPE) and size-exclusion chromatography. The first approach used Strata® C18-E solid phase extraction (SPE) cartridges (2 g/12 ml, 55 µm, 70 Å). The SPE column was conditioned with 4 column volumes (CV) of 100% acetonitrile (ACN) then equilibrated with 4 CV of HPLC-grade H_2O. FR2 was dissolved in 1 ml of 10% ACN prior to loading on the column. The column was eluted stepwise with 4 CV of solvent mixture from 10% H_2O-ACN to 100% ACN to collect 10 fractions labeled as FR2-1 through FR2-10. Finally, all 10 fractions were concentrated, screened for activity by the disc diffusion method described earlier, and analyzed by LC-MS/MS.

The active fraction FR2 was also fractionated using a LH-20 size exclusion column (24 cm x 1.5 cm) that was hydrated in 1:1 MeOH:H_2O. FR2 was suspended in 0.5 mL of 1:1 MeOH:H_2O and loaded onto the column and eluted with methanol to give nine fractions (FR2-1S to FR2-9S). Fractions were concentrated to dryness under centrifugal evaporation (Speedvac) and screened for activity by the disc diffusion assay.

3.3.12. UPLC and LC-MS analyses.

UPLC analysis of extracts and fractions from clone P17B06 and the empty vector control was performed on a Waters ACQUITY BEH C18 column (2.1 x 100mm, I.D. 1.7 mm; Waters Co., Milford, CT, USA) and gradient elution (flow rate 0.45 mL/min; mobile phase A: water (0.1

% formic acid); mobile phase B: ACN (0.1 % formic acid); hold at 5 % B for 0.5 minute, 5-100 % B over 8 min., 65-100 % B over 1 min., hold at 100 % B for 1.5 min., re-equilibrate at 5 % B for 1.5 minutes). Mass spectrometry data were acquired on a Waters G2-XS QToF mass spectrometer operated in electrospray positive mode (ESI^+) and negative mode (ESI^-). The capillary voltage set at 2.5kV, source and desolvation temperatures at 100°C and 550°C, respectively, and desolvation gas flow at 800 L/h. Simultaneous acquisition of a leucine encephalin (*m/z* 556.2771) as a lockmass solution was acquired but not applied until post-processing. Data was collected in continuum mode with the MSe function active (mass range 50-1800 Da; scan time 0.1 scans/sec; function 1 CE off; function 2 CE ramp 45-50 V; lockmass data acquired but not applied) for further analysis including spectral alignment, lockmass calibration, and peak picking. Alignment was performed on QC samples containing aliquots of all samples and vendor-recommended defaults were applied for peak picking. Analysis of LC-MS data was performed using Progenesis QI software (Nonlinear Dynamics, UK) for alignment, peak picking, and deconvolution. Alignment was performed on pooled QC samples. Run times were restricted to the 0.5–8.5 min retention time window. Vendor specified - default parameters for peak picking (automatic thresholds, minimum peak width = 3 s) and alignment were applied. Specifically, LC-MS-QToF data from bioactive clone extracts and fractions was compared visually and statistically to samples from the empty vector control to identify differentially expressed compounds. Molecular masses corresponding to peaks of interest identified by LC-MS-QToF analysis were submitted to the Global Natural Product Social networking (GNPS) (https://gnps.ucsd.edu) (M. Wang, Carver, et al., 2016) platform to compare against curated spectral libraries.

3.3.13. Data availability.

Raw sequencing data is available via NCBI GenBank under the NCBI BioProject accession number PRJNA669376.

3.4. Results

3.4.1. Construction of *S. coelicolor* metagenomic clones.

A metagenomic library comprised of 19,200 clones was previously constructed from topsoil collected from the Auburn University Cullars Rotation agricultural plot (Auburn, AL, United States) (Nasrin et al., 2018). The library was constructed in *E. coli* DH10B using the pSMARTBAC-S vector, and in a previous study, we screened the metagenomic clones for BGCs encoding PKSs or NRPSs using PCR or a next-generation sequencing (NGS) multiplexed pooling strategy (Santana-Pereira et al., 2020). We identified 151 metagenomic clones containing a PKS and/or NRPS gene cluster and these were selected for conjugal transfer to the heterologous host *S. coelicolor* M1154. We selected *S. coelicolor* M1154 as the expression host for this study because it has previously shown improved heterologous expression capabilities, has been used to produce a range of compound classes with higher metabolite yields, and has been engineered with a cleaner metabolic background (Gomez-Escribano & Bibb, 2011; Myronovskyi & Luzhetskyy, 2019).

Of the 151 metagenomic clones, a total of 90 clones were successfully conjugated from *E. coli* to *S. coelicolor* M1154 via tri-parental mating and stored as mycelial glycerol stocks. Two unsuccessful attempts were made to transfer the remaining clones to *S. coelicolor* M1154, suggesting the BGCs were incompatible with the *Streptomyces* host potentially from toxicity of the expressed products. The 90 *S. coelicolor* clones were subsequently targeted for functional screening (Supplementary Table S3.3).

3.4.2. Heterologous expression of metagenome-derived BGCs by *S. coelicolor*.

To screen *S. coelicolor* clones for heterologous expression of metagenome-derived PKS and/or NRPS clusters, supernatants from 7-day old cultures grown in modified YEME broth were evaluated for inhibitory activity against the Gram-positive bacterial strain *S. aureus* Xen29 and the Gram-negative MDR strain *A. baumannii* WR3806. Of the 90 *S. coelicolor* clones screened, 8 clones showed inhibitory activity against *S. aureus* Xen29 and 8 were active against *A. baumannii* WR3806 (Supplementary Table S3.3). Two clones were found to inhibit the growth of both tester strains. In total, primary screening identified 18 bioactive clones from a collection of 90 clones, which reflects a hit rate of 20%. The gene cluster types predicted by antiSMASH analysis revealed that three putative hybrid PKS-NRPS clusters, five PKS clusters, and 10 NRPS clusters from the library were expressed by *S. coelicolor* M1154.

We then sought to evaluate if GC% content similarities between the host and the BGC had an effect on the likelihood of expression using a generalized linear mixed effects model. In our analyses, we included the present heterologous expression data from *S. coelicolor* M1154 as a host with a higher GC% content (~72%), along with previously unpublished data from a different host with a lower GC% content (~50%) - *E. coli* BTRA. This unpublished data has also been listed in Supplementary Table S3.3. We detected no significant effect of GC% content difference between the host and the BGC on the likelihood of expression for a given clone (Fig. 3.1). There was an estimated 1.97% decrease in likelihood of expression for every 1 unit decrease in absolute GC% content difference between the BGC and the host, but this effect was not significant (β = -1.97; Z = -0.693; p = 0.488). We checked for overdispersion of this model and found it to be 0.93, indicating no effect of overdispersion according to the “blmeco” package recommendations.

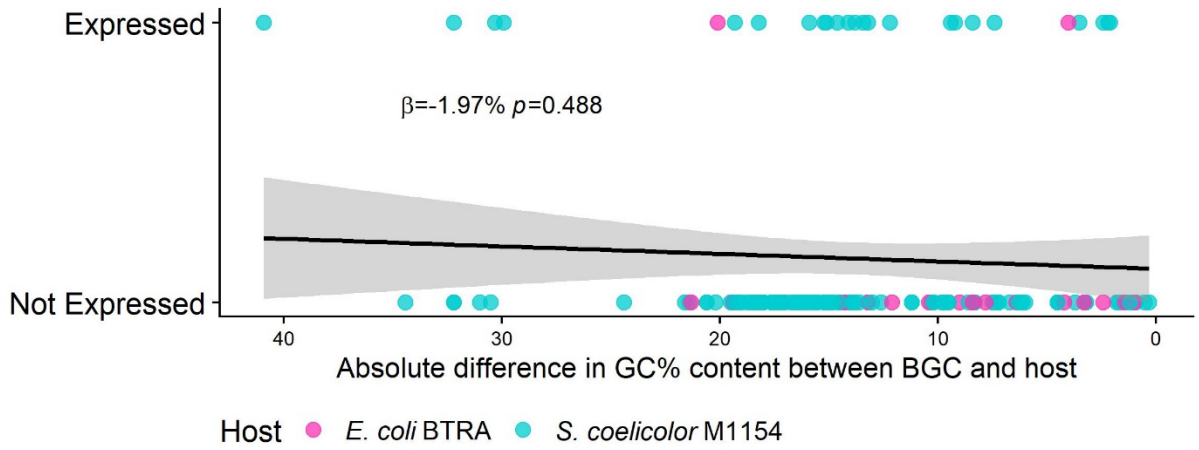

Figure 3.1. A scatterplot showing the change in likelihood of expression versus non-expression as the difference between the GC% content of the gene cluster and GC% content of the host decreases. The model estimate and p-value from the generalized linear mixed effects model is shown inlaid. The model estimate (β) is the odds ratio, and thus represents the percent change in the likelihood of expression per 1 unit decrease in GC% content difference between the BGC and the host. Each colored dot represents a single BGC, and these are color-coded by the host in which they were screened for expression.

3.4.3. Effect of the pDualP expression system on antibiotic activity of *S. coelicolor* clones.

In an attempt to optimize heterologous expression of metagenomic BGCs by *S. coelicolor*, BGCs from clones with activity against *A. baumannii* (*n*=8) were captured and subcloned in the dual-inducible promoter pDualP vector using CRISPR-Cas9 restriction and Gibson assembly as described by Stankey et al. (2022). This vector uses two inducible promotor elements, P_{otr} and P_{nitA} that flank each side of the cloning site and can be induced by OTC and ƐCL, respectively. We focused on the clones expressing activity against *A. baumannii* since the expressed metabolites were able to overcome the intrinsic antibiotic resistance mechanisms of this pathogen, and may therefore, be better candidates for new therapeutic drugs. Subcloned BGCs and an empty pDualP vector were reintroduced into *S. coelicolor* M1154 by triparental

mating and exconjugants were confirmed by colony PCR. Exconjugants from one clone could not be validated and was subsequently removed from downstream analyses.

We evaluated the pDualP inducible expression system for each metagenomic BGC (*n*=7) by culturing each clone in modified YEME for 7 days in the presence and absence of the inducers 2.5 µM OTC and/or 0.1% (w/v) ƐCL. Cell-free supernatants were then screened for antimicrobial activity against *A. baumannii* WR3806 using a 96-well microplate assay. In the same experiment, BGCs cloned in the non-inducible pSmartBAC-S vector were grown under the same conditions but without inducers to compare pDualP expression with the expression from native promoters in *S. coelicolor* M1154. Inducible expression of metagenomic gene clusters was demonstrated for four of the seven clones when cloned in pDualP and induced with one or both inducers, in comparison to the expression from native promoters (Fig. 3.2a). There was no difference in the expression of antibacterial activity across all treatments for clones P17F15, P19L22, and P50M18. No significant activity was observed from the inducers OTC and ƐCL against the target pathogen *A. baumannii* WR3806 (data not shown).

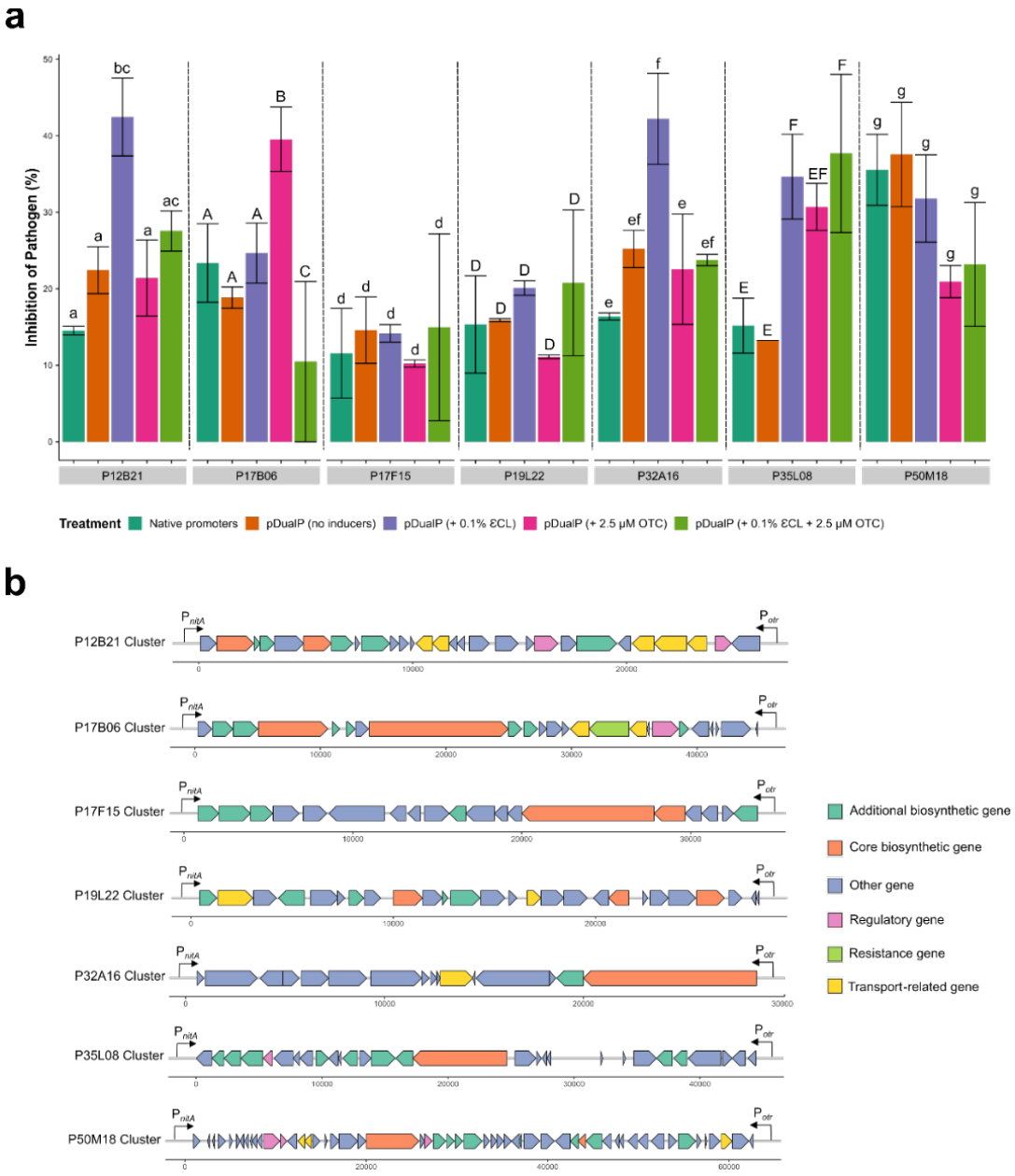

Figure 3.2. Inducible expression analysis of metagenomic biosynthetic gene clusters by *S. coelicolor* M1154 and schematic representations of each gene cluster. (a) The mean percent inhibition of *A. baumannii* WR3806 by supernatants of seven metagenomic BGCs cloned in *S. coelicolor* M1154 with the pDualP inducible expression system or native promoters and treated with or without 0.1% (w/v) ƐCL and 2.5 µM OTC. Percent inhibition of pathogen is calculated with respect to *S. coelicolor* M1154 cloned with an empty vector and values represent the means of technical replicates. Black lines indicate the standard error among technical replicates. Letters represent significant differences ($p < 0.05$) between the percent inhibitions of the pathogen associated with each treatment for each clone as determined by a beta regression model with Tukey's post-hoc correction for pairwise comparisons. (b) antiSMASH annotations of the biosynthetic gene clusters from each metagenomic clone. The location of the inducible promoters P_{nitA} and P_{otr} with respect to each gene cluster is shown.

For the four BGCs with inducible expression, the inducer treatment that led to the greatest antibacterial activity exhibited by each clone was variable. Significantly higher inhibition of *A. baumannii* (> 40% mean inhibition) was observed from clone P12B21 when induced with 0.1% ɛCL compared to the activity from the OTC inducer treatment ($p = 0.018$) and the uninduced treatments by both pDualP ($p = 0.035$) and native promoters ($p < 0.001$). Mean percent inhibition from clone P35L08 was greatest from induction with both ɛCL and OTC, and was significantly higher than the expression from native promoters ($p = 0.005$) but not by OTC alone ($p = 0.09$). For clone P32A16, the mean percent inhibition was 42% when induced with ɛCL, which was significantly higher than the activity from expression by native promoters ($p = 0.01$). Only clone P17B06 showed the greatest mean percent inhibition (~39%) from the addition of the OTC inducer alone and this was statistically significant from expression by native promoters ($p = 0.04$). The addition of both inducers to clone P17B06 significantly decreased the expression of antibacterial activity compared to the native promoters ($p < 0.001$).

To determine if the differences in expression were linked to gene orientation and location relative to the inducible promoters, we manually inspected the antiSMASH annotation of each metagenomic BGC sequence. The annotation of each gene cluster is depicted in Figure 3.2b. Clone P12B21 demonstrated increased expression from induction with ɛCL and many of the predicted biosynthetic genes including two putative genes encoding regulatory elements were in the same orientation as the P_{nitA} promoter. Expression by clone P32A16 was also greater when induced with ɛCL; however, the core biosynthetic genes were in the opposite direction of P_{nitA} and were located closer to the P_{otr} promoter. For clone P17B06, induction with OTC increased expression of antibacterial activity although many of the predicted biosynthetic genes were oriented in the opposite direction of P_{otr}. Moreover, a predicted regulatory element in the cluster

of P17B06 was in the same direction as P_{nitA}. Clone P35L08 showed the greatest activity when induced with ƐCL or both ƐCL/OTC and many of the biosynthetic genes were located closer to P_{nitA}. However, the core biosynthetic gene and regulatory gene were in the same direction as P_{otr}. Thus, no clear pattern in biosynthetic gene orientation and location with respect to the inducible promoters was observed for most of the metagenomic BGCs.

3.4.4. Extraction of antimicrobial metabolites using polymeric resins.

Polymeric resins added during fermentation have been shown to enhance the production of various secondary metabolites from actinomycetes (González-Menéndez et al., 2014). We evaluated four polymeric resins including Amberlite® XAD-4, Amberlite® XAD-7, Amberlite® XAD-16, and Diaion® HP-20 that differ in polymer chemistry, particle and pore sizes, and surface area for their ability to bind antimicrobial metabolites from the metagenomic clones. Resins were added at the inoculation stage at a concentration of 1% (w/v) to small-scale cultivations of the four metagenomic clones with significant inducible expression of activity – P12B21, P17B06, P32A16, and P35L08 (Fig. 3.2a). Desorption of metabolites from the resins was performed using 100% methanol and dried extracts ranged in weights from 5 – 33 mg and were yellow to golden brown in color for all clones.

Disc diffusion assays of the resin-based crude extracts identified active extracts against the target pathogen *A. baumannii* WR3806 from two metagenomic clones, P12B21 and P17B06 (Table 3.1). Activity was not observed from any of the crude extracts tested for clones P32A16 and P35L08. In the same assay, cell-free supernatants were also screened for activity to monitor expression (not shown in Table 3.1) and the results showed that the supernatants from each of the four metagenomic clones were active against *A. baumannii*. This finding suggested that clones P32A16 and P35L08 were still capable of producing antimicrobial metabolites under the

growth conditions used for small-scale cultivation; however, these metabolites may have low affinity for the resins or were not eluted from the resins when 100% methanol was used as the solvent.

For clone P12B21, extracts from XAD-4, XAD-16, and HP-20 resins all inhibited the growth of *A. baumannii*. Clone P12B21 extracts from XAD-16 and HP-20 resins were slightly more active against the target pathogen compared to extracts from XAD-4 resins. Antimicrobial metabolites were also extracted from cultures of clone P17B06 using XAD-16 and HP-20 resins. Extracts from XAD-7 resins were not active from any of the metagenomic clones screened. To screen for any potential antimicrobial activity from the *S. coelicolor* host, small-scale cultivations were also prepared using the *S. coelicolor* empty vector clone and extractions were prepared under the same conditions as the bioactive clones. No activity was observed from any of the extracts derived from the empty vector clone. Activity of crude extracts using HP-20 resins from each metagenomic clone is depicted in Supplementary Figure S3.1.

Of the two clones with bioactive extracts, P12B21 and P17B06, we selected clone P17B06 for further bioinformatic and chemical analysis. This selection was based on the following: 1) we hypothesized that the larger zones of inhibition observed from P17B06 extracts could mean that the antimicrobial metabolite(s) were more abundant and, as a result, this could improve dereplication and purification of the target metabolites, and 2) preliminary sequence analysis of clone P17B06 indicated that this clone harbored a unique NRPS cluster which may suggest that a novel antimicrobial compound was produced.

Table 3.1. Antibacterial activities of metagenomic clone extracts from various polymeric resins.

	Zones of growth inhibition (mm) against *A. baumannii* WR3806			
Clone	**XAD-4 extract**	**XAD-7 extract**	**XAD-16 extract**	**HP-20 extract**
P12B21	10	-	13	12
P17B06	-	-	14	15
P32A16	-	-	-	-
P35L08	-	-	-	-
Empty vector	-	-	-	-

3.4.5. Gene cluster analysis of metagenomic clone P17B06.

To identify the biosynthetic genes responsible for antimicrobial metabolite production by clone P17B06 and to predict the chemical class of the produced metabolites, insert sequences obtained by NGS and *de novo* assembly were annotated by antiSMASH v6.0.1 and BLASTp search comparisons against the GenBank database. Clone P17B06 was predicted to encode a 44,974 bp cluster consisting of 24 ORFs including core biosynthetic genes linked to an NRPS, regulatory genes, and a resistance gene (Fig. 3.3; Supplementary Table S3.4). The NRPS cluster showed limited homology (40% of gene similarity) to a puwainaphycin cluster (MIBiG accession: BGC 0001125) that encodes lipophilic cyclic decapeptides produced by *Anabaena* sp. with cardioactive functions (Gregson et al., 1992). Similarity with the puwainaphycins cluster and other clusters in the MIBiG database were found to be mainly within the core NRPS region (Supplementary Fig. S3.2).

The core NRPS region comprised five ORFs encoding five NRPS modules (M1 to M5). Annotation of the NRPS found that four of the modules (M2 to M5) were complete with the minimum core domains – a condensation (C) domain, an adenosine monophosphate-binding

(AMP) domain, and a peptidyl-carrier protein (PCP) domain. Module M1 was incomplete with only an AMP and PCP domain annotated. The predicted amino acid monomers for modules M3 to M5 were val/leu/ser (Fig. 3.3). Amino acid monomers could not be predicted for modules M1 and M2 since these domains were predicted to be inactive. Together these NRPS modules are predicted to catalyze the formation of a pentapeptide. However, the cluster was not predicted to encode a thioesterase (TE) domain that are typically present at the terminal end of the cluster and responsible for releasing the mature peptide from the NRPS machinery (Kohli et al., 2002). Thus, it is unknown at this time how the produced pentapeptide is released.

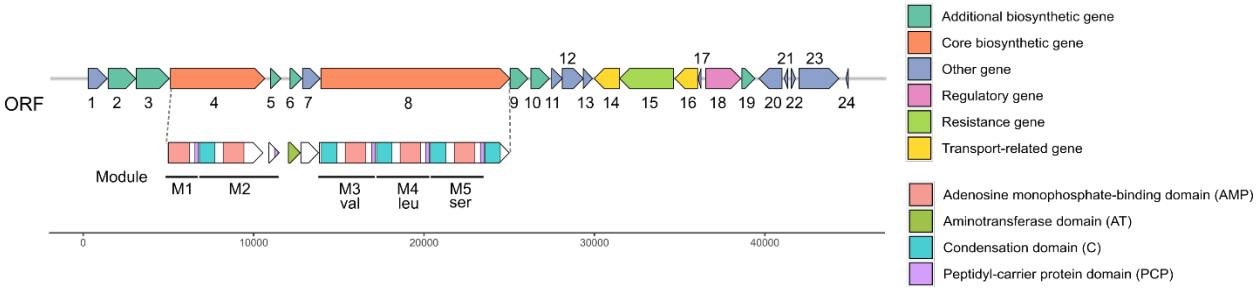

Figure 3.3. Organization of the predicted biosynthetic gene cluster of metagenomic clone P17B06. Colors represent putative broad gene functions as shown in the legend. Within the core biosynthetic region, different biosynthetic domains are represented by a colored block following the domain designation in the legend. The predicted NRPS module and any predicted amino acid monomers for a given module are also depicted. A detailed description of the predicted gene function is given in Supplementary Table S3.4.

Interestingly, the metagenome cluster was also predicted to encode two genes that comprise the characteristic biosynthetic cassette for capreomycidine. Capreomycidine is an amino acid found in the tuberactinomycin class of peptide antibiotics including viomycin and capreomycin. These antibiotics are active against mycobacteria, Gram-positive, and Gram-negative bacteria (Toyohara et al., 1969). Specifically, ORF7 showed homology to an alpha ketoglutarate-dependent dioxygenase that can convert (2*S*)-arginine to (3*S*)-hydroxy-(2*S*)-arginine. This product may be converted to (2*S*,3*R*)-capreomycidine by a capreomycidine synthase encoded in ORF6. Asparagine

was another unusual monomer predicted for the NRPS cluster as ORF3 showed homology to an asparagine synthase (Supplementary Table S3.2).

Near the right border of the P17B06 cluster, a putative regulatory gene with homology to a σ^{54} transcriptional regulator was identified. σ^{54} is capable of positively regulating heterologous production in *E. coli* and has been shown to regulate natural product genes in *Myxococcous xanthus* (Ma et al., 2021); however, σ^{54} promoters are non-functional in *Streptomyces* (David Cole Stevens, Conway, et al., 2013) and it is unlikely that this regulator would be functional in *S. coelicolor*. No other regulatory genes were predicted within this cluster, which may suggest that the regulation of this BGC is controlled by a pleiotropic regulator located outside of the cluster. A resistance gene (ORF15) with homology to a multidrug-efflux transporter was also identified within the cluster that may function to provide self-resistance to the produced antibiotic.

The source of each putative gene product from BLAST analysis was analyzed to predict the phylogenetic origin of the metagenome cluster. Many of the NRPS-associated ORFs had closest GenBank matches to proteins from Cyanobacteria taxa and uncultured bacteria (Supplementary Table S3.2). The % identity of these genes ranged from 46 – 66%. Genes that flanked the NRPS region had a relatively higher % identity (range of 61 – 83%) to gene products from the phylum Acidobacteria. Specifically, the highest % identities were found to homologs of the class *Blastocatellia*. Plotting of the G+C% content across the insert sequence provided support for a *Blastocatellia* origin of the metagenome cluster as the median (~55% GC; Supplementary Fig. S3.3) was within the range observed for other *Blastocatellia* genomes deposited in GenBank (50 – 60%). Three ORFs showed no significant homology to any sequences in the GenBank database.

3.4.6. Bioactivity-guided fractionation of antimicrobial compounds.

A bioactivity-guided fractionation approach was followed in an attempt to isolate the antimicrobial compounds from clone P17B06 (Fig. 3.4). Crude extracts were prepared from upscaled cultures of *S. coelicolor* clone P17B06 and the *S. coelicolor* empty vector clone cultured in the presence and absence of the inducer 2.5 µM OTC. The antibacterial activity of crude extracts (10 mg/ml) were evaluated against the target MDR pathogen *A. baumannii* WR3806 and an antibiotic-susceptible strain of *E. coli* WTZ17006. *E. coli* WTZ17006 was included as a tester strain since previous work has shown that this strain is susceptible to antimicrobial peptides and other compounds due to mutations in *degP* and *tolC* (Sulavik et al., 2001; Ulvatne et al., 2002). Thus, this strain may be more sensitive to compounds that are in lower abundance, which is a common occurrence in heterologous expression studies (Herai et al., 2004).

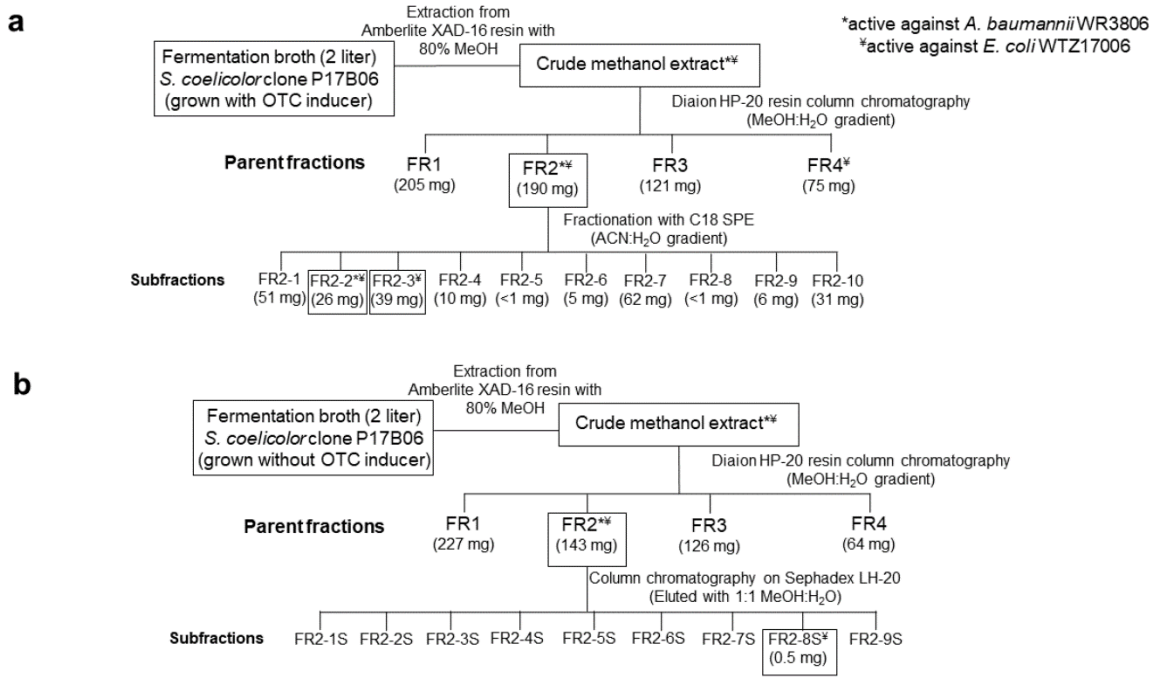

Figure 3.4. Flowchart of the extraction and fractionation of bioactive compounds produced by *S. coelicolor* clone P17B06. Two separate approaches were followed from fermentations of (a) *S. coelicolor* clone P17B06 cultured with OTC inducer, and (b) *S. coelicolor* clone P17B06 cultured without OTC are depicted.

The results showed that the crude extracts derived from clone P17B06 cultured with OTC inducer exhibited 12 mm zone of inhibition (ZOI) against *A. baumannii* and 20 mm ZOI against *E. coli* as shown in Fig. 3.5. Despite our results showing an increase in antibacterial activity from supernatants when clone P17B06 was grown with OTC, there was no measurable difference in the ZOI from crude extracts of clone P17B06 cultured with and without OTC. No antibacterial activity was observed against either tester strain from crude extracts of the empty vector control.

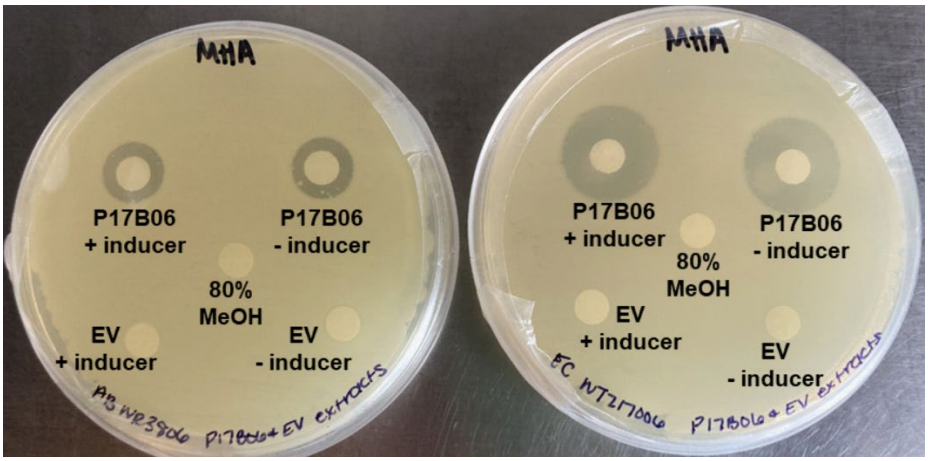

Figure 3.5. Zones of inhibition against *A. baumannii* WR3806 (left) and *E. coli* WTZ17006 (right) of crude extracts (10 mg/ml) from *S. coelicolor* clone P17B06 grown with and without OTC inducer are shown. Discs loaded with crude extracts (10 mg/ml) from *S. coelicolor* empty vector (EV) grown in the presence and absence of OTC were included as negative controls. Discs loaded with the resuspension solvent, 80% methanol (MeOH), were also included as a separate negative control.

The extracts of clone P17B06 and the empty vector were fractionated using column chromatography with HP-20 resin. Four parent fractions (FR1 – FR4) per extract were collected and tested at a concentration of 10 mg/ml. Fraction 2 (FR2) derived from clone P17B06 with and without OTC and eluted with 25% MeOH was active against *A. baumannii* WR3806 (11 mm ZOI) and *E. coli* WTZ17006 (20 mm ZOI) as shown in Fig. 3.6. Fraction 4 (FR4) derived from clone P17B06 cultured with OTC was active against *E. coli* but not *A. baumannii*. In

comparison, FR4 derived from clone P17B06 grown without OTC was not active against either strain suggesting that the activity from this fraction may be linked to the inducer. FR2 from the empty vector was not active against either tester strain and FR4 inhibited the growth of *E. coli*. This was further supported by LC-MS analyses, which detected OTC in the highest abundance in FR4.

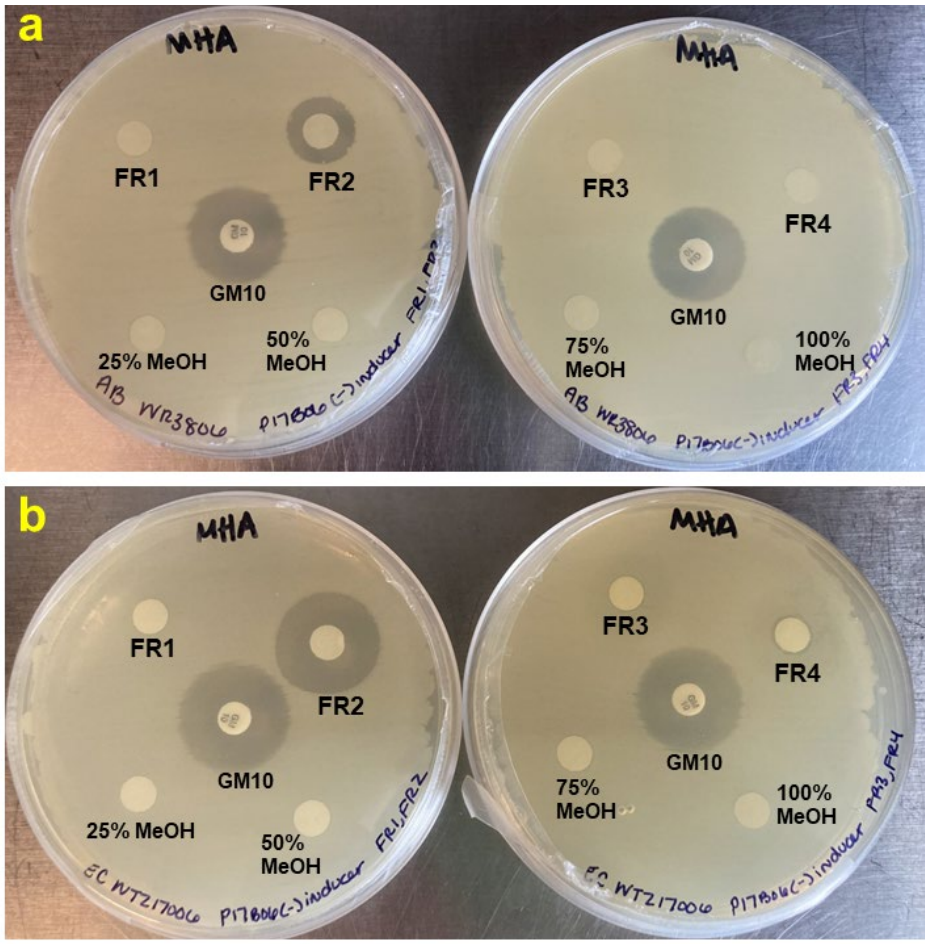

Figure 3.6. Zones of inhibition of parent fraction 2 (FR2; 10 mg/ml) against (a) *A. baumannii* WR3806 and (b) *E. coli* WTZ17006 are shown. This fraction was derived from *S. coelicolor* clone P17B06 cultured in the presence of OTC inducer. Discs loaded with the resuspension solvents (25% - 100% MeOH) were the negative controls. Positive controls were gentamicin 10 µg antibiotic discs (GM10).

A control experiment was subsequently performed to screen for inhibitory activity against the tester strains from four concentrations of OTC (0.1 mg/ml, 0.5 mg/ml, 1 mg/ml, 5

mg/ml) by a disc diffusion assay. All four concentrations of OTC significantly inhibited the growth of *E. coli* WTZ17006. Only the highest concentration of OTC at 5 mg/ml inhibited the growth of *A. baumannii* WR3806, which is a higher concentration than present in FR2. Collectively, these findings demonstrated that OTC eluted in FR4 and could be responsible for the activity of FR4 against *E. coli* WTZ17006, but was less likely to be influencing the activity observed in FR2. As a result, FR2 was deemed the most likely fraction to contain the antimicrobial compounds of interest and was subsequently targeted for further purification.

Bioactive fraction FR2 derived from clone P17B06 grown with OTC was further fractionated using a C18 SPE cartridge and ten sequential subfractions (FR2-1 – FR2-10) that varied in polarity (10 – 100% acetonitrile in water) were collected. The first three subfractions (FR2-1 – FR2-3) collected were colored from golden yellow to yellow and the remaining subfractions were colorless. All ten subfractions were assayed for antibacterial properties and one weakly active subfraction (FR2-2) against *A. baumannii* (7 mm ZOI) and *E. coli* (8 mm ZOI) was identified. The neighboring subfraction (FR2-3) showed inhibitory activity against *E. coli* only (7 mm ZOI). The degree of antibacterial activity from these two subfractions was significantly lower than observed from the parent fraction FR2. Recombining subfractions FR2-1 through FR2-4 resulted in larger zones of inhibition against *A. baumannii* and *E. coli* (Supplementary Fig. S3.4).

A second fractionation approach was attempted of FR2 grown without OTC. Extracts from P17B06 without OTC inducer were used to mitigate the potential for false positives in disc diffusion assays from undesirable antimicrobial effects of the inducer against the tester strains. Moreover, there was no significant difference in activity from extracts grown with and without OTC suggesting that the antimicrobial compounds in these samples were present in similar

abundances. Column chromatography with Sephadex LH-20 afforded nine subfractions (FR2-1S through FR2-9S). Only FR2-8S showed limited inhibitory activity against *E. coli* with a faint halo (~ 9 mm) observed around the disc and notable growth inside the zone (Supplementary Fig. S3.5). No subfractions were active against *A. baumannii*. Recombining all nine subfractions did not result in any ZOIs against either tester strain suggesting that size-exclusion chromatography could not purify the metabolites of interest. A summary of the results from disc diffusion assays of all extracts and fractions is provided in Table 3.2.

Table 3.2. Zones of inhibition of crude extracts and fractions from *S. coelicolor* P17B06 on the test organisms.

Metagenomic clone	Inducer treatment	Sample (10 mg/ml)	Diameter of inhibition zone (mm)	
			***A. baumannii* WR3806**	***E. coli* WTZ17006**
S. coelicolor P17B06	+ OTC	Crude extract	12	20
		FR1	-	-
		FR2	11	20
		FR3	-	-
		FR4	-	8
		FR2-1	-	-
		FR2-2	7	8
		FR2-3	-	7
		FR2-4	-	-
		FR2-5	-	-
		FR2-6	-	-
		FR2-7	-	-
		FR2-8	-	-
		FR2-9	-	-
		FR2-10	-	-
		Combined FR2-1 through FR2-4	10	21
S. coelicolor P17B06	- OTC	Crude extract	12	20
		FR1	-	-
		FR2	12	19
		FR3	-	-
		FR4	-	-
		FR2-1S	-	-
		FR2-2S	-	-
		FR2-3S	-	-
		FR2-4S	-	-
		FR2-5S	-	-
		FR2-6S	-	-
		FR2-7S	-	-
		FR2-8S	-	9*
		FR2-9S	-	-
		Combined FR2-1S through FR2-9S	-	-

Note: *denotes growth of the tester strain observed within the zone of inhibition.

3.4.7. LC-MS analysis of crude extracts and bioactive fractions.

Metabolites present in crude extracts and fractions from *S. coelicolor* clone P17B06 and the *S. coelicolor* empty vector control were analyzed by UPLC-MS-QToF analysis. Select metabolites were annotated by matching spectra against the GNPS public spectral library. UPLC chromatograms were first compared between crude extracts of clone P17B06 and the empty vector clone. No significant, unique peaks to the bioactive extract of clone P17B06 could be detected by visual inspection of chromatograms from ESI^+ and ESI^- modes. Base peak ion (BPI) chromatography under ESI^+ contained clearer chromatographic peaks and less background interference and was subsequently used for multivariate statistical analysis by Progenesis QI software. OPLS-DA was employed to statistically assess differences in the chemical constituents from three separate runs of clone P17B06 and the empty vector extracts. As shown in Fig. 3.7a, clone P17B06 crude extract was distinct from the empty vector crude extract in the OPLS-DA score plot. The S-plot generated from the OPLS-DA model (Fig. 3.7b) aided in identifying ions that were distinct from the empty vector extract. In this plot, the points closer to the lower left and upper right corners showed high variable importance in the projection (VIP) scores (VIP > 1) and more significantly driving the observed separation between the samples. The detailed information for the metabolic features enriched in P17B06 extracts is listed in Supplementary Table S3.5. Many of the ions that were enriched in extracts from clone P17B06 were in very low abundance (below 100 counts).

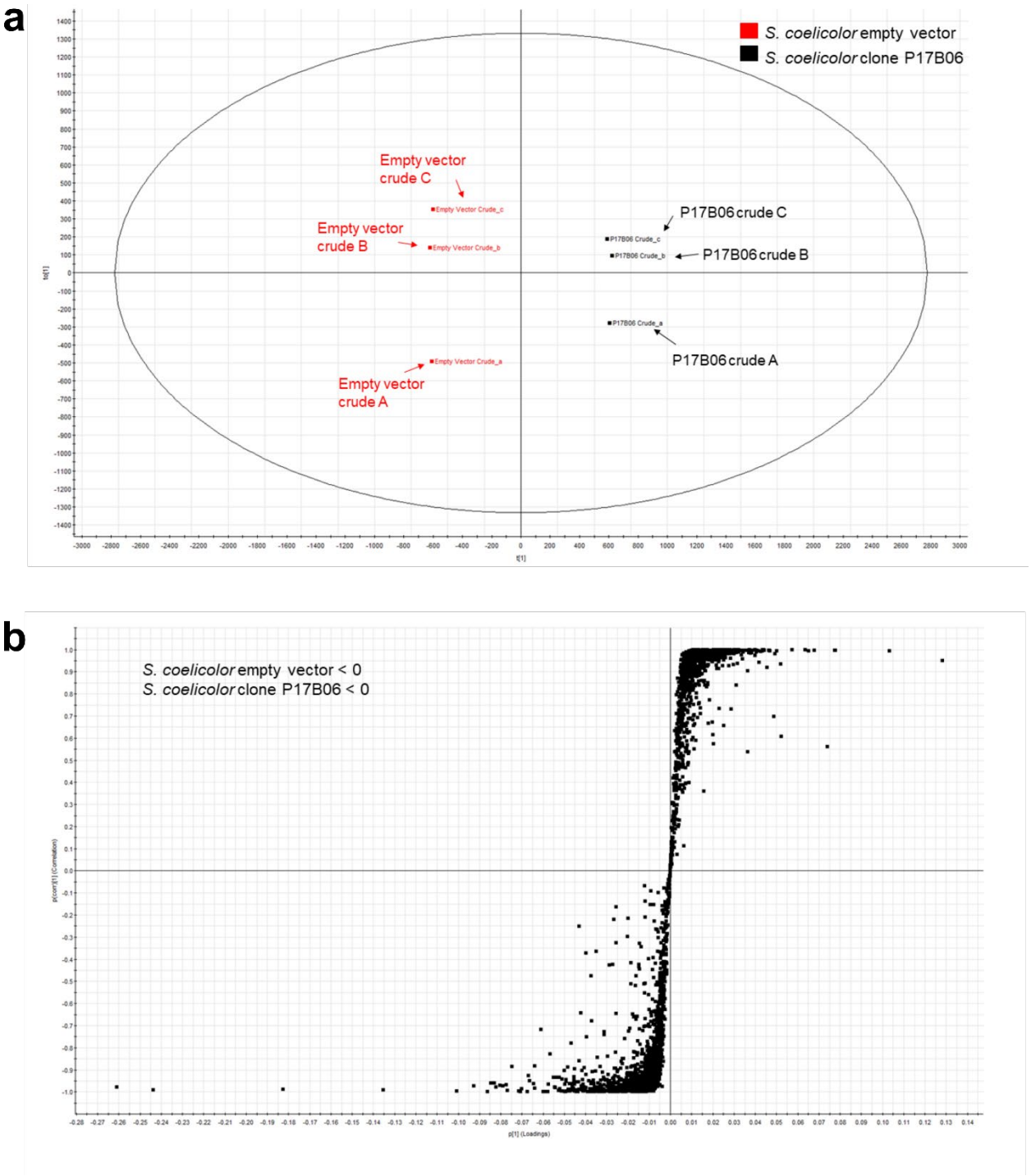

Figure 3.7. OPLS-DA of metabolites analyzed by UPLC-MS-QToF in crude extracts of clone P17B06 and the empty vector clone. (a) Score plot of crude extracts from the empty vector (red box) and crude extracts from clone P17B06 (black box) is shown. Separation along the X-axis represents separation between clones, while separation along the Y-axis represents variation between UPLC runs for individual extracts. (b) S-plot derived from OPLS-DA model indicating differences in metabolic features from crude extracts of the empty vector (lower left) and clone P17B06 (upper right).

The active fraction FR2 and subfractions were subsequently analyzed by LC-MS with the aim of identifying compounds responsible for antimicrobial activity. An overlay of the BPI chromatograms from FR2 and crude extracts from P17B06 and the empty vector did not identify any unique features to FR2 and/or the crude extract from P17B06 that were absent from the control (Fig. 3.8). The metabolic features enriched in P17B06 extracts from OPLS-DA (Supplementary Table S3.5) were then cross-referenced with MS data from FR2. As a result, seven ions that were present in both the fraction FR2 and the P17B06 extract but absent from the empty vector extract were identified. The mass-to-charge ratio (*m/z*) of these ions were *m/z* 182, *m/z* 136, *m/z* 254, *m/z* 146, *m/z* 120, *m/z* 992, *m/z* 698. However, none of these ions appeared to be enriched in fraction FR2 compared to the crude extract of P17B06.

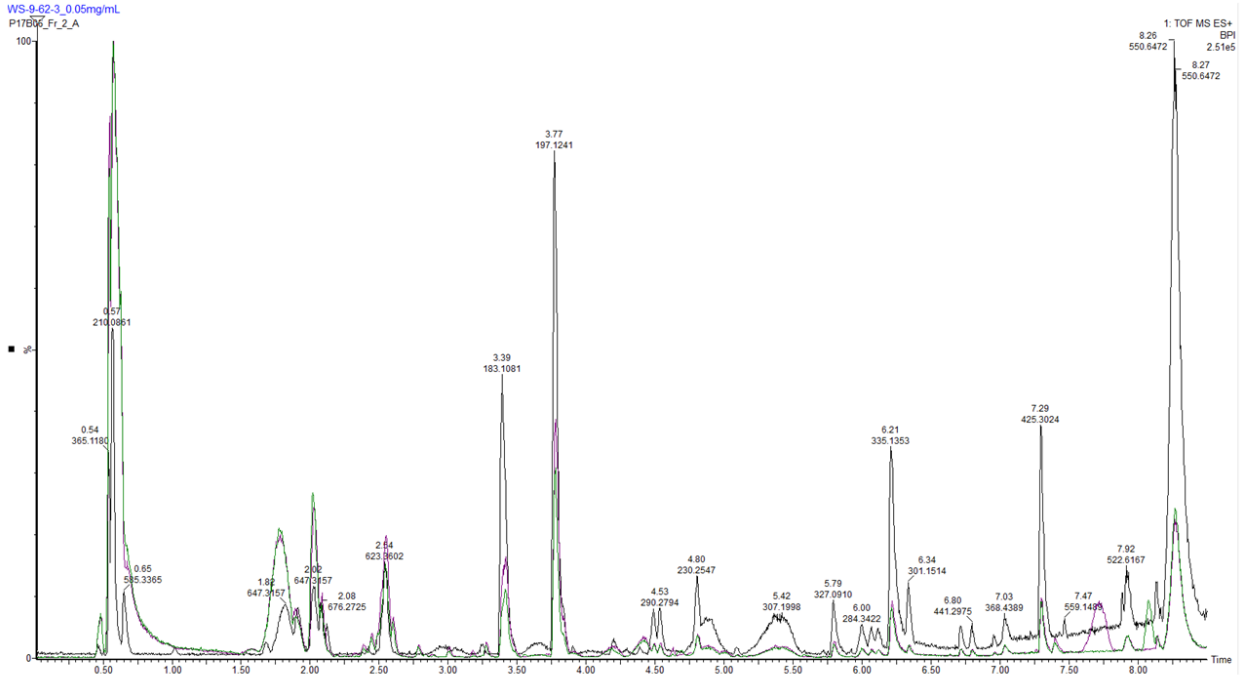

Figure 3.8. Representative (overlay) of base peak chromatograms from UPLC-MS analysis of crude extracts and active fraction from clone P17B06. The UPLC chromatogram of the crude extract of clone P17B06 (purple line) was compared with the crude extract from the empty vector clone (green line) and the active fraction FR2 derived from clone P17B06 (black line) to identify differentially expressed compounds.

To determine if these candidate ions could be detected in active subfractions, MS data was generated from the subfractions collected by C18 SPE including FR2-1 to FR2-4. The MS data were then compared between the active subfractions FR2-3 and FR2-4 to the inactive subfractions FR2-1 and FR2-4. Representative BPI chromatographs of all four subfractions are illustrated in Fig. 3.9. The metabolic profiles were similar between all subfractions and were relatively simple with only 12 major peaks observed in the active subfraction FR2-2. Two differentially expressed peaks were identified between the active and inactive subfractions. The first candidate ion at Rt 2.38 min gave a $[M+Na]^+$ ion at *m/z* 607 and a $[M+H]^+$ ion at *m/z* 585. These ions were present in the active subfractions FR2-2 and FR2-3 in similar abundances, but not the inactive subfractions. The second candidate ion at Rt 2.48 min gave a $[M+Na]^+$ ion at *m/z* 623 and a $[M+H]^+$ ion at *m/z* 601 and was present in subfractions FR2-2, FR2-3, and FR2-4. Although this ion was present in the inactive subfraction FR2-4, the relative abundance was much lower compared to the active subfractions FR2-2 and FR2-3. Thus, this ion was still considered a candidate ion since the abundance of this compound in subfraction FR2-4 may have been too low to inhibit the growth of the tester strains in disc diffusion assays.

Parent masses and fragment spectra of the candidate ions were then compared to spectral libraries in the GNPS database for compound identification. The masses 623 and 585 tentatively corresponded to desferrioxamine E and dehydroxynocardamine, respectively. Desferrioxamine E is a nonpeptide hydroxamate siderophore and has been identified as the major desferrioxamine siderophore produced by *S. coelicolor* (Barona-Gómez et al., 2004). Anti-tumor activity has been observed from the iron-chelator desferrioxamine E (Kalinovskaya et al., 2011), but to our knowledge, antibacterial activity has not been demonstrated. Dehydroxynocardamine is a cyclic peptide and a derivative of desferrioxamine that is produced by *Streptomyces* and other bacteria.

This compound has been described as an enzyme B sortase inhibitor (H.-S. Lee, Shin, et al., 2005) and has not been linked to antibacterial activity.

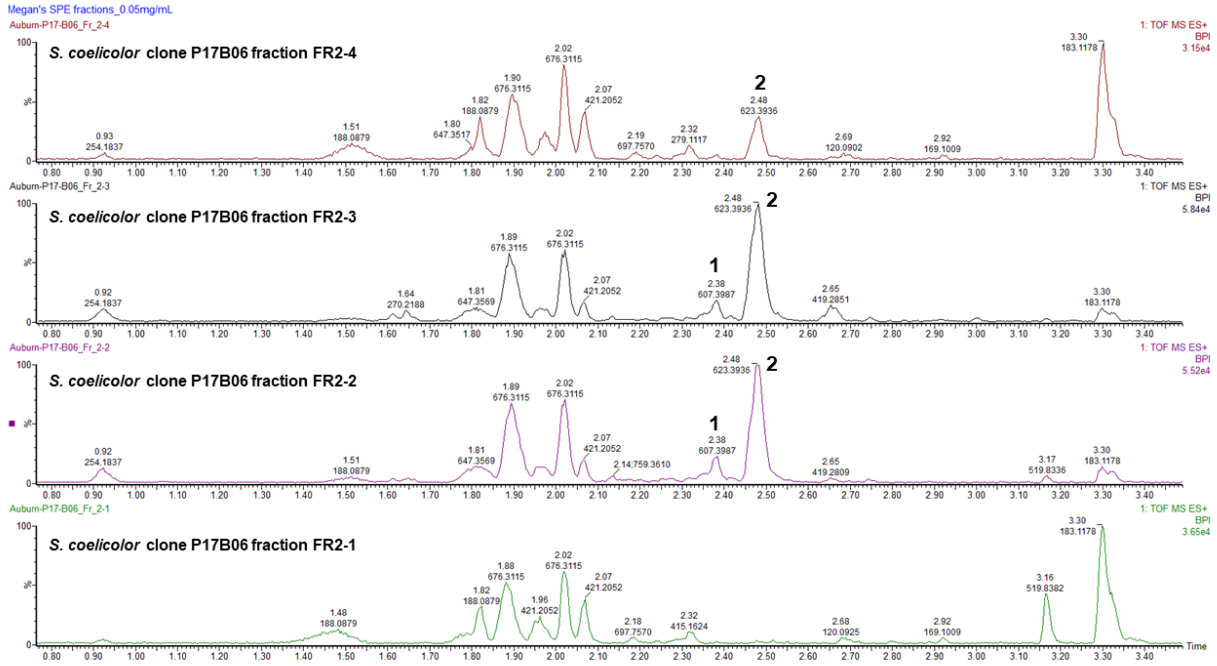

Figure 3.9. Base peak chromatograms of four subfractions derived from clone P17B06. Two peaks unique to or in higher abundance in active subfractions FR2-2 and FR2-3 were identified. These compounds were tentatively identified as desferrioxamine E (1), and dehydroxynocardamine (2) according to spectral library searches against the GNPS database.

The masses corresponding to desferrioxamine E (*m/z* 623) and dehydroxynocardamine (*m/z* 585) were searched against the LC-MS profiles of the empty vector extracts to determine if they were uniquely present in the bioactive samples. These mass ions were present in the empty vector extracts in similar abundance to extracts from clone P17B06. Thus, it is unlikely that these known metabolites are responsible for the antibacterial activity exhibited by clone P17B06. Preliminary LC-MS analysis of crude extracts and bioactive fractions could not identify the antimicrobial peptide produced by metagenomic clone P17B06 associated with inhibitory activity against *A. baumannii*.

3.5. Discussion

The field of metagenomics has become a powerful approach to study the natural product potential of uncultured microorganisms, yet the challenges associated with expressing metagenomic DNA has restricted its applicability as an efficient drug-discovery platform. The identification of new compounds from a metagenomic source is dependent on the ability to express often large and diverse biosynthetic pathways, like PKS and NRPS gene clusters, in a heterologous host (Charles et al., 2017). Challenges associated with expressing these non-native pathways are a driving factor for the low hit rates observed from functional metagenomic screenings, which is estimated to be as low as $10^4 – 10^5$ variants (Colin et al., 2015). In order to successfully express these pathways, the host machinery must be compatible with the DNA of interest (M. Katz et al., 2016). As such, choosing the right host for functional screenings is crucial but which host is best suited for the job remains understudied.

Engineered strains of *Streptomyces* have proven their worth as hosts for native and heterologous expression of natural products derived from closely-related sources. However, application of these optimized hosts has been limited in functional metagenomics and most examples to date have used *S. lividans* Δ*red*Δ*act* and *S. albus* J1074 hosts. In this study, we utilized an engineered *Streptomyces* host, *S. coelicolor* M1154, which is understudied in function-based metagenomics. We targeted 151 metagenome-derived gene clusters for transfer to the engineered host *S. coelicolor* M1154. These gene clusters were derived from a 19,200 clone soil metagenomic library and predicted to encode PKS and/or NRPS biosynthetic pathways (Santana-Pereira et al., 2020). Exconjugants were obtained for 90 BGCs and we assume that the remaining gene clusters were not transferrable due to incompatibility with the host, or in other words, may have produced secondary metabolites that were toxic to *S. coelicolor*. Although we

were not able to test this hypothesis, these BGCs may be a source of interesting antimicrobials that target Gram-positive strains if their expression could be achieved in an alternative host.

PKS and NRPS clusters are known to produce many diverse secondary metabolites that often exhibit biological activity (H. Wang et al., 2014). With the goal of identifying clones that produce antimicrobial metabolites, we screened all *S. coelicolor* clones harboring PKS and/or NRPS pathways (*n*=90) for expression of inhibitory activity against the MDR pathogen *A. baumannii* and a strain of *S. aureus*. We predicted that if we targeted these clusters for expression rather than screening the entire library and used an optimized *S. coelicolor* strain as a heterologous host, then we would achieve a greater hit rate of antibiotic-producing clones compared to the typical hit rate estimated for functional metagenomics (Colin et al., 2015). In support of our prediction, we identified 18 clones (out of 90) expressing antibacterial activity against *A. baumannii* and/or *S. aureus* for a high hit rate of ~20%.

Codon usage plays a key role in heterologous expression and can be a significant cause for poor expression by a host (Gustafsson et al., 2004). Disparities in codon usage can impede expression or result in a protein that is truncated, misfolded, and potentially non-functional (Angov et al., 2008). It is also known that GC% content is associated with codon usage patterns (Wan et al., 2004). Therefore, our study evaluated if the GC% content difference between the host and the BGC influenced the likelihood of expression for an individual clone. Surprisingly, we found no significant effect between the GC% content of the BGC, the host, and the likelihood of expression. Clones displaying bioactivity against the target strains harbored BGCs from a range of low and high GC% contents. This finding suggested that *S. coelicolor* M1154 was robustly capable of expressing BGCs from more distant phylogenetic origins.

Previous studies have suggested that only 10% of BGCs cloned from a non-*Streptomyces* organism within the Actinobacteria phylum will be expressed in a *Streptomyces* host (Myronovskyi & Luzhetskyy, 2019). GC% content of Actinobacteria ranges from 51% to more than 70% (Ventura et al., 2007), and some of the expressed BGCs in our study were below this range indicating that they were derived from non-Actinobacteria taxa. Thus, our data suggest that the likelihood of expressing BGCs from close and distant origins may actually be greater when using an optimized host like *S. coelicolor* M1154. It is possible that some of the 17 gene clusters expressed by *S. coelicolor* and derived from non-Actinobacterial origins encoded all of the machinery needed for regulation, expression, and production of an active metabolite. This would, in turn, decrease the odds of incompatibilities between the host and cloned DNA. However, the exact origin of these BGCs is not known at this time and further studies are needed to investigate this relationship.

In addition to showing that *S. coelicolor* M1154 could express metagenomic BGCs, we found that the antimicrobial activity of some of the clones could be enhanced by an Actinobacteria inducible expression vector (pDualP). This vector takes advantage of recently developed regulatory expression systems in *Streptomyces* including P_{nitA}-NitR (Herai et al., 2004) and P_{otr} (Weishan Wang, Yang, et al., 2016) that are capable of turning gene expression "on" at desired points with the addition of the chemical inducers ƐCL and OTC, respectively. Our study is the first to evaluate these technologies for the expression of metagenomic BGCs. Increased antibacterial activity was observed for four metagenomic BGCs when cloned in pDualP and induced with ƐCL and/or OTC in comparison to the expression by the native promoters. We showed that the inducer treatment leading to the greatest antibacterial activity varied amongst BGCs. We predicted that this was due to the direction of the gene cluster and

location with respect to P_{nitA} and P_{otr}. In contrast to our prediction, we found no significant effect between BGC orientation and location with respect to the inducible promoters on the likelihood of induced expression by ɛCL or OTC. Many of the BGCs are highly unique from known gene clusters and it could be that they possess an unannotated regulatory gene that more strongly influences transcription. Indeed, the expression of BGCs can be controlled by cluster-situated regulators, cross-cluster regulation, or global regulatory systems that are encoded either directly within the gene cluster or in another operon (McLean et al., 2019).

Although we were not able to induce expression of every BGC tested, this technology has demonstrated its advantages for function-based metagenomics due to its ability to express BGCs with varied architectures. Alternative methods to enhance expression of BGCs rely on knowing the architecture of the encoded genes in order to manipulate the pathway by overexpressing regulators or modifying transcription elements (Huo et al., 2019; Pereira, 2019). For metagenomic studies, the pathway organization is often unknown prior to expression. In the case, such as ours, where multiple hits are identified in random functional screens, refactoring each BGC in order to improve expression can be costly and laborious. The approach described in this study provides an opportunity to subclone a gene cluster of interest in an inducible expression system and rapidly screen each clone for enhanced heterologous expression. As a result, clones that are better candidates for downstream chemical analyses can be more rapidly identified.

The next major aim of our work was to identify the antimicrobial metabolites produced by bioactive *S. coelicolor* clones. Several studies have shown that the addition of polymeric adsorptive resins during fermentation can improve production of secondary metabolites by Actinobacteria spp. through sequestering metabolites and decreasing cytotoxic effects of the produced compounds (González-Menéndez et al., 2014). Selecting the appropriate resin can

affect the yield of the targeted product (Dzhavakhiya et al., 2020), and in random, functional metagenomic studies the compound identity is often unknown prior to purification. In our study, we attempted to extract metabolites from four bioactive *S. coelicolor* clones using the resins XAD-4, XAD-7, XAD-16, and HP-20. By testing resins of different adsorptive properties, we were able to obtain methodological information on the nature of the antimicrobial compounds. Our bioactivity evaluations indicated that extracts from nonpolar resins, XAD-16 and HP-20, led to the greatest antibacterial activity from clones P12B21 and P17B06 and no activity was observed from the moderately-polar XAD-7 resin. This data suggested that the antimicrobial compounds produced by these two clones may be more nonpolar in nature since they had a higher affinity for the nonpolar resins when a polar solvent (80% MeOH) was used. Using this data, we continued our "activity-forward" approach and selected clone P17B06 for further analyses since extracts from this clone displayed the greatest degree of inhibition against *A. baumannii*.

Sequence analysis of clone P17B06 revealed that the clone harbored an NRPS-like cluster with ~40% homology to a puwainaphycin cluster. Although the threshold for determining gene cluster novelty is debatable, previous studies have used sequence identity scores below 85% to indicate a potentially novel gene cluster (Ziemert et al., 2012). Thus, the low homology of the P17B06 cluster to any known cluster in the MIBiG or antiSMASH databases provides evidence that it is unique and could encode a novel antimicrobial compound. We were unable to predict a chemical structure from analysis of the BGC, but the five encoded NRPS modules are predicted to catalyze the formation of a pentapeptide. Further examination of the ORFs encoded in the cluster identified a resistance gene with homology to a multidrug-efflux transporter and we predict that this gene may provide self-resistance of the produced metabolite to the host. We also

identified a regulatory gene, yet closer analysis suggested that this is a σ^{54} transcriptional regulator that is not thought to be active in the *S. coelicolor* host (David Cole Stevens, Conway, et al., 2013). We were unable to identify a gene encoding a TE domain that is needed to cleave the peptide from the NRPS machinery (Kohli et al., 2002). This could be attributed to limitations of the prediction tools, particularly if the encoded domain is significantly unique from TE domains in the databases. More work is needed to better understand the architecture of this cluster, how it is regulated, and the mechanisms responsible for the synthesis and release of the final product.

Intriguingly, initial bioinformatics analysis suggested that the P17B06 gene cluster originated from Cyanobacteria taxa. Upon closer inspection of the entire clone insert, we found that many of the genes flanking the NRPS region had higher % similarity (61 – 83%) to genes from the phylum Acidobacteria, whereas the NRPS-related genes had lower % similarity (46 – 66%) to genes from Cyanobacteria. In particular, many of the flanking genes showed greatest similarity to the *Blastocatellia* class of the phylum Acidobacteria. This is a similar finding to a study by Parsley et al. (2011) in which KS domains from metagenomic clones showed low sequence similarity to Cyanobacteria taxa and the complete clone insert sequence was more similar to Acidobacteria. The authors concluded that the BGC was actually derived from Acidobacteria, but was initially misidentified due to the lack of Acidobacteria-derived PKS sequences and genomes in the GenBank database. The %G+C content of the P17B06 insert sequence was within the range of %G+C content for *Blastocatellia* genomes. Additionally, NRPS-related sequences from Acidobacteria are significantly lacking in the GenBank database. Thus, we suspect that the BGC of clone P17B06 is derived from a *Blastocatellia* species. To our knowledge, there has not been a single nonribosomal peptide natural product characterized from

the phylum Acidobacteria. This provides further evidence that the antimicrobial compound(s) produced by clone P17B06 are novel and could be the first example of a natural product identified from the Acidobacteria phylum.

A bioassay-guided fractionation and metabolomics approach was followed in an attempt to identify the antimicrobial metabolites produced by clone P17B06 and assess their novelty. UPLC-MS data with statistical analysis by OPLS-DA aided the detection of metabolites that were unique in crude extracts from clone P17B06 in comparison to an empty vector control. We also successfully obtained an active fraction FR2 from clone P17B06 that showed potent inhibition of the target strains. However, the abundance of the unique and/or enriched compounds in FR2 were very low which inhibited our ability to isolate such low-yielding bioactive compounds. The major compounds detected by UPLC-MS analyses in active subfractions dereplicated to the metabolites desferrioxamine E and dehydroxynocardamine that are known to be produced by the host *S. coelicolor*. Antibacterial activity has not been found from either of these metabolites suggesting that they are not linked to the bioactivity of clone P17B06, and this was further confirmed through the detection of these masses in empty vector extracts. We were unable to detect any peptide compounds in the active extract or fractions despite bioinformatics analysis suggesting that clone P17B06 harbors a NRPS pathway.

The inability to detect and purify the active compounds from clone P17B06 may be due to insufficient quantities or properties of the compound that render it inactive during purification. It is possible that the chemical products escaped detection, due to inherent limits of analytical methods. Another possible explanation is that multiple constituents are required for the biological effect and these components are diluted or separated in further rounds of fractionation. Other studies have reported that the activity of extracts can be attributed to mixtures of

compounds that act synergistically, additively, or antagonistically (Caesar & Cech, 2019). Since subfractions were less active than the parent fraction FR2, we hypothesized that this phenomenon was occurring for clone P17B06. We attempted to test this by recombining subfractions and screening a consolidated sample for activity, but the results were inconclusive. Recombined subfractions from C18 SPE fractionation showed ZOIs more similar to that of FR2. The same was not true for LH-20 recombined subfractions which showed no activity. Thus, we cannot support or refute this hypothesis at this stage. We also found no evidence of degradation products in MS analyses of P17B06 samples indicating that chemical stability was not a major factor contributing to the failed compound characterization.

Despite the advancements made in this study, our inability to obtain a pure compound from clone P17B06 further emphasizes the challenges with heterologous expression. Although we attempted to increase heterologous expression through optimizing media conditions (data not shown) and inducing expression of the BGC, the metabolite yield was still insufficient for accurate detection. There are several future directions that could be explored to overcome these challenges and build upon the research established in this study. Rational engineering of the BGC, the use of alternative hosts, and up-scaling fermentations of bioactive clones may be useful techniques. For clone P17B06, synthetic biology tools could be used to optimize promoters, ribosome binding sites, or codon usage. Refactoring regulatory genes is another common approach shown to improve heterologous expression levels (Bauman et al., 2019). For example, putative repressors could be targeted for deletion. The pathway could also be refactored with strong constitutive or inducible promoters that are compatible with the host. These promoters could be placed in different locations in the cluster and recombinant clones screened for transcript abundance or degree of antibiotic activity. One such location could be near the putative

cluster-situated regulator, ORF18. Alternatively, ORF21, which encodes a SOS response-associated peptidase, may be an interesting target for overexpression or mutation since stress responses have been found to activate secondary metabolism in *Streptomyces* and other bacteria (F. Xu et al., 2017).

Expression of the BGC from clone P17B06 could be compared in other heterologous hosts. Ideally, a host that is more closely related to the phylogenetic origin of this cluster would be used, but there is not a genetically amenable host from the Acidobacteria phyla. Instead, other engineered *Streptomyces* hosts, such as *S. coelicolor* M1154 4x*attB* (unpublished strain) or *S. albus* Del14 (Myronovskyi et al., 2018), could be screened. These hosts have been shown to improve production levels through the addition of multiple *phi*C31 *attB* integration sites that can increase the copy number of the heterologous gene clusters. By refactoring the BGC for more controlled expression and using a host with multiple integration sites for the targeted cluster, we predict that expression could be substantially increased compared to the levels achieved in this study.

Clearly, more work is required to decipher the chemical nature of the antimicrobial metabolites produced by clone P17B06, as well as further studies to characterize other bioactive clones identified in this study. We also acknowledge that a bias in this work is attributed to our functional screening approach. From our collection of PKS and NRPS-containing clones, we were only able to identify those displaying antibiotic activity when expressed in *S. coelicolor* M1154. There may be many interesting metabolites from BGCs that were not expressed in this host. Future studies could perform more detailed sequence analysis of the metagenomic BGCs and attempt to express these gene clusters in hosts that more closely match the predicted origin. This targeted approach could result in the identification of a new set of clones that are capable of

expressing novel antimicrobial compounds. Additionally, we only identified clones that exhibit activity against *A. baumannii* WR3806 and *S. aureus* Xen29. Using a different set of Gram-positive and Gram-negative test strains may affect the hit rate observed from functional screenings of the library.

In conclusion, this study demonstrated the relevance of *S. coelicolor* M1154 as a host for the expression of biosynthetic pathways from a metagenome. We showed that BGCs of low and high GC% could be expressed by *S. coelicolor* M1154 providing support that this host is capable of expressing metagenomic DNA from distant and close phylogenetic origins. In support of this finding, an expressed cluster (P17B06) was predicted to encode an NRPS-like biosynthetic pathway that appeared to be derived from a *Blastocatellia* species. Furthermore, we demonstrated that a dual-inducible expression system could improve heterologous expression for some clones including clone P17B06, but the effect was BGC-dependent. Bioactivity-guided fractionation and LC-MS enabled the detection of unique metabolites in extracts and fractions of clone P17B06. However, we were unable to purify the antimicrobial metabolites and more work is needed to further enhance the metabolite yield by this clone to support purification and chemical structure determinations. Despite this shortcoming, we positively identified a high hit rate of metagenomic clones capable of producing metabolites that inhibit the growth of a MDR pathogen. Thus, we assert this study provides techniques and tools that can advance the field of functional metagenomics.

3.6. Chapter 3 Supplementary Material

Table S3.1. Bacterial strains and plasmids used in this study.

Strain/plasmid/BAC	Description/relevant characteristics[a]	Reference
Strains		
Escherichia coli DH10B BAC-Optimized Replicator v2.0	F- *mcrA Δ(mrr-hsdRMS-mcrBC) endA1 recA1 Φ80dlacZ ΔM15 ΔlacX74 araD139 Δ(ara,leu)7697 galU galK rpsL (Str^R) nupG (attL araC-PBADtrfA250 bla attR) λ*; Large-insert DNA cloning host	Lucigen Corporation
E. coli HB101	F- *recA13*; General cloning host	(Boyer & Roulland-Dussoix, 1969)
E. coli WTZ17006	F- *Δ(araD-araB)567*, *ΔdegP775::kan*, *ΔlacZ4787*(::rrnB-3), *λ*$^-$, *ΔtolC832::FRT*, *rph-1*, *Δ(rhaD-rhaB)568*, *hsdR514*; antibiotic-susceptible strain used for bioassays	(Baba et al., 2006)
Streptomyces coelicolor M1154	M145 Δ*act* Δ*red* Δc*pk* Δ*cda rpoB*(C1298T) *rpsL*(A262G); Host strain for heterologous expression	(Gomez-Escribano & Bibb, 2011)
Staphylococcus aureus Xen29	Biofilm-producing strain used for bioassays	(Park et al., 2012)
Acinetobacter baumannii WR3806	Multidrug-resistant strain used for bioassays	(Taitt et al., 2014)
Plasmids		
pSmartBAC-S	Cm^R, Apra^R, conjugative, integrative, inducible-copy	(Nasrin et al., 2018)
pRK2013	Tra$^+$; Kan^R; Conjugative helper plasmid in triparental matings	(Figurski & Helinski, 1979)
pDualP	Apra^R	(Stankey et al., 2022)

[a]Abbreviations: Str^R, streptomycin resistant; Cm^R, chloramphenicol resistant; Tra$^+$, conjugal transfer ability; Kan^R, kanamycin resistant; Apra^R, apramycin resistant

Table S3.2. Primers used in this study.

Primer name	Sequence	Amplicon length
UniversalBAC F	CGGCAGGTATATGTGATGGGTTAAAATG	200 bp
UniversalBAC R	ACCTCTTACGTGCCGATCAG	

Table S3.3. Validated BGC-containing clones from the Cullars soil metagenomic library that were transferred to *S. coelicolor* M1154 (this study) and/or *E. coli* BTRA (previous study). Clones that expressed bioactivity in *S. coelicolor* M1154 or *E. coli* BTRA are noted. Predictions for BGC types, lengths, and number of biosynthetic domains were analyzed by antiSMASH 5.0 (Blin et al., 2019).

Clone	Contig(s)	GC (%)	Identified by:	Cloned Insert Length (bp)	Predicted BGC type	Predicted BGC length (bp)	Number of predicted biosynthetic domains	Heterologous expression host	Bioactivity
P01A07	P01A1.48	69.0	NGS-PCR	128761	PKS-NRPS	96149	33	*S. coelicolor* M1154	-
								E. coli BTRA	-
P01H15	P01A1.21	71.3	NGS	64339	PKS-NRPS	4722	15	*S. coelicolor* M1154	-
P01N03	P01B1.33 P01B1.73	58.1	NGS	40763	NRPS	41349	21	*S. coelicolor* M1154	-
P02K21	P02A1.196	65.5	NGS	107433	NRPS	14523	2	*S. coelicolor* M1154	-
P02O05	P02A1.12	56.3	NGS	106305	NRPS	13762	9	*S. coelicolor* M1154	-
P03C03	P03A1.52	54.6	NGS	62165	NRPS	118628	6	*S. coelicolor* M1154	-
P04F05	P04B1.68	71.1	NGS	129359	NRPS	98035	8	*S. coelicolor* M1154	-
P04H02	P04B1.15 P04B1.67	54.8	NGS	140742	PKS-NRPS	97614	72	*S. coelicolor* M1154	-
P04O01	P04A1.29	57.7	NGS	53001	PKS	2609	8	*S. coelicolor* M1154	-
P05D13	P05B1.29	54.7	NGS	71074	NRPS	51314	35	*S. coelicolor* M1154	-
P05G19	P05A1.676	54.2	NGS	90869	NRPS	39049	20	*S. coelicolor* M1154	-
P05K23	P05A1.752	54.6	NGS	108335	NRPS	1278	4	*S. coelicolor* M1154	-
P05P18	P05B2.4	57.0	NGS	121566	NRPS	58216	5	*E. coli* BTRA	-
P06A05	P06A1.18	57.0	NGS	87898	PKS-NRPS	83133	9	*S. coelicolor* M1154	-
P06G05	P06A1.1	56.0	NGS	102475	PKS-NRPS	68150	18	*S. coelicolor* M1154	-
P06G11	P06A1.285	55.8	NGS	129397	PKS-NRPS	52348	52	*S. coelicolor* M1154	-
P07D12	P07B2.21	57.2	NGS	115226	NRPS	24317	20	*S. coelicolor* M1154	-
P07H03	P07B1.70	70.7	NGS-PCR	31569	PKS-NRPS	79592	16	*S. coelicolor* M1154	-
								E. coli BTRA	Antifungal (*C. neoformans*)
P07H07	P07B1.72	54.2	NGS	30585	NRPS	37239	20	*S. coelicolor* M1154	-
P07L16	P07B2.36	65.0	NGS	70028	PKS	93770	8	*S. coelicolor* M1154	-
P07M04	P07A1.173	64.8	NGS-PCR	11583	PKS-NRPS	39662	5	*E. coli* BTRA	-
P07O04	P07A2.48	66.0	NGS	61286	PKS-NRPS	106177	10	*S. coelicolor M1154*	-
P07O09	P07A1.56	55.6	NGS	108907	PKS-NRPS	87047	16	*S. coelicolor M1154*	-
P08A19	P08A1.2	56.0	NGS	96916	PKS-NRPS	43697	38	*S. coelicolor M1154*	-
P08C17	P08A2.46 P08A1.540 P08A1.626 P08A1.683	55.2	NGS	124248	NRPS	6693	33	*S. coelicolor M1154*	-

Clone	Contig(s)	GC (%)	Identified by:	Cloned Insert Length (bp)	Predicted BGC type	Predicted BGC length (bp)	Number of predicted biosynthetic domains	Heterologous expression host	Bioactivity
P09P03	P09B1.20	63.8	NGS-PCR	110970	PKS-NRPS	73677	11	*S. coelicolor* M1154	Antibacterial (*S. aureus* Xen29; *A. baumannii* WR3806)
P11H15	P11B1.15	54.8	NGS	77085	NRPS	66479	25	*S. coelicolor* M1154	-
P11I13	P11A1.18	63.0	NGS	110052	PKS	78569	11	*S. coelicolor* M1154	Antibacterial (*S. aureus* Xen29)
P11I24	P11A2.58	54.0	NGS	68208	PKS-NRPS	57478	45	*S. coelicolor* M1154	-
P11J18	P11B2.361 P11B2.716 P11B2.149 P11B2.133 P11B2.172 P11B2.474	50.6	NGS	42994	NRPS	87614	11	*S. coelicolor* M1154	-
P11N21	P11B1.351 P11B1.159	50.8	NGS	49818	NRPS	31688	11	*S. coelicolor* M1154	-
P11O04	P11A2.44	67.7	NGS-PCR	77067	PKS-NRPS	74462	31	*S. coelicolor* M1154	-
								E. coli BTRA	-
P11P20	P11B2.532	65.0	NGS-PCR	150553	PKS-NRPS	31118	59	*S. coelicolor* M1154	-
								E. coli BTRA	-
P12B21	-	69.8	NGS	60007	NRPS	36531	2	*S. coelicolor* M1154	Antibacterial (*A. baumannii* WR3806)
P15C22	P15A1.77 P15A2.23	62.8	NGS	83712	NRPS	39703	18	*S. coelicolor M1154*	Antibacterial (*S. aureus* Xen29)
P15P01	P15B1.63	56.1	NGS	132498	NRPS	45875	8	*S. coelicolor M1154*	-
P15P18	P15B2.5	54.6	NGS	81084	NRPS	70052	25	*S. coelicolor M1154*	-
P16O06	P17B2.38	54.6	NGS	67482	NRPS	42261	17	*S. coelicolor M1154*	-
P17B06	P17B2.114	60.0	NGS	100101	NRPS	105964	12	*S. coelicolor M1154*	Antibacterial (*A. baumannii* WR3806; *E. coli* WTZ17006)
P17F15	P17B1.340	70.0	NGS-PCR	71189	PKS-NRPS	88261	10	*S. coelicolor M1154*	Antibacterial (*A. baumannii* WR3806)
								E. coli BTRA	-
P17G08	P17A2.40	56.5	NGS	65444	NRPS	2884	11	*S. coelicolor M1154*	-
P17I02	P17A2.30	66.2	NGS	125265	NRPS	43874	10	*S. coelicolor M1154*	-
P19H07	P19B1.221	52.7	NGS	16484	NRPS	16484	2	*S. coelicolor M1154*	-
P19L22	P19A2.82 P19B2.34	56.3	NGS	87188	NRPS	51653	7	*S. coelicolor M1154*	Antibacterial (*A. baumannii* WR3806)
P19N16	P19B2.404	58.3	NGS	60112	NRPS	2302	20	*S. coelicolor M1154*	-
P19O08	P19A2.56	41.9	NGS	93369	PKS	58388	11	*S. coelicolor M1154*	Antibacterial (*A. baumannii* WR3806)
P20F15	P20B1.242 P20B1.293	68.5	NGS	75278	PKS-NRPS	53104	30	*S. coelicolor M1154*	-
P20P17	P20B1.264 P20B1.518	63.6	NGS	26094	NRPS	49110	11	*S. coelicolor M1154*	-
P21E03	P21A1.38	54.8	NGS	91188	NRPS	53798	15	*S. coelicolor M1154*	-
								E. coli BTRA	-
P22N20	P22B2.30	52.0	NGS-PCR	103714	PKS	98719	8	*S. coelicolor M1154*	-
								E. coli BTRA	-
P22P12	P22B2.239	53.2	NGS	44041	NRPS	3014	13	*S. coelicolor M1154*	-
P24H02	P24B2.16	47.8	NGS	87129	PKS	67260	6	*S. coelicolor M1154*	-
P24H13	P24B1.29	51.6	NGS-PCR	99579	PKS-NRPS	43697	16	*S. coelicolor M1154*	-
								E. coli BTRA	

Clone	Contig(s)	GC (%)	Identified by:	Cloned Insert Length (bp)	Predicted BGC type	Predicted BGC length (bp)	Number of predicted biosynthetic domains	Heterologous expression host	Bioactivity
P24L06	P24B2.26	40.0	NGS	97272	PKS	97771	8	*S. coelicolor* M1154	Antibacterial (*S. aureus* Xen29)
P25F10	P25B2.32	62.7	NGS	64797	NRPS	45350	5	*S. coelicolor* M1154	-
P25M03	P25A1.10	54.8	NGS	87165	NRPS	71332	29	*S. coelicolor* M1154	-
P25O21	P25A1.147	62.4	NGS	37775	NRPS	29896	8	*S. coelicolor* M1154	-
P26D23	P26B1.113	57.3	NGS	51817	NRPS	24124	3	*S. coelicolor* M1154	-
P27C17	P27A1.12	41.2	NGS	102918	NRPS	47322	7	*S. coelicolor* M1154	-
P27K06	P27A2.75	57.0	NGS	139647	NRPS	20072	41	*S. coelicolor* M1154	Antibacterial (*S. aureus* Xen29)
P27L24	P27B2.506	55.1	NGS	112758	NRPS	5913	12	*S. coelicolor* M1154	-
P27O06	P27A2.39	57.4	NGS	139647	NRPS	2970	41	*S. coelicolor* M1154	-
P28I16	P28A2.16	55.0	NGS	74990	PKS-NRPS	38208	10	*S. coelicolor* M1154	-
P28L15	P28B1.439 P28B1.133 P28B1.257 P28B1.63 P28B1.237 P28B1.390	55.7	NGS	110926	NRPS	23644	20	*S. coelicolor* M1154	-
P28O04	P28A2.202	61.0	NGS	85938	NRPS	77087	30	*S. coelicolor* M1154	-
P28O18	P28A2.23	61.0	NGS	85938	NRPS	-	30	*E. coli* BTRA	-
P29M03	P29A1.109	63.9	NGS	97291	PKS	47644	5	*S. coelicolor* M1154	-
P29O18	P29A2.15	56.6	NGS	114397	PKS-NRPS	117998	77	*S. coelicolor* M1154	-
P32A16	P32A2.546	68.7	NGS	59698	NRPS	83726	9	*S. coelicolor* M1154	Antibacterial (*A. baumannii* WR3806)
P33A17	P33A1.19	54.2	NGS	114632	NRPS	34816	8	*S. coelicolor M1154*	-
P33I07	P33A1.40	40.0	NGS	91908	PKS	53131	30	*S. coelicolor M1154*	-
P35G10	P35A2.492 P35A2.22	56.7	NGS	69371	NRPS	5532	42	*S. coelicolor M1154*	-
P35K06	P35A2.331	66.0	NGS	142359	PKS-NRPS	50486	13	*S. coelicolor M1154*	-
P35L08	P35B2.21	59.0	NGS-PCR	81817	PKS	4886	10	*S. coelicolor M1154*	Antibacterial (*A. baumannii* WR3806)
								E. coli BTRA	-
P35M06	P35A2.5	40.0	NGS	96287	PKS	30405	5	*S. coelicolor M1154*	-
P36D09	P36B1.54 P36B1.60	58.8	NGS	76453	NRPS	75565	30	*S. coelicolor M1154*	Antibacterial (*S. aureus* Xen29)
P37L16	P37L16	63.8	NGS-PCR	57334	PKS-NRPS	4799	17	*E. coli BTRA*	-
P38B06	P38B2.77	70.1	NGS	91764	NRPS	47012	7	*S. coelicolor M1154*	Antibacterial (*S. aureus* Xen29)
P38F23	P38B1.47	64.8	NGS	35834	NRPS	3270	9	*S. coelicolor M1154*	Antibacterial (*S. aureus* Xen29)
P38G02	P38G02	54.6	NGS-PCR	91476	PKS-NRPS	3192	24	*S. coelicolor M1154*	-
								E. coli BTRA	Anticancer
P38H14	P38B2.118	53.8	NGS	106831	NRPS	32332	6	*S. coelicolor M1154*	-
P38J21	P38B1.620	62.5	NGS	97705	NRPS	34263	8	*S. coelicolor M1154*	-
P38N20	P38B2.22	54.0	NGS	96293	NRPS	5942	5	*S. coelicolor M1154*	Antibacterial (*S. aureus* Xen29)
P39I02	P39A1.8	57.6	NGS	135991	PKS-NRPS	70579	15	*S. coelicolor M1154*	Antibacterial (*A. baumannii* WR3806)

Clone	Contig(s)	GC (%)	Identified by:	Cloned Insert Length (bp)	Predicted BGC type	Predicted BGC length (bp)	Number of predicted biosynthetic domains	Heterologous expression host	Bioactivity
P39I17	P39A1.32 P39B1.24	64.7	NGS	92944	NRPS	73906	42	*S. coelicolor* M1154	-
P40C07	P40A1.53	58.3	NGS	142632	NRPS	103818	9	*S. coelicolor* M1154	-
P40C19	P40A1.32	65.9	NGS	113503	NRPS	9815	5	*S. coelicolor* M1154	-
P40O18	P40A2.5	69.0	NGS	136500	NRPS	1137	12	*S. coelicolor* M1154	-
P41O18	P41.7	62.7	NGS	61544	PKS	52280	9	*S. coelicolor* M1154	-
								E. coli BTRA	-
P42P24	P42.2475	53.9	NGS	19065	NRPS	65394	13	*S. coelicolor* M1154	-
								E. coli BTRA	-
P49O09	P49.2215 P49.1560 P49.2119	71.9	NGS	124799	NRPS	2251	57	*S. coelicolor* M1154	-
								E. coli BTRA	-
P50M18	P50.95	57.1	NGS	84718	PKS	1755	6	*S. coelicolor* M1154	Antibacterial (*S. aureus* Xen29; *A. baumannii* WR3806)

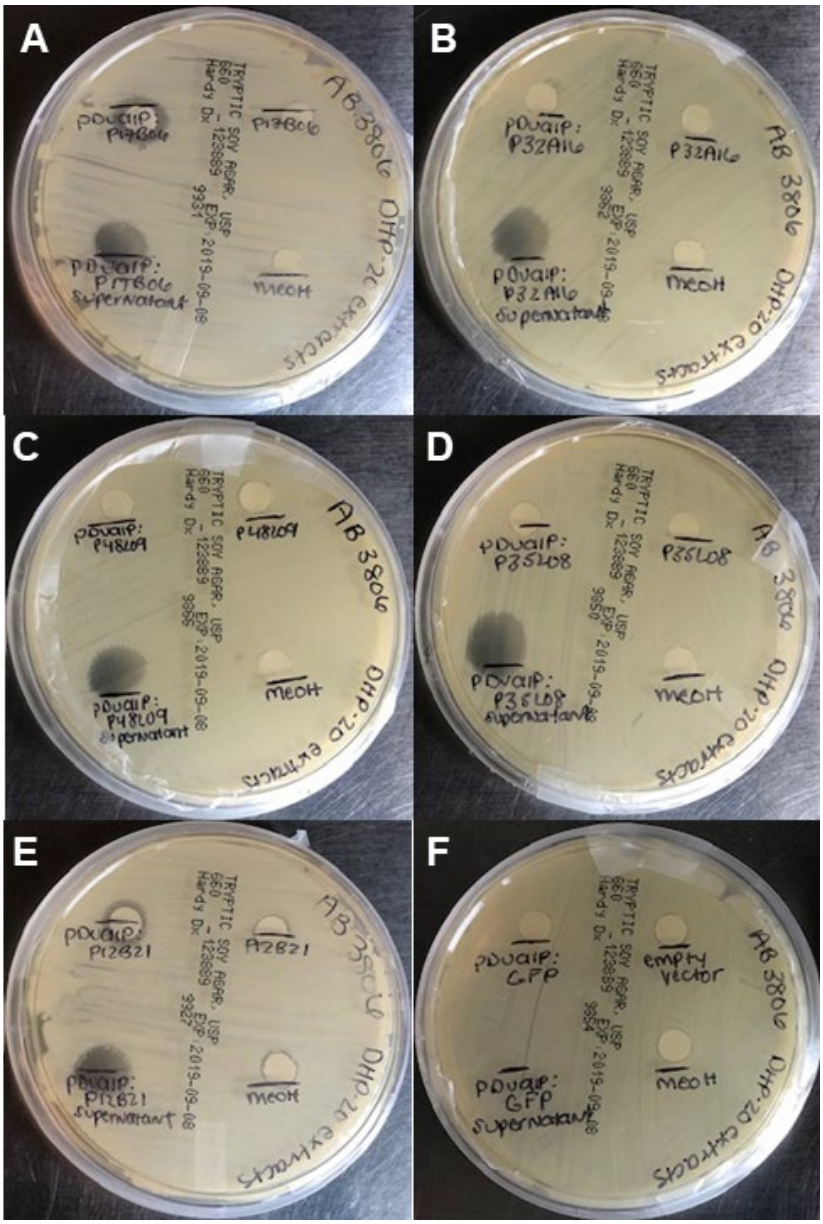

Figure S3.1. Antibacterial activity of HP-20 extracts and cell-free supernatants from *S. coelicolor* metagenomic clones. Zones of inhibition against *A. baumannii* WR3806 are shown for each clone and include: discs loaded with 10 mg/ml of HP-20 extract from the pDualP version of each clone (top left disc), discs loaded with 10 mg/ml of HP-20 extract from the pSmartBAC-S version of each clone (top right disc), 10 µl of cell-free supernatant spotted (bottom left disc), and discs loaded with 100% methanol (bottom right disc). Each plate corresponds to a clone: A) Clone P17B06, B) Clone P32A16, C) Clone P48L09, D) Clone P35L08, E) Clone P12B21, and F) Empty vector clone.

Table S3.4. Predicted function of the genes from the P17B06 BGC.

ORF	Gene length (bp)	Closest homolog [source]	Predicted function	E-value	Protein ID (%)	NCBI Accession
1	1119	Hypothetical protein [uncultured bacterium]	Unknown	0.0	67.75	QEO74846.1
2	1620	PhpK family radical SAM P-methyltransferase [*Acidobacteria* bacterium]	P-methyltransferase	0.0	67.30	MCA1612824.1
3	1947	Asparagine synthase (glutamine-hydrolyzing) [uncultured bacterium]	Asparagine synthesis	0.0	66.40	QEO74760.1
4	5538	AMP-dependent synthetase and ligase [uncultured bacterium]	AMP synthesis	0.0	46.51	QEO74753.1
5	609	Nonribosomal peptide synthetase [*Microcystis* sp. M04BS1]	Peptide synthesis	2e-50	54.19	MCA2554884.1
6	690	Capreomycidine synthase [*Nostoc flagelliforme*]	Capreomycidine synthesis	4e-57	57.85	WP_190944349.1
7	1029	TauD/TfdA family dioxygenase [*Nostoc* sp. WHI]	Alpha-ketoglutarate-dependent dioxygenase	2e-138	54.76	WP_196519570.1
8	11091	Nonribosomal peptide synthetase [*Nostoc* sp. UHCC 0702]	Peptide synthesis	0.0	47.42	WP_206268250.1
9	1050	Dioxygenase TauD/TfdA [uncultured bacterium]	Alpha-ketoglutarate-dependent dioxygenase	3e-149	60.42	QEO74275.1
10	1083	TonB-dependent receptor [*Pyrinomonadaceae* bacterium]	Outer membrane receptor	6e-74	69.09	MBC8031683.1
11	630	TVP38/TMEM64 family protein [*Pyrinomonadaceae* bacterium]	Inner membrane protein	2e-82	65.07	MBC8031685.1
12	1242	PQQ-dependent sugar dehydrogenase [*Pyrinomonadaceae* bacterium]	Glucose dehydrogenase	0.0	74.45	MBC8031693.1
13	540	Manganese efflux pump [*Pyrinomonadaceae* bacterium]	Manganese exporter	8e-55	62.80	MBC8031692.1
14	1491	TPA: transporter [*Blastocatellia* bacterium]	Transport	0.0	78.76	HBB97703.1

ORF	Gene length (bp)	Closest homolog [source]	Predicted function	E-value	Protein ID (%)	NCBI Accession
15	3165	Multidrug-efflux RND transporter permease subunit [*Acidobacteria* bacterium]	Resistance	0.0	75.78	MBS1791895.1
16	1335	TPA: efflux transporter periplasmic adaptor subunit [*Blastocatellia* bacterium]	RND family efflux transporter MFP subunit	8e-180	73.33	HBB97705.1
17	162	None	Unknown	-	-	-
18	2058	Fis family transcriptional regulator [*Blastocatellia* bacterium]	σ^{54} dependent transcriptional regulator	0.0	83.07	PWT89797.1
19	771	Isocitrate lyase/phosphoenolpyruvate mutase family protein [*Pyrinomonadaceae* bacterium]	Lyase activity/metal ion binding	3e-90	61.90	MBA3323073.1
20	1356	Hypothetical protein [candidate division NC10 bacterium]	Unknown	1e-40	41.04	MBI4391632.1
21	210	SOS response-associated peptidase [*Acidobacteria* bacterium]	Peptidase	4e-27	76.81	MCA1579262.1
22	240	None	Unknown	-	-	-
23	2349	Hypothetical protein C5B55_01185 [*Blastocatellia* bacterium]	Unknown	0.0	70.18	PWT95511.1
24	141	None	Unknown	-	-	-

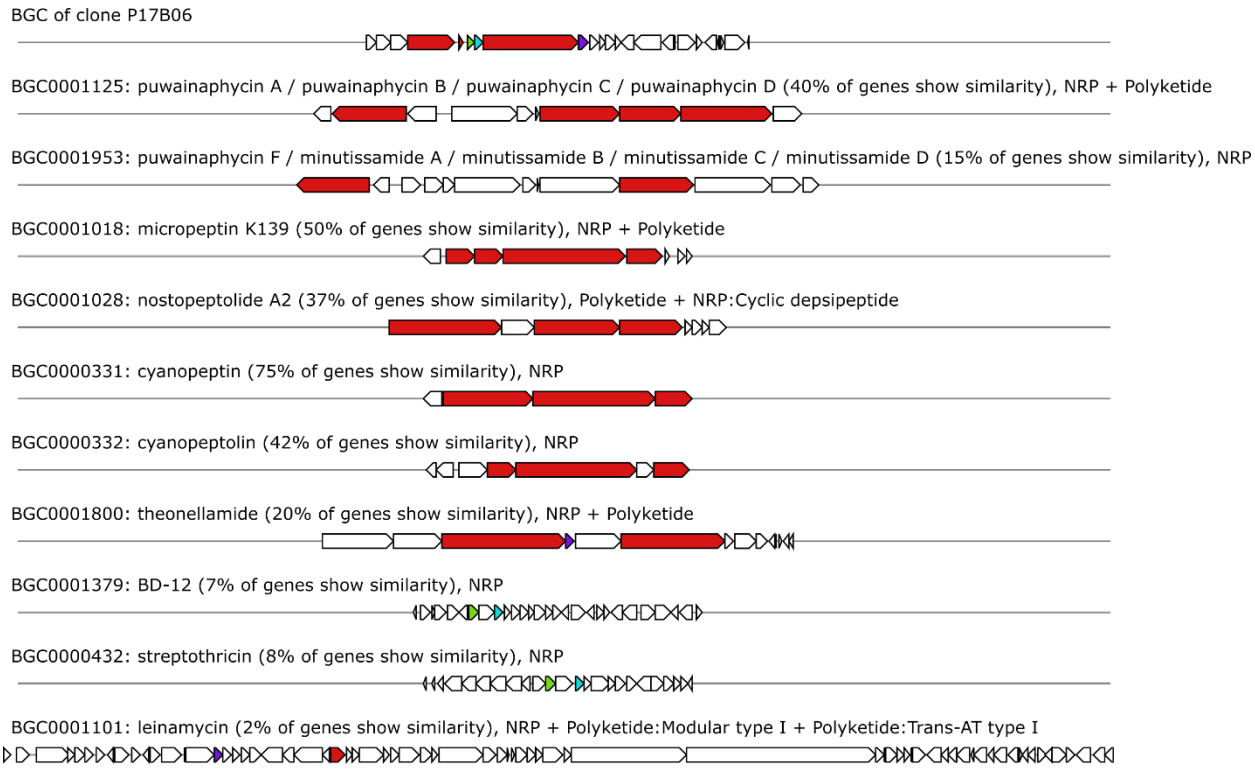

Figure S3.2. Comparison of the BGC sequence from clone P17B06 with genomic regions that are similar in other experimentally-validated clusters in the MiBIG database. Colored genes are putative homologs according to antiSMASH.

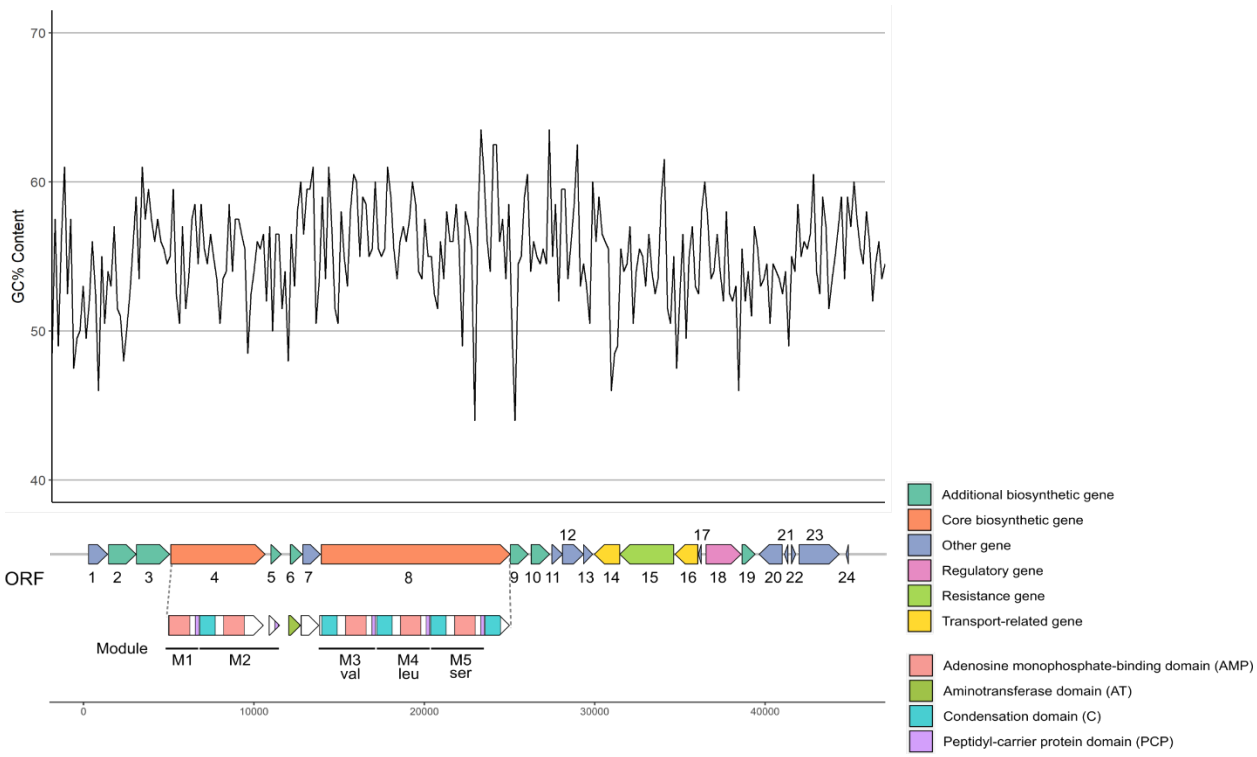

Figure S3.3. Distribution of the GC content across the predicted gene cluster of clone P17B06.

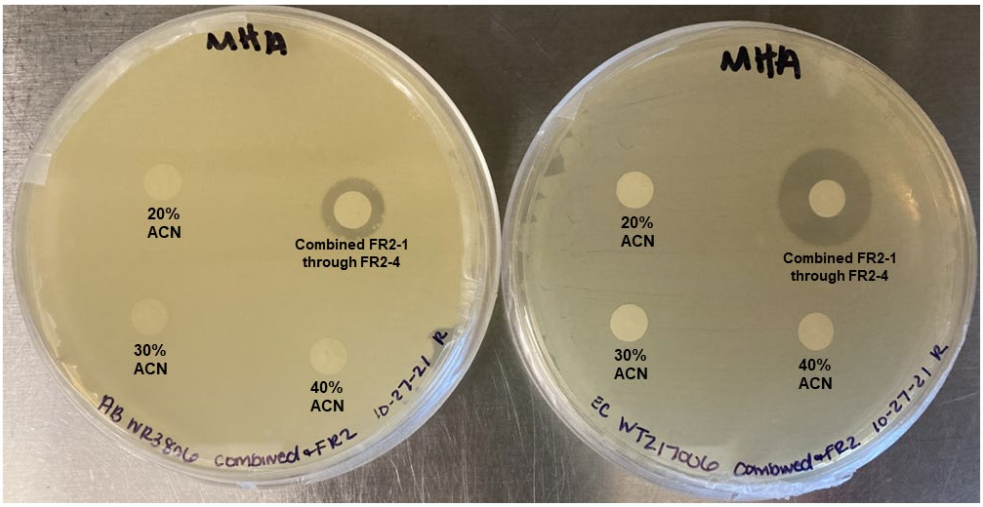

Figure S3.4. Zones of inhibition against *A. baumannii* WR3806 (left) and *E. coli* WTZ17006 (right) of combined subfractions FR2-1, FR2-2, FR2-3, and FR2-4 from SPE fractionation. These subfractions were derived from *S. coelicolor* clone P17B06 grown with OTC inducer are shown. Discs loaded with the resuspension solvents, 20% - 40% acetonitrile (ACN) were included as a negative control.

Figure S3.5. A faint zone of inhibition against *E. coli* WTZ17006 of subfraction FR2-8S from Sephadex LH-20 column chromatography. This subfraction was derived from *S. coelicolor* clone P17B06 grown without OTC inducer. Growth of the tester strain *E. coli* WTZ17006 was observed within the zone of inhibition. Discs loaded with the resuspension solvents, 20% - 40% acetonitrile (ACN) were included as a negative control.

Table S3.5. Detailed information of metabolic features enriched in clone P17B06 crude extracts obtained by OPLS-DA analysis.

No.	*m/z*	RT (min)	Max abundance	Adducts	Neutral mass (Da)	Normalised abundance					
						P17B06 crude A	P17B06 crude B	P17B06 crude C	Empty vector crude A	Empty vector crude B	Empty vector crude C
1	287.10	2.30	41	M+H, M+Na	264.11	49.52116	24.83397	47.68489	4.17278	0	0
2	583.36	3.38	51			51.14513	60.57226	39.99926	0	0	4.15221
3	182.08	3.19	125			120.0427	132.2461	123.3043	2.26675	0	3.10303
4	336.12	3.72	422	M+H, M+Na	313.13	432.3495	420.5893	412.0133	5.09952	8.23815	0.06785
5	136.08	3.19	119			121.0404	119.8053	117.5766	0	2.54212	0
6	254.12	3.19	153			162.3791	159.8439	136.9159	2.58090	0	0
7	219.07	2.39	39			33.95891	31.80453	51.81952	0	0.65615	0
8	233.06	1.21	4			4.803039	2.492698	5.180557	0	0	0
9	749.86	4.66	22			26.27919	25.55921	15.57881	0	0	0
10	710.38	4.65	37			36.99689	33.82153	40.00008	0	0	0
11	586.24	4.22	25			29.26649	25.04978	19.89126	0	0	0
12	1257.33	3.38	41			56.80935	34.69923	31.89364	0	0	0
13	887.34	3.23	76			77.94028	62.69609	86.38754	0	0	0
14	306.13	3.19	33			40.6932	23.91043	33.52774	0	0	0
15	165.06	3.19	27			28.31868	22.86335	30.59412	0	0	0
16	206.08	2.91	27			36.13236	18.90935	25.76562	0	0	0
17	291.07	2.89	32			31.75777	21.30961	44.17189	0	0	0
18	146.06	2.89	26			26.20717	23.47531	28.63393	0	0	0
19	144.08	2.89	16			15.22095	16.94799	14.68163	0	0	0
20	132.08	2.89	24			27.5226	22.35519	23.58133	0	0	0
21	120.05	2.71	70			69.32465	69.40378	72.19883	0	0	0
22	992.98	2.71	46			23.81736	51.55125	64.09412	0	0	0
23	722.08	2.70	13			22.69958	11.79697	5.366821	0	0	0
24	514.04	2.39	43			45.25903	50.94775	32.97864	0	0	0
25	698.03	2.37	10			4.633228	16.4747	9.742828	0	0	0
26	654.47	2.09	27			30.24181	40.67164	8.9893	0	0	0
27	803.92	5.13	55			55.83949	57.58405	51.79915	0	0	0
28	523.60	7.81	11			22.34546	5.43896	4.008545	0	0	0

Chapter 4

Characterization and taxonomy of a metagenome-derived biosynthetic gene cluster that encodes an antimicrobial secondary metabolite

4.1. Abstract

Microorganisms can produce a diverse array of secondary metabolites that have been valuable for medical and biotechnological applications. The ecological role of microbial secondary metabolites is poorly understood, but it is likely that these compounds are important mediators of microbial interactions in nature. The genomics era has made it possible to study the genes encoding the synthesis of these metabolites, called biosynthetic gene clusters (BGCs), from cultured bacteria and metagenomes which can provide fundamental insights into their function, diversity, and evolution. In this study, we characterized a BGC, called P12B21 BGC, derived from a soil metagenomic library that encodes the synthesis of an antimicrobial secondary metabolite. We determined that the P12B21 BGC spans 26-kb and contains 28 open readings frames (ORFs), some of which likely encode a hybrid nonribosomal peptide synthetase-polyketide synthase (NRPS-PKS) biosynthetic pathway. Furthermore, we identified several genes flanking the BGC boundary that are involved in glycosylation, and by deleting those genes, we observed a loss in bioactivity which may indicate that the functional secondary metabolite is glycosylated. Bioinformatic analysis revealed that the P12B21 BGC is highly homologous to a cluster from *Microbacterium hatanonis* JCM 14558^{T}, suggesting that the cluster originated from a *Microbacterium* species. Through PCR and dot blot hybridizations, we did not identify a homologous cluster in 65 other endophytic *Microbacterium* species providing evidence that this cluster is not conserved across the entire genus. We propose that this cluster has been maintained by selection in some *Microbacterium* species due to the competitive advantages of the encoded antimicrobial secondary metabolite, but may be lost in other species if

the cluster is silent. These results support the concept that BGCs can be an important avenue to better understand the natural role of microbial secondary metabolites and their evolution.

4.2. Introduction

Biosynthetic gene clusters (BGCs) are a group of spatially clustered genes that are responsible for the production of secondary metabolites with diverse functions (Bibb, 2005). Microbial BGCs can span from 20- to over 100-kb in size and may contain over 30 genes that encode enzymes and proteins required for the biosynthesis, regulation, transport, self-resistance, and modification of secondary metabolites (Greule et al., 2017). Secondary metabolites are not directly required for the growth and development of the producing organism, but instead, are thought to provide a competitive advantage through nutrient acquisition, inhibition of competitors, or cell communication (Sharrar et al., 2020). These compounds have been extensively studied for clinical and industrial applications due to their diverse biological activities. However, the ecological roles of most microbial secondary metabolites remains poorly understood (O'Brien & Wright, 2011).

Unraveling the natural role of microbial secondary metabolites is a difficult task since the gene clusters that govern their synthesis are often silent or cryptic under laboratory conditions (Kim et al., 2021). Additionally, most environmental microorganisms cannot be readily cultured in the laboratory and current estimates suggest that more than 99% of microbial species are uncultured to date (Locey & Lennon, 2016). To address some of these challenges, several bioinformatics tools have been developed to detect and characterize BGCs in (meta)genome sequences of cultured and uncultured bacteria. These include tools such as antiSMASH (Blin et al., 2019), PRISM (Skinnider et al., 2017), CLUSEAN (Weber et al., 2009), ClustScan (Starcevic et al., 2008), and RiPPMiner (Agrawal et al., 2017). Through these technologies, we have

discovered that BGCs are extensively distributed in various soil, marine, and microbiome environments and many of these gene clusters are uncharacterized (Gavriilidou et al., 2022). As a result, our understanding of the diversity of microbial biosynthetic pathways is rapidly expanding. However, far less progress has been made to understand why microbes produce these compounds (O'Brien & Wright, 2011).

The prevailing theory on the functional role of microbial secondary metabolites is that they provide a competitive advantage to the host since many of these metabolites display antimicrobial properties (O'Brien & Wright, 2011). In fact, most clinically-relevant natural products are derived from soil bacteria and have shown diverse antibacterial, antifungal, and antiparasitic activities (D'Costa et al., 2007). Additionally, some microbial secondary metabolites have been shown to play roles in metal transport and microbial differentiation (Demain & Fang, 2000). These bioactive compounds are believed to important mediators of microbe-microbe and microbe-host interactions, likely with complex coevolutionary histories (Braga et al., 2016). It is clear that many of these bioactive metabolites are beneficial to the producing organism and may provide the selective advantage needed to form and maintain BGCs that encode their biosynthesis (Johnston et al., 2020).

Comparative genomic studies have identified that partial or entire gene clusters can be exchanged between and among prokaryotic species through vertical and horizontal gene transfer (Chase et al., 2021; Fischbach et al., 2008). These clusters can be highly divergent or conserved across different species which is likely influenced by their importance to the organism. Therefore, BGCs can be ideal candidates to study microbial chemical ecology and evolutionary dynamics (Fischbach et al., 2008; Greule et al., 2017). To date, most studies in this field have narrowly focused on the genus *Streptomyces*, largely due to the overwhelming sequence data available from

cultured isolates of this large taxonomic group. However, these studies are still primarily motivated by the discovery of compounds for pharmaceutical interests rather than guided by evolutionary or ecological questions (O'Brien & Wright, 2011). Thus, the current state of knowledge on the function and evolution of BGCs is limited and more studies are needed that seek to understand these relationships across diverse taxa.

The field of metagenomics combined with modern bioinformatics provides an opportunity to explore the biological function of secondary metabolites from uncultured microorganisms (Vecherskii et al., 2021). The application of these techniques have advanced our ability to identify, capture, and express previously "hidden" or "silent" gene clusters (Pereira, 2019). Many unique BGCs have been detected through these culture-independent strategies and led to the discovery of new bioactive secondary metabolites (Hover et al., 2018; Santana-Pereira et al., 2020). The role of these metabolites for the native organism is unknown, but more in-depth characterization of metagenome-derived gene clusters can provide clues on their evolutionary history and distribution. In turn, this knowledge will not only provide a more comprehensive understanding on microbial chemical ecology and their influence on ecosystems (O'Brien & Wright, 2011), but may also guide efforts to find the most promising reservoirs of new therapeutics.

Here, we sought to explore the functional importance and taxonomy of a soil metagenome-derived gene cluster, P12B21 BGC. In a previous study, we identified that this BGC encodes antibacterial secondary metabolites when expressed in the heterologous host, *Streptomyces coelicolor* M1154 (unpublished, Chapter 3 of this dissertation). In this study, we used *in silico* sequence analysis to characterize the genetic elements encoded in the P12B21 BGC and investigate its' taxonomic origin. Next, we set out to determine the functional role of some of the encoded genes by generating targeted P12B21 deletion constructs and screening their expression in *S.*

coelicolor. Preliminary evidence indicated that the P12B21 gene cluster is derived from a *Microbacterium* species. *Microbacterium* are Gram-positive bacteria belonging to the phylum Actinobacteria and are found in various environments including soils and water (X. Li et al., 2021). Other studies have identified that *Microbacterium* can harbor nine secondary metabolite-encoding gene clusters on average per genome (Corretto et al., 2020), but little is known about their functional role. To elucidate the biological and functional importance of the P12B21 gene cluster, efforts were made to explore its conservation in cultured *Microbacterium* species.

4.3. Materials and Methods

4.3.1. Bacterial strains and plasmids.

Bacterial strains and plasmids used in this study are listed in Supplementary Table S4.1. The construction, cloning, and subcloning of *Streptomyces coelicolor* M1154 metagenomic clone P12B21 is described in (Santana-Pereira et al., 2020) and Chapter 3 of this dissertation. BAC-Optimized Replicator v2.0 *Escherichia coli* DH10B (Lucigen Corporation, Middleton, WI) was used for construction and cloning of P12B21 deletion constructs. The P12B21 deletion constructs were subcloned in the pDualP vector (Varigen Biosciences) that contains two inducible promoters flanking the cloning site for expression in *Streptomyces* (Stankey et al., 2022). *E. coli* HB101 (Boyer & Roulland-Dussoix, 1969) bearing the helper plasmid pRK2013 (Figurski & Helinski, 1979) was used to facilitate the intergeneric triparental conjugation of P12B21 deletion constructs from *E. coli* to *S. coelicolor* M1154. For comparative BGC analyses, *Microbacterium hatanonis* DSM 19179^{T} was obtained from DSMZ (German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany). The rest of the 65 *Microbacterium* strains were isolated from inside rhizomes (endophytic bacteria) and are part of an internal collection of Auburn University. Multidrug-resistant *Acinetobacter baumannii* WR3806 (Taitt et

al., 2014), methicillin-sensitive *Staphylococcus aureus* Xen29 (Park et al., 2012), and antibiotic-susceptible *E. coli* WTZ17006 (Baba et al., 2006) were used as tester strains for the bioassay experiments.

4.3.2. Media and culture conditions.

For conjugations, *E. coli* strains were cultured in liquid Luria-Bertani (LB) medium or solid agar LB plates containing the appropriate antibiotics at 37°C with shaking at 180 rpm. Apramycin (50 µg/ml), kanamycin (30 µg/ml), and nalidixic acid (25 µg/ml) were supplemented in the medium when necessary. *E. coli – Streptomyces* conjugations were performed on soya flour-mannitol (SFM) agar [2% (w/v) soybean flour, 2% (w/v) mannitol, and 2% (w/v) agar] containing 20 mM $MgCl_2$. Germination of *S. coelicolor* M1154 spores was performed in 2xYT medium [1.6% (w/v) tryptone, 1% (w/v) yeast extract, 0.5% (w/v) NaCl]. *S. coelicolor* clone P12B21 and its genetic variants were grown on SFM containing 50 µg/ml apramycin and incubation at 30°C for 5 – 7 days for sporulation and colony PCR. For genomic DNA isolation, *S. coelicolor* clone P12B21 was grown in TSB at 30°C with shaking at 180 rpm for 5 days with 3mm glass beads. For heterologous expression studies, liquid TSB with 50 µg/ml apramycin and 3mm glass beads were used for *S. coelicolor* pre-cultures. Modified liquid yeast extract-malt extract (modified YEME) medium [0.3% (w/v) yeast extract, 0.3% (w/v) malt extract, 0.5% (w/v) peptone, 1% (w/v) glucose] was used for heterologous expression of *S. coelicolor* clone P12B21 and deletion strains.

Microbacterium isolates were routinely grown on TSA at 30°C or in liquid TSB with shaking at 180 rpm at 30°C for 48 hr. For bioactivity assays, *Microbacterium* isolates were grown in liquid International *Streptomyces* Project (ISP) 2 agar (Shirling & Gottlieb, 1966) at 30°C with shaking at 180 rpm for 48 hr. Strains used for bioassays including *A. baumannii*

WR3806, *S. aureus* Xen29, and *E. coli* WTZ17006 were cultured in liquid or solid LB medium and incubated at 37°C with shaking at 180 rpm.

4.3.3. Sequence analysis and annotation of the P12B21 gene cluster.

The insert sequence of clone P12B21 from the metagenomic library was sequenced and assembled as previously described (Santana-Pereira et al., 2020). The BGC encoded in the insert sequence was characterized by antiSMASH 5.0 (Blin et al., 2019) and manually inspected by importing the antiSMASH output into the Geneious software suite (Kearse et al., 2012). Annotation of predicted open reading frames (ORFs) and genes/domains from the putative P12B21 BGC as predicted by antiSMASH were visualized using the R package "gggenes" (Wilkins & Kurtz, 2019) and "ggplot2" (Wickham, 2016). The function of putative gene products were predicted by comparing the translated nucleotide sequences against the GenBank database using BLASTx. Similar BGCs determined by the ClusterBlast algorithm of antiSMASH were aligned in the MAUVE program with default settings (Darling et al., 2004).

4.3.4. Construction and subcloning of pDualP:P12B21 constructs.

Deletion constructs of the P12B21 gene cluster were generated by Varigen Biosciences (Madison, WI) using CRISPR-Cas9 restriction. Bioinformatic analysis of the P12B21 gene cluster was used to identify two regions to target for deletion. For the first deletion construct (P12B21_Deletion1), the downstream DNA sequence after ORF31 was targeted for deletion. The downstream DNA sequence after ORF28 was targeted for deletion in the second deletion construct (P12B21_Deletion2). Genomic DNA from *E. coli* DH10B with the BAC-Optimized Replicator harboring the unmodified P12B21 gene cluster was restricted *in vitro* using Cas9 and two guide RNAs which targeted sites upstream and downstream of the selected regions. A Gibson assembly reaction was used to assemble the pDualP vector with the restricted DNA from

each deletion construct. The pDualP vector containing the restricted DNA was transformed to *E. coli* DH10B with the BAC-Optimized Replicator and transformants were selected with apramycin 50 µg/ml. Four constructs were generated: P12B21 (unmodified, positive control), Empty vector (negative control), P12B21_Deletion1, and P12B21_Deletion2.

4.3.5. Conjugal transfer of the pDualP:P12B21 constructs to *S. coelicolor* M1154.

The P12B21 and empty vector constructs were transferred from *E. coli* to *S. coelicolor* M1154 by conjugation. Donor cells of individual *E. coli* DH10B clones containing the subcloned pDualP constructs were prepared by overnight growth at 37°C in 2 ml of LB supplemented with 50 µg/ml apramycin. The helper strain *E. coli* HB101/pRK2013 was prepared by overnight growth at 37°C in 2 ml of LB supplemented with kanamycin (30 µg/ml). *E. coli* donors and helpers were diluted 1:100 in fresh LB with appropriate antibiotics and grown to an OD_{600} of 0.4. Cells were then washed twice in ice-cold LB and resuspended in 1 ml of ice-cold LB. Approximately 10^8 spores of *S. coelicolor* M1154 were added to 0.5 ml of 2xYT medium per conjugation, rinsed once in 2xYT medium, and resuspended in 0.5 ml of 2xYT. Spores were heat-shocked at 59°C for 10 minutes to promote germination. After cooling to room temperature, 100 µl of *E. coli* DH10B donor and 100 µl of *E. coli* HB101/pRK2013 were added to the spore suspension. The mixture was centrifuged at 5,000 x *g* for 5 min, the supernatant was poured off, and the pellet was resuspended in the residual liquid. The mating mixture was plated on MS agar medium containing 20 mM $MgCl_2$ and incubated at 30°C for 24 hr. Plates were overlaid with 1 mL of sterile water containing 0.5 mg of nalidixic acid and 1 mg of apramycin and incubated further at 30°C until exconjugants were visible. Single colonies of triplicate exconjugants per construct were picked from the plates and isolated on SFM agar containing 25 µg/ml nalidixic acid and 50 µg/ml apramycin. *S. coelicolor* exconjugants were then plated on 50 µg/ml

apramycin plates and confirmed by colony PCR using the primers UniversalBACF and UniversalBACR (Supplementary Table S4.2). Validated exconjugants were stored as mycelia in 25% glycerol at -80°C and as spores at -20°C.

4.3.6. Heterologous expression studies of the DualP:P12B21 constructs.

For heterologous expression studies, pre-cultures of triplicate exconjugants from each *S. coelicolor* P12B21 and empty vector construct were prepared by inoculating spores of each exconjugant from glycerol stocks into 5 ml of TSB containing 50 µg/ml apramycin and 0.3 g of 3mm glass beads. Cultures were incubated at 30°C, 180 rpm for 3 – 4 days. Cells were pelleted by centrifugation at 7, 700 x *g* for 10 min, washed twice in TSB, and resuspended in 5 ml of modified YEME medium. A 1% inoculum (100 µl) of washed pre-cultures was added to 10 ml of modified YEME. For induction of pDualP promoters, cultures were supplemented with the following inducer treatments at the start of incubation: 1) no inducers, 2) 0.1% (w/v) Ɛ-caprolactam (ƐCL) inducer, 3) 2.5 µM oxytetracycline dihydrate (OTC) inducer, or 4) 0.1% ƐCL and 2.5 µM OTC inducers. All cultures were incubated at 30°C for 7 days and shaking at 180 rpm. Cell-free supernatants were harvested from the cultures by centrifugation at 7,700 x *g* for 10 minutes and the resulting supernatants were filtered through an 0.2-µm filter (VWR).

Cell-free supernatants were assayed for bioactivity against *A. baumannii* WR3806 by a 96-well microplate assay. *A. baumannii* WR3806 was prepared by inoculating a single colony in 2 ml of LB and shaken overnight at 37°C, 200 rpm. The overnight culture was diluted 1:100 in fresh LB, grown to an optical density at 600 nm (OD_{600}) of ~ 0.5 – 0.6, and diluted 1:100 in LB. A volume of 100 µl of each cell-free supernatant was added to triplicate wells of sterile 96-well microplates. 100 µl of the diluted log-phase bacterial test strains were added and mixed with the supernatants. Control wells were included with 100 µl of the bacterial test strain mixed with 100

µl of modified YEME medium. The OD_{600} of each well was determined in a microplate reader before incubation (T_0) and following incubation at 37°C for 16 hours with shaking at 200 rpm (T_{16}). Growth inhibition, expressed as a percent inhibition of the test strain wells relative to control wells, was determined using the formula:

$$\text{Inhibition (\%)} = \left(1 - \left(\frac{OD_{600} \text{ test well at } T_{16} - OD_{600} \text{ test well at } T_0}{OD_{600} \text{ control well at } T_{16} - OD_{600} \text{ control well at } T_0}\right)\right) \times 100$$

Percent inhibitions for each triplicate well were averaged per *S. coelicolor* pDualP construct prior to statistical analyses to avoid pseudoreplication.

Statistical analyses of the bioactivity data were performed in R (R_Core_Team 2019) using the following packages: 'lme4' (Bates et al., 2014), 'lmerTest' (Kuznetsova et al., 2017b), 'emmeans' (Lenth et al., 2018), 'betareg' (Cribari-Neto & Zeileis, 2010). Data wrangling was performed in R with the help of the following packages: 'tidyverse' (Wickham, 2017b), 'dplyr' (Wickham et al., 2015), 'reshape2' (Wickham, 2007b) and a bar graph was produced using the following packages: 'ggpubr' (Kassambara, 2020), 'cowplot' (Wilke, 2019), and 'RColorBrewer' (Erich, 2014). Pairwise comparisons were analyzed between the percent inhibition associated with pDualP:P12B21 and pDualP:EV constructs. Pairwise comparisons were corrected by using the default "Tukey method" in the 'emmeans' package to correct for Type 1 error. Statistically significant differences between treatments were established at $\alpha = 0.05$.

4.3.7. Colony PCR-based screening of *Microbacterium* isolates.

Primers were designed to amplify the conserved A3 motif within NRPS adenylation domain of the P12B21 gene cluster which was annotated by antiSMASH. Five primer pairs were generated using NCBI Primer-BLAST and checked for specificity to the *S. coelicolor* genome

(Supplementary Table S4.2). Primers were synthesized by Eurofins Genomics LLC (Louisville, KY). The primer pairs were first tested for their functionality with *S. coelicolor* clone P12B21 by PCR. For this, SFM agar plates containing 50 µg/ml were streaked with the strain and incubated at 30°C for 4 days. A single colony was picked and used as a template for PCR reactions. Each reaction consisted of 25 µl of CloneID 1X Master Mix (Lucigen Corporation, Middleton, WI), 0.25 µL of each primer at 50 µM, and a single colony. The amplification was conducted with a touchdown PCR protocol, with denaturation at 98°C for 2-min, followed by 10 cycles of a touchdown PCR (98°C for 15 s, annealing beginning at 72°C and ramping down 1°C per cycle to 62°C for 30 s, and 72°C for 30 s), and then 20 cycles of amplification (98°C for 15 s, 62°C for 30 s, and 72°C for 30 s), and final elongation at 72°C for 5 min. PCR products were visualized through gel electrophoresis. The primer pair yielding the desired amplicon was then used to screen the *Microbacterium* isolates listed in Supplementary Table S4.1 by colony-PCR. Each *Microbacterium* isolate was prepared by streaking on TSA and incubating at 30°C for 48 hr. A single colony of each isolate was used for PCR following the protocol described above. Reactions were considered positive if a ~511 bp amplicon was visible upon agarose gel electrophoresis.

4.3.8. Dot blot hybridization analyses.

To prepare biomass of *S. coelicolor* clone P12B21, the culture was grown in 5 ml of TSB with 3mm glass beads and incubated at 30°C with shaking at 180 rpm for 5 days. Biomass of *Microbacterium* isolates was prepared by culturing each isolate in 2 ml of TSB and incubation at 30°C with shaking at 180 rpm for 48 hr. Genomic DNA was isolated from all strains using the E.Z.N.A® Bacterial DNA Kit (Omega Bio-Tek, Doraville, CA) following the manufacturer's instructions with slight modifications for difficult-to-lyse bacteria. Samples were left for 1 hr for

the first incubation with lysozyme and 2 hr for incubation with proteinase K. The DNA of clone P12B21 was used as a template in a PCR to generate a probe for downstream dot blot screenings. The primer pair P12B21_core_P1_F and P12B21_core_P1_R was used for probe synthesis (Supplementary Table S4.2). The amplification was conducted with a 2-min denaturation at 98°C, followed by 30 cycles of amplification (98°C for 15 s, 62°C for 30 s, and 72°C for 30 s) and final elongation at 72°C for 5 min. The resultant amplicon was labeled with digoxigenin (DIG) using the PCR DIG probe synthesis kit (Roche Applied Science, Indianapolis, IN) and the probe was used for dot blot analyses. Genomic DNA of each isolate was diluted to a concentration of 10 ng/µl and 10 µl was spotted in duplicate on positively-charged nylon Zeta-Probe membranes (Bio-Rad, Hercules, CA). The DNA was denatured by placing the membrane (DNA side facing up) on a Whatman 3MM filter paper for 10 min. The membrane was then transferred on top of a filter paper saturated in 2X SSC buffer to rinse and left for 5 min. The membrane was left to air dry completely and stored between two pieces of 3MM filter paper at room temperature prior to dot blot assays. Dot blot assays were performed following the manufacturer's recommendations (Science, 2009) with hybridization overnight at 52°C and the recommended stringency washes for 80 – 100% homology to the target and high GC-content. Detection was performed using the DIG nucleic acid detection kit (Roche Applied Science, Indianapolis, IN) by colorimetric reaction with NBT/BCIP according to the manufacturer's instructions.

4.3.9. Biological activity assays of *Microbacterium* isolates.

Target organisms for antimicrobial activity assays of *Microbacterium* isolates included *S. aureus* Xen29, *E. coli* WTZ17006, and *A. baumannii* WR3806 (Supplementary Table S4.1). The antibacterial activity of each *Microbacterium* isolate was evaluated using a spot-on-lawn assay

with cell-free supernatants. The isolates were cultured in 2 mL of ISP 2 broth and incubated for 48 hr at 30°C with 180 rpm shaking. After incubation, the cells were pelleted by centrifugation at 5,000 x *g*. Supernatants were filtered through a 0.2 µm microporous membrane, and 10 µl of cell-free supernatant was spotted on the surface of agar plates containing the target organisms in triplicate. A 10 µl spot of sterile ISP 2 broth was used as a negative control. Plates containing cell-free supernatants from *S. coelicolor* clone P12B21 were also included as a control. After spots were dried in a clean bench, plates were incubated at 37°C for 24 hr. The diameters of the zones of inhibition were subsequently measured.

4.4. Results

4.4.1. Bioinformatic analysis of the P12B21 biosynthetic gene cluster.

Sequence analysis by antiSMASH and BLASTx determined that metagenomic clone P12B21 contained a 60,007 bp insert sequence consisting of 62 ORFs encoding a non-ribosomal peptide synthetase (NRPS)-like gene cluster, putative sugar-related enzymes, aromatases/cyclases, transcriptional regulators, transporters/permeases, resistance factors, and hypothetical proteins as shown in Fig. 4.1a. Manual inspection of the annotation provided by antiSMASH revealed that the NRPS-like gene cluster spanned 26,196 bp and was encoded within the first 28 ORFs of the entire 60-kb insert sequence (Fig. 4.1b). Functional predictions of genes encoded in the P12B21 BGC is listed in Supplementary Table S4.3.

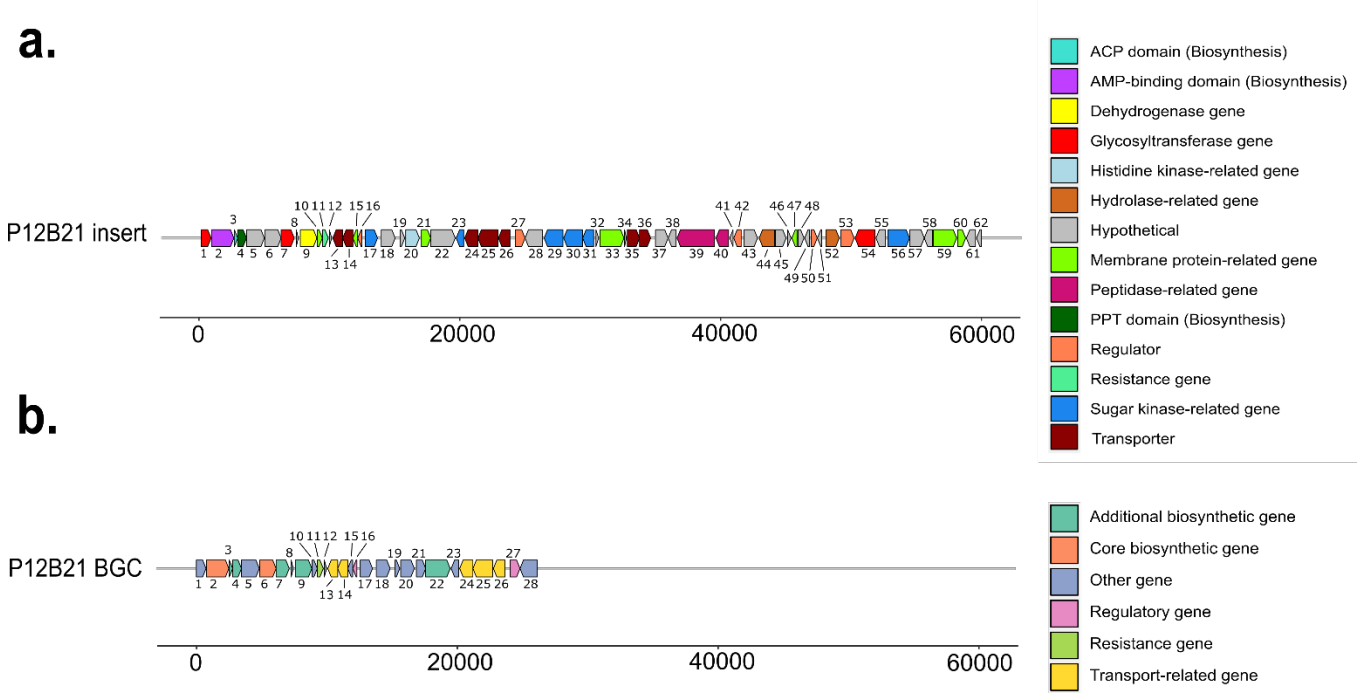

Figure 4.1. Annotation of the genes encoded in metagenomic clone P12B21. (a) Functional prediction of genes within the 60-kb insert sequence of clone P12B21 as determined by BLASTp analysis. (b) AntiSMASH annotation of the 20-kb BGC encoded in clone P12B21. Colors represent putative broad gene functions as shown in the legends.

A typical NRPS cluster consists of three catalytic domains: adenylation (A) domain, thiolation (T) domain, and condensation (C) domain (Hur et al., 2012). Of these three core domains, only a partial A domain was annotated within the P12B21 BGC. Within the A domain, a conserved motif designated as A3 was identified that comprises the ANL enzyme superfamily of NRPSs (Miller et al., 2014). Several accessory domains were encoded in the core BGC region including a phosphopantetheinyl transferase (PPT) domain (ORF4), an acyl carrier protein (ACP) domain (ORF3), and an adenosine monophosphate (AMP)-binding domain (ORF2) (Supplementary Table S4.3). PPT domains are frequently observed in NRPS pathways where they play a role in the post-translational modification of the T domain (Lundy et al., 2020). On the other hand, ACP and AMP-binding domains function to activate and incorporate acyl substrates to the elongating protein in PKS pathways (H. Chen et al., 2015), but have also been

identified in hybrid NRPS-PKS pathways (Fisch, 2013). According to these findings, P12B21 BGC may represent a hybrid NRPS-PKS pathway with an unusual architecture.

Other notable features within the P12B21 gene cluster included a PadR transcription regulator (ORF16) and a GntR family transcription regulator (ORF27) (Supplementary Table S4.3). Additionally, a resistance gene was identified in ORF11 with similarity to a *vanZ* gene (Fig. 4.1; Supplementary Table S4.3). *VanZ* has been identified to confer resistance to lipoglycopeptide antibiotics in other bacteria (Vimberg et al., 2020) and may be responsible for conferring self-resistance to the native producer and the *S. coelicolor* M1154 heterologous host.

Manual inspection of the genes flanking the putative BGC boundaries identified several enzymes predicted to be involved in sugar modification and glycosylation (Fig. 4.1a). These included three ORFs encoding sugar kinases, three ORFs encoding glycosyltransferases, one ORF encoding a nucleotide sugar dehydrogenase, three ORFs encoding sugar transporters, and three ORFs encoding glycoside hydrolases. The encoded glycosyltransferases are presumed to be involved in the attachment of sugars, but it was unknown if they act upon the secondary metabolite produced by clone P12B21. Putative enzymes involved in saccharide and triacylglycerol syntheses were also identified. As shown in Fig. 4.1a, many of the genes located between ORF29 to ORF62 were assigned as hypothetical proteins with unknown function.

4.4.2. Heterologous expression of P12B21 gene deletion constructs.

To further elucidate the functional role of some of the encoded genes in the P12B21 cluster, we generated two gene deletion constructs and screened their expression in *S. coelicolor* by bioactivity assays. The first deletion construct (P12B21_Deletion1) was designed to keep the three ORFs (ORF29, ORF30, ORF31) that directly flanked the right edge of the BGC and were predicted to encode sugar kinases that we presumed play a role in glycosylation of the antibiotic

molecule. The remaining genes (ORF32 to ORF62) were deleted in this construct. For the second deletion construct (P12B21_Deletion2), the entire DNA sequence flanking the right border of the BGC was deleted (ORF29 to ORF62). A schematic representation of the P12B21 constructs generated in this study is shown in Figure 4.2.

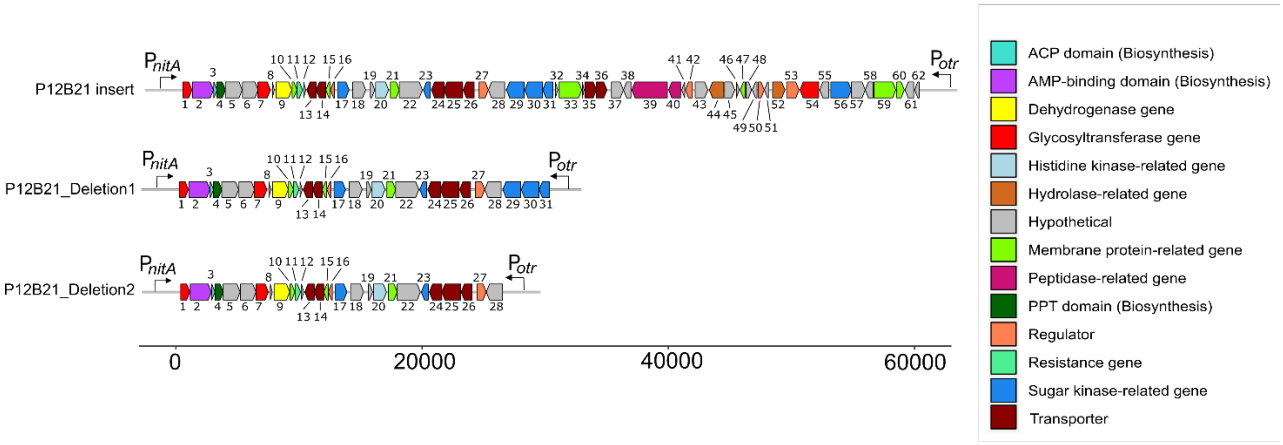

Figure 4.2. Genetic organization and annotation of the P12B21 insert sequence and schematic representation of deletion constructs generated in this study.

Deletion constructs were subcloned in the pDualP vector which contains the inducible promoters P_{nitA} and P_{otr} that flank the left and right edge of the cloned sequence, respectively (Stankey et al., 2022). Our previous experiments identified that the expression of an unmodified P12B21 construct could be enhanced by induction with 0.1% (w/v) ƐCL which acts on the P_{nitA} promoter. Therefore, in the present experiments, we also wanted to determine if bringing the P_{otr} promoter closer to the right edge of the BGC could further alter expression of this cluster by *S. coelicolor*. Resultant deletion constructs were subsequently introduced into *S. coelicolor* M1154 by conjugation from *E. coli* DH10B and triplicate exconjugants treated with ƐCL and/or OTC inducers were screened for antimicrobial activity against *A. baumannii*. Percent inhibitions of

each deletion construct were compared to the positive control of the unmodified P12B21 construct and the empty vector negative control.

No significant inhibition of *A. baumannii* was observed from clones containing the P12B21_Deletion2 constructs with any inducer treatment (Fig. 4.3). The average inhibition of *A. baumannii* from each exconjugant with the P12B21_Deletion2 construct was < 20% across all inducer treatments. In comparison, activity was highly variable between exconjugants of clones harboring the P12B21_Deletion1 constructs (Fig. 4.3a). Two exconjugants with the P12B21_Deletion1 construct consistently exhibited < 20% inhibition of the target pathogen across all treatments, whereas one exconjugant showed much greater antimicrobial activity (> 50% inhibition) when treated with the ɛCL inducer, ɛCL and OTC inducers combined, or no inducers (Fig. 4.3b). With the exception of treatment with OTC alone, the inhibitory activity of this exconjugant was significantly higher than the activity from any exconjugant with an empty vector, unmodified P12B21, or P12B21_Deletion2 construct.

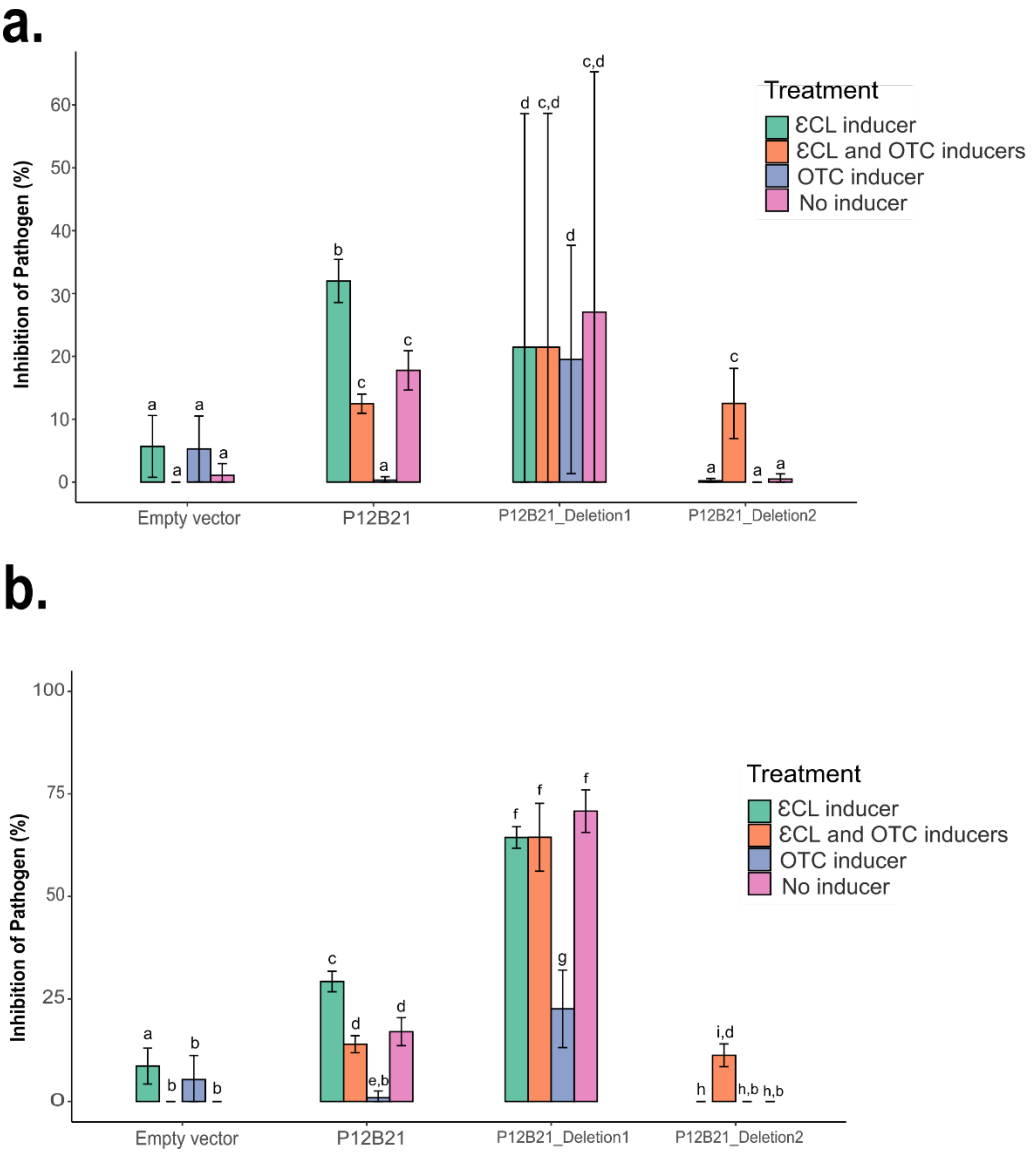

Figure 4.3. Antibacterial activity of *S. coelicolor* exconjugants harboring modified and unmodified P12B21 gene clusters. (a) Percent inhibition of *A. baumannii* WR3806 by cell-free supernatants of triplicate *S. coelicolor* exconjugants carrying modified and unmodified P12B21 constructs. Constructs were cloned in the pDualP inducible expression system and exconjugants were treated with or without the inducers 0.1% (w/v) ECL and 2.5 µM OTC. Bars represent the mean inhibition from three exconjugants. (b) Percent inhibition of *A. baumannii* WR380 from cell-free supernatants of a single exconjugant of *S. coelicolor* carrying modified and unmodified P12B21 constructs. Bars represent the mean inhibition of a single exconjugant and three technical replicates. For both plots, black lines indicate the standard error among technical replicates. Letters represent significant differences ($p < 0.05$) between the percent inhibitions of the pathogen associated with each treatment for each construct as determined by beta regression with Tukey's post-hoc correction for multiple comparisons.

4.4.3. Comparison of the P12B21 gene cluster with other *Microbacterium*-derived clusters.

Our next studies aimed to elucidate the taxonomic origin of the BGC and determine if this gene cluster was conserved in other species. AntiSMASH analysis revealed that 96% of the genes within the P12B21 BGC were similar to genes within a NRPS cluster (antiSMASH accession#: VRSV01000002.1) from *M. hatanonis* JCM 14558^T. Furthermore, each of the annotated ORFs within the cluster had closest GenBank matches to proteins from *M. hatanonis* (Supplementary Table S4.3). *M. hatanonis* JCM 14558^T is a bacterium of the phylum Actinobacteria and was originally isolated as a contaminant of hair spray (Bakir et al., 2008). Although other *Microbacterium* strains have been observed to produce bioactive metabolites (Bano et al., 2022; Kanagasabhapathy et al., 2008), antimicrobial activity has not been reported from *M. hatanonis* and a secondary metabolite has not been linked to this particular NRPS cluster encoded in its genome.

In addition to *M. hatanonis*, the most similar clusters to P12B21 as determined by the ClusterBlast algorithm of antiSMASH were also derived from *Microbacterium* species including *M. pygmaeum* DSM 23142^T, *M. yannicii* PS01^T, and *M. sediminis* YLB-01^T. However, these clusters were less similar (ranged from 20% to 51% similarity) to P12B21 compared to the NRPS cluster of *M. hatanonis*. The gene cluster sequences from each of these *Microbacterium* strains were further compared to the P12B21 NRPS cluster using a collinearity analysis performed by MAUVE. Results were consistent with antiSMASH and showed homologous and non-homologous regions between each BGC with the most similarity between P12B21 and *M. hatanonis* (Fig. 4.4). Thus, the synteny of the P12B21 gene cluster with other *Microbacterium*-derived BGCs suggests that the cluster is derived from a *Microbacterium* species.

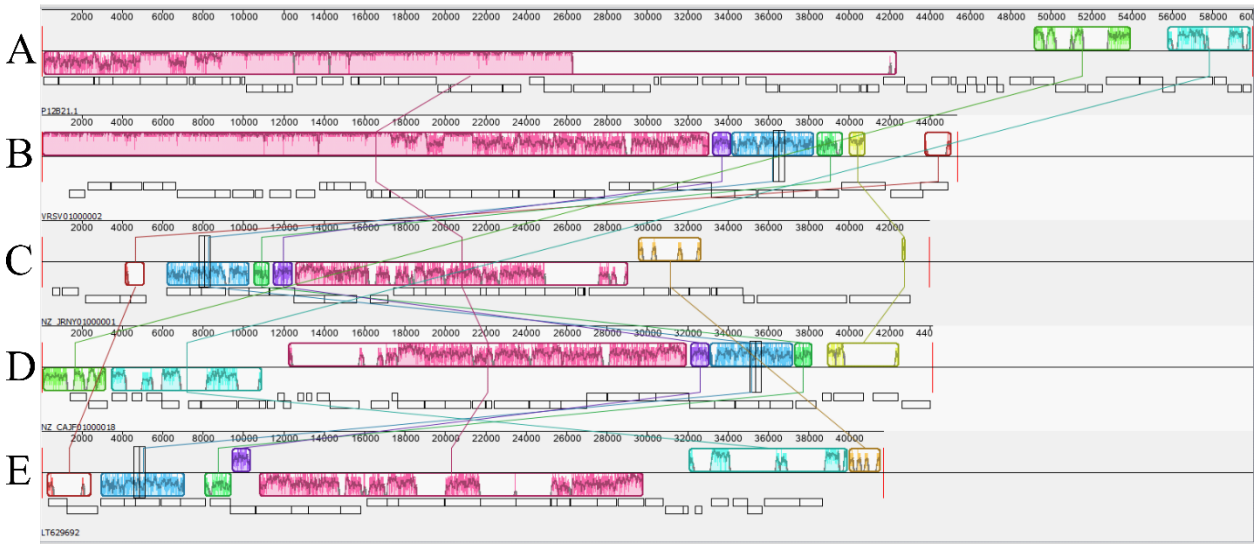

Figure 4.4. MAUVE alignment of the biosynthetic gene clusters of *S. coelicolor* clone P12B21 (A), *M. hatanonis* JCM 14558^{T} (B), *M. sediminis* YLB-01^{T} (C), *M. yannicii* PS01^{T} (D), and *M. pygmaeum* DSM 23142^{T} (E). Each colored block is a region of collinear sequence representing homologous regions among each strain. Where colors match among gene clusters, homology is detected. Boxes under each sequence map represent protein-coding genes (white boxes).

These findings prompted us to investigate if we could identify homologous BGCs to P12B21 from *M. hatanonis* and other soil *Microbacterium* isolates using *in vitro* techniques and if these isolates exhibited a similar antibacterial activity profile to *S. coelicolor* clone P12B21. For the analyses described in the following sections, we screened the type strain *M. hatanonis* DSM 19179^{T} and 65 cultured *Microbacterium* strains that were isolated from inside rhizomes (endophytic bacteria) and are part of the internal collection of Auburn University. Genome sequences were not available from the 65 *Microbacterium* isolates which led us to use PCR and dot blot analyses to screen for homologous gene clusters. Although the exact identity of these endophytic *Microbacterium* isolates is not known, BLASTn of 16S rRNA gene sequences revealed that many of these isolates show > 98% similarity to the species *M. azadirachtae*, *M. foliorum*, *M. arborescens*, *M. testaceum*, *M. oleivorans*, *M. oxydans*, or *M. hydrothermale*. We

were not able to acquire *M. pygmaeum*, *M. yannicii*, or *M. sediminis*, and therefore, these strains could not be included in these analyses.

4.4.4. Antibacterial activity screening of *Microbacterium* isolates.

Antibacterial activity assays of each *Microbacterium* isolate and *S. coelicolor* clone P12B21 were performed against target organisms (*A. baumannii* WR3806, *E. coli* WTZ17006, and *S. aureus* Xen29) using a spot-on-lawn technique. These target strains were selected since *S. coelicolor* clone P12B21 is known to inhibit the growth of *A. baumannii* WR3806 and *E. coli* WTZ17006 based on our previous studies. We also wanted to include a Gram-positive strain (*S. aureus* Xen29) to screen for any differences in bioactivity displayed between the *Microbacterium* isolates and clone P12B21. Other studies have identified production of bioactive secondary metabolites from *Microbacterium* isolates when ISP 2 broth was used as the growth medium (Abdelmohsen et al., 2010; Bano et al., 2022). For this reason, we chose to culture each of the *Microbacterium* isolates in ISP 2 broth in an attempt to promote secondary metabolism. Additionally, these isolates were not observed to grow well in the modified YEME that is used for fermentation of clone P12B21. Bioactivity was not observed from any of the *Microbacterium* isolates against any of the target strains when cultured in ISP 2 broth for 48 hr. Consistent with our prior findings, supernatants from the control *S. coelicolor* clone P12B21 were inhibitory to the growth of *A. baumannii* WR3806 and *E. coli* WTZ17006, but not *S. aureus* Xen29.

4.4.5. Design of P12B21 gene cluster-specific primers.

Following inconclusive results from bioactivity assays, we decided to screen the *Microbacterium* isolates for a homologous BGC to P12B21 by PCR and dot blot hybridization. The design of PCR primers specific for the P12B21 gene cluster was based on the annotation by antiSMASH (Fig. 4.1). Five PCR primer pairs were designed so that they amplified a ~200- to

600-bp region of the P12B21 core adenylation domain (Supplementary Table S4.2). More specifically, the primers targeted the conserved A3 motif within the predicted A domain. Motif A3 is part of the ten motifs (A1 – A10) that comprise the ANL enzyme superfamily and are conserved across NRPS clusters (Miller et al., 2014).

The primer pairs were first tested for their functionality with *S. coelicolor* clone P12B21 by colony PCR using a touchdown PCR protocol. Amplification products of the expected size ranges were obtained from three of the primer pairs which included P12B21_core_P1, P12B21_core_P4, and P12B21_core_P5 (Fig. 4.5). The P12B21_core_P1 bands were selected for downstream PCR and dot blot analyses since the intensity of the bands was greater than the PCR products from the other primer pairs.

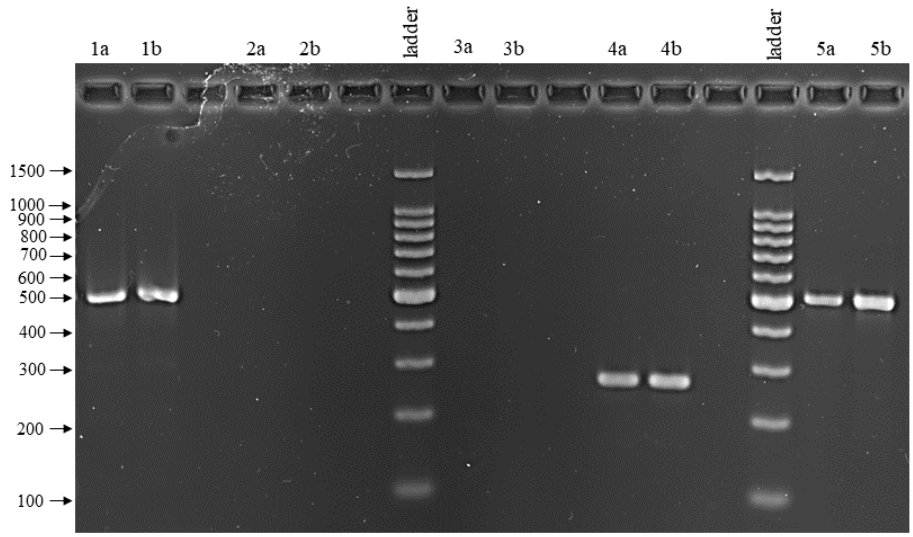

Figure 4.5. Agarose gel electrophoresis of PCR products from colony PCR screening of *S. coelicolor* clone P12B21. Selective amplification of ~200- to 600-bp fragments using primers specific for the conserved A3 motif of the adenylation domain in the P12B21 biosynthetic gene cluster. The desired amplicon was obtained for three of the five primer pairs tested. Lanes, 1a and 1b: *S. coelicolor* clone P12B21 with P12B21_core_P1_F/R primers; 2a and 2b: *S. coelicolor* clone P12B21 with P12B21_core_P2_F/R primers; 3a and 3b: *S. coelicolor* clone P12B21 with P12B21_core_P3_F/R primers; 4a and 4b: *S. coelicolor* clone P12B21 with P12B21_core_P4_F/R primers; 5a and 5b: *S. coelicolor* clone P12B21 with P12B21_core_P5_F/R primers.

4.4.6. PCR-based screening of cultured *Microbacterium* for a homologous BGC to P12B21.

The PCR primer set P12B21_core_P1_F and P12B21_core_P1_R was used as a molecular probe to screen *M. hatanonis* DSM 19179^T and the cultured *Microbacterium* isolates for the presence of a homologous P12B21 BGC. Single colonies of each isolate, including *S. coelicolor* P12B21 as a control, were used as a template for colony PCR. PCRs resulted in successful amplification of the expected ~511 bp fragment from *S. coelicolor* clone P12B21 and *M. hatanonis* DSM 19179^T further supporting the data from antiSMASH analysis. Of the other 65 cultured *Microbacterium* isolates that were screened, none yielded a PCR product suggesting that these bacteria did not contain a homologous BGC to clone P12B21 or their respective A domain sequences are too divergent to be amplified with the primers.

4.4.7. Dot blot analysis of cultured *Microbacterium* isolates.

In addition to PCR screening, genomic DNA from each *Microbacterium* isolate was also screened by a less-sensitive approach, dot blot hybridization, for homologous BGCs to the gene cluster of clone P12B21. The amplicon from colony PCR of *S. coelicolor* clone P12B21 with the BGC-specific primer set P12B21_core_P1_F/R was labeled with DIG and used as a probe for dot blot hybridization. The probe was first validated to hybridize with genomic DNA from *S. coelicolor* clone P12B21, *E. coli* clone P12B21, and *M. hatanonis* DSM 19179^T which all served as positive controls. Additionally, DNA from *S. coelicolor* M1154 was used as a negative control to monitor the sensitivity of the probe. As desired, the DIG-labeled probe hybridized with DNA from each positive control and did not hybridize with the DNA from the negative control (Supplementary Fig. S4.1). After validating the probe, DNA from the 65 *Microbacterium* isolates was screened following the same hybridization conditions. Consistent with the results

from PCR screening, none of the *Microbacterium* isolates besides *M. hatanonis* DSM 19179^T exhibited a colorimetric reaction after hybridization with the DIG-labeled P12B21-specific probe, indicating that these isolates did not possess a homologous gene cluster.

4.5. Discussion

Advancements in next-generation sequencing and (meta)genome mining efforts have demonstrated that nature holds a wealth of secondary metabolite chemical diversity which is largely undiscovered (Gavriilidou et al., 2022). New techniques are rapidly developed to access these compounds for medical research and have greatly expanded our understanding on the mechanisms that control their biosynthesis in microorganisms. Contrarily, the ecological role of microbial secondary metabolites remains poorly understood, but is important to enhance our knowledge on microbial communities and their influence on ecosystems (O'Brien & Wright, 2011). Although it is difficult to define the precise role of such diverse compounds, the application of culture-independent techniques (i.e. metagenomics) have made it possible to explore the distribution and function of BGCs that encode the synthesis of secondary metabolites from uncultured microorganisms (Vecherskii et al., 2021). This information can be useful to understanding the importance of these biosynthetic pathways and their chemical products to the producing microbes.

In this work, we have characterized a secondary metabolite gene cluster derived from a soil metagenomic library in order to predict its function and taxonomic origin. We specifically focused on the P12B21 BGC, since our previous work showed that this gene cluster could be expressed in a *S. coelicolor* heterologous host which facilitated further downstream analyses. Through bioinformatics analysis, we identified that the P12B21 cluster spans 26-kb and is harbored in the first 28 ORFs of the entire 60-kb cloned DNA insert. These ORFs encode several

putative functions required for secondary metabolite biosynthesis, transcriptional regulation, resistance, and transportation.

Initial annotation by antiSMASH identified the P12B21 cluster as an "NRPS-like" BGC. NRPSs are a well-defined class of gene clusters capable of producing secondary metabolites with diverse bioactivities. A minimal NRPS module consists of an A domain, C domain, T domain, and PCP domain (Miller & Gulick, 2016). Interestingly, the P12B21 BGC was predicted to contain only a partial A domain but lacks the other typical NRPS domains. Instead, an ACP domain and AMP-binding domain were identified in the cluster which are two domains commonly associated with polyketide synthases (PKSs), another important class of gene clusters frequently observed in bacterial and fungal genomes (Walsh, 2004). The BGC was not predicted to contain the other catalytic modules common in minimal PKS pathways.

Previous studies have provided evidence for hybrid NRPS-PKS systems encoded in bacterial genomes that contain catalytic enzymes from both systems (Mizuno et al., 2013). These hybrid NRPS-PKS gene clusters can synthesize a compound by direct or indirect interactions between the NRPS and PKS modules, and these pathways have been linked to the production of bioactive metabolites (L Du & Shen, 2001). Based on this evidence, we propose that the P12B21 gene cluster is a hybrid NRPS-PKS pathway. The regulation of this pathway may be controlled by homologous PadR and/or GntR transcriptional regulators which were identified in ORF16 and ORF27, respectively, of the cluster. PadR is a global transcriptional regulator that can function as an environmental sensor and contributes to antibiotic biosynthesis, multidrug resistance, and detoxification in diverse bacteria (Florez et al., 2015). GntR regulators can also act as environmental sensors and have been observed to function in the control of antibiotic synthesis by *Streptomyces* and other bacteria (Hillerich & Westpheling, 2006). The role of these genes in

the P12B21 BGC is unknown at this time, but we predict that these putative regulators take part in biosynthesis control by the native producer and possibly the heterologous host.

To better understand the function of the P12B21 BGC, we attempted to determine the type of compound synthesized by its pathway. The unusual architecture impeded our ability to predict a chemical structure through antiSMASH. However, annotation of the 60-kb cloned DNA insert sequence suggested that P12B21 may be encoding the synthesis of a glycosylated compound due to the presence of several ORFs putatively encoding glycosylation and sugar-modifying enzymes. In support of this prediction, we found that deleting all genes downstream of the BGC boundary (ORF29 to ORF62), several of which were predicted to be involved in glycosylation and sugar modification, resulted in a loss of bioactivity. Furthermore, we identified a resistance gene carried in the BGC with homology to a *vanZ* gene. *VanZ* is a member of the *vanA* glycopeptide resistance cluster that confers resistance to the glycopeptide antibiotics vancomycin and teicoplanin (Vimberg et al., 2020). The *vanZ* homolog identified in the P12B21 cluster may be responsible for providing resistance to the host and provides further evidence that the P12B21 gene cluster is encoding a glycopeptide-like secondary metabolite. Future studies could employ chemical analysis by LC-MS and NMR to accurately determine the compound structure.

We further characterized the functional role of some of the genes encoded in the P12B21 cluster by deleting three genes putatively acting as sugar kinases. Screening the expression of this modified cluster in *S. coelicolor* gave inconclusive results. Only one *S. coelicolor* exconjugant with this modified cluster showed significant antibacterial activity whereas the other two exconjugants did not display any detectable activity. Other studies have observed phenotypic and genotypic heterogeneity between *Streptomyces* colonies that form during growth (Z. Zhang,

Du, et al., 2020). We believe a similar phenomenon may be affecting the degree of BGC expression between *S. coelicolor* exconjugants, and further work will aim to assess this hypothesis.

By characterizing genes encoded in the P12B21 gene cluster, we were also able to predict its taxonomic origin. The closest homologous clusters were derived from *Microbacterium* species and MAUVE alignment revealed syntenic regions between the P12B21 BGC and clusters from *M. hatanonis* JCM 14558^T, *M. pygmaeum* DSM 23142^T, *M. yannicii* PS01^T, and *M. sediminis* YLB-01^T. Based on this evidence, we postulated that the P12B21 gene cluster is derived from a *Microbacterium* species. *Microbacterium* are Gram-positive bacteria within the Actinobacteria phylum that are ubiquitous in many soil and aquatic environments (X. Li et al., 2021). Studies have shown that these bacteria can harbor diverse BGCs in their genomes, but the average number of gene clusters is fairly low compared to other genera in the same phylum (Corretto et al., 2020). Many of the characterized gene clusters from *Microbacterium* species are thought to produce metabolites that function in metal homeostasis since these bacteria are often found in heavy metal-contaminated sites (Brown et al., 2012; Corretto et al., 2020). Less is known about other functions of these gene clusters, but marine *Microbacterium* have been found to synthesize antimicrobial or cytotoxic metabolites (Graça et al., 2013; Kanagasabhapathy et al., 2008). Our study provides evidence that soil *Microbacterium* may also possess gene clusters that function to synthesize antimicrobial compounds.

Although it is difficult to demonstrate the role of the P12B21 BGC since it was derived from an uncultured bacterium, we predict that the secondary metabolites encoded in its pathway provide a competitive advantage to the presumed host *Microbacterium* species. Secondary metabolites with antimicrobial properties are widely thought to play significant roles in microbial

warfare where they can inhibit the growth of competitors that occupy the same ecological niche (O'Brien & Wright, 2011). One example is that of plant growth-promoting bacteria that can produce antimicrobial compounds to suppress the growth of bacterial plant pathogens (Tyc et al., 2017). Consequently, these bioactive metabolites can be key drivers of microbe-microbe and microbe-plant ecological interactions in soil (Braga et al., 2016). Based on this knowledge, we predicted that the BGC serves a beneficial function to the genus *Microbacterium*, and therefore, may be conserved in other soil *Microbacterium* species.

In other members of the Actinobacteria phylum, such as *Salinispora* and *Amycolatopsis*, a strong relationship has been shown between phylogeny and BGC diversity, with most BGCs being shared between close relatives (Adamek et al., 2018; Ziemert et al., 2014). Horizontal gene transfer of entire BGCs has been documented in Actinobacteria, but it is argued that these events are much more rare compared to vertical inheritance (McDonald & Currie, 2017; van Bergeijk et al., 2020). Therefore, it is probable that the P12B21 cluster, or a variation of it, may have been present in a common ancestor of the genus *Microbacterium* and has been maintained in geographically distant and related organisms through vertical inheritance. To test this theory, we used PCR and dot blot hybridizations to screen *M. hatanonis* JCM 14558^T and 65 endophytic *Microbacterium* species for a homologous P12B21 gene cluster.

Through PCR and dot blot analyses, we found further evidence that the P12B21 BGC is homologous to a cluster from *M. hatanonis*. Interestingly, *M. hatanonis* JCM 14558^T was originally isolated as a contaminant of hair spray from Japan (Bakir et al., 2008). We speculate that *M. hatanonis* is actually a soil-borne microorganism that has contaminated a cosmetic product during the manufacturing process. Due to the high homology between the P12B21 BGC and the gene cluster from *M. hatanonis*, we predicted that *M. hatanonis* would display a similar

antimicrobial activity profile to the P12B21 clone. However, we did not observe bioactivity from *M. hatanonis* against any of the tester strains. The lack of bioactivity could mean that the homologous cluster is silent in *M. hatanonis* under the given laboratory conditions or the expression of the BGC was too low for detection. Future studies could seek to identify the growth conditions necessary to "turn on" expression of this BGC in *M. hatanonis* or use chemical analysis to determine if these two strains produce a similar metabolite.

Surprisingly, we were not able to identify a homologous P12B21 BGC to any of the 65 endophytic *Microbacterium* isolates when using hybridization conditions designed for 80 – 100% homology. It is possible that these *Microbacterium* species had portions of the P12B21 BGC, but the homology was too low to be detected by our approach. Genome sequencing of these isolates could help address that possibility. Alternatively, these isolates may not possess this cluster and there are several plausible explanations that may be contributing to the patchy distribution of this BGC across the *Microbacterium* genus.

In Actinobacteria taxa, selective pressure is thought to influence the localization of gene clusters to regions on the chromosome that are more likely to be maintained (van Bergeijk et al., 2020; Ventura et al., 2007). In the absence of that selective pressure, these BGCs may localize away from the core genome to genomic islands which are more susceptible to gain/loss events through horizontal gene transfer (Letzel et al., 2017; van Bergeijk et al., 2020). We suspect that the selective pressure to maintain the P12B21 BGC is influenced by interspecific competition in a niche. Depending on the surrounding microbial community, maintaining and expressing a BGC may be favorable for a particular species' survival or be too costly if not needed. Therefore, we predict that the P12B21 BGC is conserved in its native host and *M. hatanonis* because it confers an advantage to them. The other *Microbacterium* species in this study may have lost significant

portions or all of the gene cluster because its function is not needed or has been replaced by another gene cluster, the latter of which has been documented in other Actinobacteria species (Bruns et al., 2018). Nevertheless, a comprehensive BGC phylogenetics study across more *Microbacterium* representatives is needed to better understand the evolution of the P12B21 gene cluster and its importance to this taxa.

In summary, our study sheds light on the function and taxonomy of a metagenome-derived gene cluster. The P12B21 BGC appears to have originated from a soil *Microbacterium* species and putatively encodes the synthesis of a glycosylated secondary metabolite. We propose that this secondary metabolite acts as a chemical defense agent to the native host based on its antibiotic properties. Furthermore, this cluster is conserved in some *Microbacterium* species and absent in others which suggests that its ecological importance is influenced by the microbial composition and dynamics in a particular niche. Future studies could explore the phylogenetic patterns of the P12B21 BGC throughout the *Microbacterium* genus to expand our knowledge on its distribution and evolution. As demonstrated in this study, characterizing the function and origin of new gene clusters will advance our understanding on the role of microbial secondary metabolites in ecosystems. In turn, this can aid efforts to find the most promising sources of new compounds for human applications.

4.6. Chapter 4 Supplementary Material

Table S4.1. Bacterial strains and plasmids used in this study.

Strain/plasmid/BAC	Description/relevant characteristics[a]	Reference
Plasmids		
pRK2013	Tra^{+}; Kan^{R}; Conjugative helper plasmid in triparental matings	(Figurski & Helinski, 1979)
pDualP	Apra^{R}	(Stankey et al., 2022)
Strains		
Escherichia coli DH10B BAC-Optimized Replicator v2.0	F- *mcrA Δ(mrr-hsdRMS-mcrBC) endA1 recA1 Φ80dlacZ ΔM15 ΔlacX74 araD139 Δ(ara,leu)7697 galU galK rpsL (Str^{R}) nupG (attL araC-PBADtrfA250 bla attR) λ*; Large-insert DNA cloning host	Lucigen Corporation
E. coli HB101	F- *recA13*; General cloning host	(Boyer & Roulland-Dussoix, 1969)
E. coli WTZ17006	F- *Δ(araD-araB)567*, *ΔdegP775::kan*, *ΔlacZ4787*(::rrnB-3), *λ*$^{-}$, *ΔtolC832::FRT*, *rph-1*, *Δ(rhaD-rhaB)568*, *hsdR514*; antibiotic-susceptible strain used for bioassays	(Baba et al., 2006)
Streptomyces coelicolor M1154	M145 Δ*act* Δ*red* Δc*pk* Δ*cda rpoB*(C1298T) *rpsL*(A262G); Host strain for heterologous expression	(Gomez-Escribano & Bibb, 2011)
Staphylococcus aureus Xen29	Biofilm-producing strain used for bioassays	(Park et al., 2012)
Acinetobacter baumannii WR3806	Multidrug-resistant strain used for bioassays	(Taitt et al., 2014)
Microbacterium hatanonis DSM 19179^{T}	Strain used for comparative BGC analysis	(Bakir et al., 2008)
Microbacterium sp. JM-72	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-90	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-97	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-109	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-110	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-112	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-113	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-163	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-164	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-171	Strain used for comparative BGC analysis	This study

[a]Abbreviations: Str^{R}, streptomycin resistant; Tra^{+}, conjugal transfer ability; Kan^{R}, kanamycin resistant; Apra^{R}, apramycin resistant

Strain/plasmid/BAC	Description/relevant characteristics	Reference
Strains		
Microbacterium sp. JM-247	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-308	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-315	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-322	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-356	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-988	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-991	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1025	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1054	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1071	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1077	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1079	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1082	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1083	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1090	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1324	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1325	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1327	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1328	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1330	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1336	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1338	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1340	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1341	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1342	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1343	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1344	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1348	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1349	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1351	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1394	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1398	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1401	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1405	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1450	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1455	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1457	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1464	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1476	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1479	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1480	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1481	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1482	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1483	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1484	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1485	Strain used for comparative BGC analysis	This study

Strain/plasmid/BAC	Description/relevant characteristics	Reference
Strains		
Microbacterium sp. JM-1490	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1493	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1494	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1495	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1496	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1497	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1499	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1500	Strain used for comparative BGC analysis	This study
Microbacterium sp. JM-1962	Strain used for comparative BGC analysis	This study

Table S4.2. Primers used in this study.

Primer name	Sequence	Amplicon length
UniversalBAC F	CGGCAGGTATATGTGATGGGTTAAAATG	200 bp
UniversalBAC R	ACCTCTTACGTGCCGATCAG	
P12B21_core_P1_F	GACCTGTCGAGTCTCACCAC	511 bp
P12B21_core_P1_R	CCTCTTTGAGTCGTCCGGTG	
P12B21_core_P2_F	AACTTCGCGTTCGGACTCAG	607 bp
P12B21_core_P2_R	TGTCCTGCGGGTAGAGATTG	
P12B21_core_P3_F	CTCATCACCGCACTACGCA	491 bp
P12B21_core_P3_R	CCCGAGGTGTAGAGCAGGAA	
P12B21_core_P4_F	TCGTTCCTGCTCTACACCTC	263 bp
P12B21_core_P4_R	CTGATCAACTGCAGCCAGAC	
P12B21_core_P5_F	GCAACACTCATCACCGCACT	502 bp
P12B21_core_P5_R	TCGACCCCGAGGTGTAGAG	

Table S4.3. Predicted function of the genes from the P12B21 BGC.

ORF	Gene length (bp)	Closest homolog [source]	Predicted function	E-value	Protein ID (%)	NCBI Accession
1	741	Glycosyltransferase family 2 protein [*Microbacterium hatanonis*]	Glycosyltransferase	2e-161	96.69	WP_147895459.1
2	1719	AMP-binding protein [*Microbacterium hatanonis*]	AMP-binding protein	0.0	95.14	WP_147895458.1
3	276	Acyl carrier protein [*Microbacterium hatanonis*]	Acyl carrier protein	2e-56	96.70%	WP_147895457.1
4	675	4'-phosphopantetheinyl transferase superfamily protein [*Microbacterium hatanonis*]	4'-phosphopantetheinyl transferase	3e-87	91.24	WP_147895456.1
5	1374	Hypothetical protein [*Microbacterium hatanonis*]	Unknown	0.0	96.06	WP_147895455.1
6	1299	Hypothetical protein [*Microbacterium hatanonis*]	Unknown	0.0	96.76	WP_147895454.1
7	1008	Glycosyltransferase [*Microbacterium hatanonis*]	Glycosyltransferase	0.0	97.31	WP_187266973.1
8	186	Hypothetical protein [*Microbacterium hatanonis*]	Unknown	3e-14	95.08	WP_147895452.1
9	1326	UDP-glucose/GDP-mannose dehydrogenase family protein [*Microbacterium hatanonis*]	UDP-glucose/GDP-mannose dehydrogenase	0.0	99.55	WP_147895451.1
10	393	DUF5684 domain-containing protein [*Microbacterium hatanonis*]	Membrane protein	4e-71	98.46	WP_147895450.1
11	450	VanZ family protein [*Microbacterium hatanonis*]	Resistance protein	8e-68	99.16	WP_147895449.1
12	180	Hypothetical protein [*Microbacterium* sp. Gd 4-13]	Unknown	7e-31	96.61	WP_116648355.1
13	765	ABC transporter permease [*Microbacterium hatanonis*]	ABC transporter permease	3e-140	97.85	WP_147895447.1
14	780	ATP-binding cassette domain-containing protein [*Microbacterium hatanonis*]	ATP-binding cassette domain	3e-171	5.35	WP_147895446.1
15	375	DUF1048 domain-containing protein [*Microbacterium hatanonis*]	Unknown	1e-69	96.36	WP_147895445.1
16	345	PadR family transcriptional regulator [*Microbacterium hatanonis*]	Transcriptional regulator	5e-76	98.25	WP_147895444.1
17	936	Diacylglycerol kinase family lipid kinase [*Microbacterium hatanonis*]	Sugar kinase	0.0	98.39	WP_147895443.1
18	1065	Hypothetical protein [*Microbacterium hatanonis*]	Unknown	0.0	91.50	WP_147895442.1
19	360	Hypothetical protein [*Microbacterium hatanonis*]	Unknown	7e-74	94.96	WP_147895441.1
20	1092	PAS domain-containing sensor histidine kinase [*Microbacterium hatanonis*]	Histidine kinase	0.0	92.82	WP_147895440.1
21	720	Cohesin domain-containing protein [*Microbacterium hatanonis*]	Cohesin domain-containing protein	7e-77	95.82	WP_147895713.1

ORF	Gene length (bp)	Closest homolog [source]	Predicted function	E-value	Protein ID (%)	NCBI Accession
22	1881	Hypothetical protein FVP77_15575 [*Microbacterium hatanonis*]	Unknown	0.0	97.28	TXK10264.1
23	567	SIS domain-containing protein [*Microbacterium hatanonis*]	Sugar isomerase	5e-134	99.47	TXK10263.1
24	1011	ABC transporter permease [*Microbacterium* spp.]	ABC transporter permease	7e-159	99.70	WP_205781275.1
25	1527	Sugar ABC transporter ATP-binding protein [*Microbacterium hatanonis*]	Sugar ABC transporter	0.0	97.24	WP_147895437.1
26	888	Sugar ABC transporter substrate-binding protein [*Microbacterium hatanonis*]	Sugar ABC transporter substrate-binding protein	0.0	100.00	TXK10261.1
27	747	GntR family transcriptional regulator [*Microbacterium hatanonis*]	Transcriptional regulator	1e-176	99.19	TXK10260.1
28	1344	Hypothetical protein FVP77_15545 [*Microbacterium hatanonis*]	Unknown	0.0	98.66	TXK10259.1

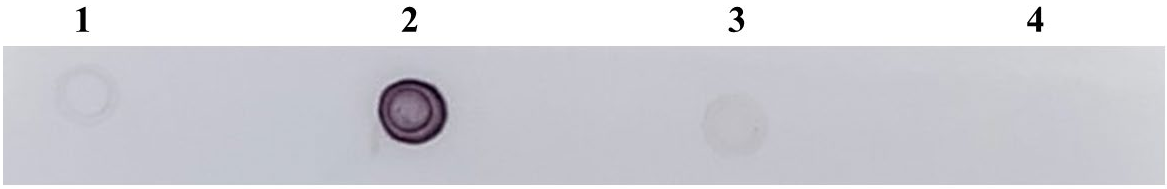

Figure S4.1. Dot blot hybridization with the DIG-labeled probe that is specific to the biosynthetic gene cluster of P12B21. A colorimetric reaction is observed after hybridization with the probe when genomic DNA from *S. coelicolor* clone P12B21 (1), *E. coli* clone P12B21 (2), *M. hatanonis* DSM 19179^{T} (3) is used as a template. Hybridization is not observed when genomic DNA from *S. coelicolor* M1154 (4), which does not harbor the gene cluster, is used as a template.

Chapter 5
Conclusions

There is a rich diversity of microbial life that is waiting to be discovered. Within that hidden diversity, lies a wealth of untapped biosynthetic potential of microbial secondary metabolites with biotechnological, industrial, or pharmaceutical applications. Modern DNA sequencing and –omics-based technologies have revolutionized our ability to study cultured and uncultured microorganisms. Microbial (meta)genomes can now be readily mined for biosynthetic gene clusters (BGCs) encoding the synthesis of secondary metabolites, thereby revealing their metabolic potential. However, harnessing this potential is still challenging due to technical barriers in isolation, screening, and expression of secondary metabolites *in vitro*. Overcoming these challenges is necessary to improve our access to therapeutic compounds and to better understand the role of these metabolites in nature.

The research in this dissertation demonstrated the value of culture-dependent and –independent approaches to characterize and express novel gene clusters encoding antimicrobial secondary metabolites from microorganisms. I also provided evidence that members of the Actinobacteria phylum, namely the genus *Streptomyces*, continue to have an important role in the discovery of natural products. Through a culture-based strategy, the work in Chapter 2 demonstrated the taxonomic novelty and biosynthetic potential of a marine sponge-derived bacterium known as *Streptomyces poriferorum* P01-B04^{T}. My results showed that this species has significant biosynthetic potential for secondary metabolite production as evident by the detection of many unique BGCs in its genome. Furthermore, strain P01-B04^{T} displayed antimicrobial activity against human and plant pathogens suggesting that it could be a promising source of beneficial natural products, but also that these metabolites may have an important role

in the sponge-symbiont relationship. Future studies should assess the clinical relevance of the antimicrobial metabolites by purification and chemical analysis. Additionally, the prevalence of *S. poriferorum* and its predicted BGCs in other sponge microbiomes should be determined to better understand the co-evolutionary history and dynamics between the host and symbiont. The work in this chapter emphasized that culture-based studies still play a necessary part in natural product discovery and we should continue to explore diverse habitats for new microorganisms.

In Chapter 3, the utility of *Streptomyces coelicolor* M1154 as a heterologous host in functional metagenomics was demonstrated. I found that diverse metagenomic BGCs could be expressed in *S. coelicolor* M1154, and through this approach, a high hit rate of clones were identified that produced antimicrobial metabolites capable of inhibiting the growth of a multidrug-resistant pathogen. Moreover, a new inducible expression system was shown to enhance heterologous expression of metagenomic gene clusters by *S. coelicolor* M1154. Therefore, these methodologies represent a means to circumvent some of the common bottlenecks in functional metagenomics that have limited their applicability in drug discovery pipelines. Although the chemical structure of the antimicrobial metabolites produced by these clones was not determined in this study, significant strides were made to establish an approach that could be used to extract and fractionate the compounds, particularly from clone P17B06. Using bioinformatics tools, I demonstrated that P17B06 expresses a BGC that is derived from the Acidobacteria phylum. To my knowledge, this is the first report of an antimicrobial metabolite from this phylum which sheds light on the importance of this taxa in future natural product discovery efforts. However, low compound yields impeded our ability to purify and characterize the antimicrobial metabolites produced by clone P17B06. Future work should seek to enhance expression of this BGC and improve compound yields, either by refactoring the

pathway or assessing its expression in other hosts. Taken together, the work in Chapter 3 provides a means to study the function of BGCs from uncultured microorganisms and find new biomolecules.

The research in Chapter 3 was largely focused on finding and expressing metagenomic BGCs for the purpose of discovering new antibiotic compounds. In Chapter 4, a biology-guided study was taken to explore the functional importance and origin of a metagenomic BGC, P12B21. My results showed that the P12B21 BGC originated from a *Microbacterium* species within the Actinobacteria phylum and found evidence that this gene cluster is highly homologous to a cluster from *Microbacterium hatanonis*. This cluster was not found in other endophytic *Microbacterium* species and I speculated that the BGC may have been lost in those isolates if it was not providing a selective advantage to them. I also provided preliminary evidence that the BGC encodes the synthesis of a glycosylated antimicrobial compound, which may serve as chemical defense to those strains that maintain and express the gene cluster. Future studies should include an in-depth BGC phylogenetic analysis to better understand the distribution and ecological relevance of this cluster across the *Microbacterium* genus.

Collectively, the work in this dissertation has the exciting potential to expand our access to the metabolic diversity of microorganisms. From one perspective, this research may help guide future efforts to discover beneficial compounds from nature which is desperately needed to combat rising rates of antimicrobial resistance. From a separate viewpoint, this research provides insight into the chemical ecology of soil and marine bacteria.

References

Abdelmohsen, U. R., Bayer, K., & Hentschel, U. (2014). Diversity, abundance and natural products of marine sponge-associated actinomycetes. *Natural Product Reports, 31*(3), 381-399.

Abdelmohsen, U. R., Pimentel-Elardo, S. M., Hanora, A., Radwan, M., Abou-El-Ela, S. H., Ahmed, S., & Hentschel, U. (2010). Isolation, phylogenetic analysis and anti-infective activity screening of marine sponge-associated actinomycetes. *Marine Drugs, 8*(3), 399-412.

Adamek, M., Alanjary, M., Sales-Ortells, H., Goodfellow, M., Bull, A. T., Winkler, A., . . . Ziemert, N. (2018). Comparative genomics reveals phylogenetic distribution patterns of secondary metabolites in Amycolatopsis species. *BMC Genomics, 19*(1), 1-15.

Agrawal, P., Khater, S., Gupta, M., Sain, N., & Mohanty, D. (2017). RiPPMiner: a bioinformatics resource for deciphering chemical structures of RiPPs based on prediction of cleavage and cross-links. *Nucleic Acids Research, 45*(W1), W80-W88.

Ahmed, Y., Rebets, Y., Estévez, M. R., Zapp, J., Myronovskyi, M., & Luzhetskyy, A. (2020). Engineering of Streptomyces lividans for heterologous expression of secondary metabolite gene clusters. *Microbial cell factories, 19*(1), 1-16.

Alam, K., Abbasi, M. N., Hao, J., Zhang, Y., & Li, A. (2021). Strategies for Natural Products Discovery from Uncultured Microorganisms. *Molecules, 26*(10), 2977.

Alanis, A. J. (2005). Resistance to antibiotics: are we in the post-antibiotic era? *Archives of medical research, 36*(6), 697-705.

Albarano, L., Esposito, R., Ruocco, N., & Costantini, M. (2020). Genome mining as new challenge in natural products discovery. *Marine Drugs, 18*(4), 199.

Alt, S., & Wilkinson, B. (2015). Biosynthesis of the novel macrolide antibiotic anthracimycin. *ACS chemical biology, 10*(11), 2468-2479.

Amann, R. I., Binder, B. J., Olson, R. J., Chisholm, S. W., Devereux, R., & Stahl, D. (1990). Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Applied and Environmental Microbiology, 56*(6), 1919-1925.

Amara, A., Takano, E., & Breitling, R. (2018). Development and validation of an updated computational model of Streptomyces coelicolor primary and secondary metabolism. *BMC Genomics, 19*(1), 1-9.

Anderson, A. S., & Wellington, E. (2001). The taxonomy of *Streptomyces* and related genera. *International Journal of Systematic and Evolutionary Microbiology, 51*(3), 797-814.

Angov, E., Hillier, C. J., Kincaid, R. L., & Lyon, J. A. (2008). Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host. *PloS one, 3*(5), e2189.

Arkin, A. P., Cottingham, R. W., Henry, C. S., Harris, N. L., Stevens, R. L., Maslov, S., . . . Canon, S. (2018). KBase: the United States department of energy systems biology knowledgebase. *Nature Biotechnology, 36*(7), 566-569.

Arocha-Garza, H. F., Canales-Del Castillo, R., Eguiarte, L. E., Souza, V., & De la Torre-Zavala, S. (2017). High diversity and suggested endemicity of culturable Actinobacteria in an extremely oligotrophic desert oasis. *PeerJ, 5*, e3247.

Asokan, G. V., & Vanitha, A. (2018). WHO global priority pathogens list on antibiotic resistance: an urgent need for action to integrate One Health data. *Perspectives in public health, 138*(2), 87-88.

Atlas, R. M. (2010). *Handbook of microbiological media*. Boca Raton, FL: CRC press.

Auch, A. F., von Jan, M., Klenk, H.-P., & Göker, M. (2010). Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Standards in genomic sciences, 2*(1), 117-134.

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., . . . Kubal, M. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics, 9*(1), 1-15.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., . . . Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology, 2*(1), 2006.0008.

Baig, U., Dahanukar, N., Shintre, N., Holkar, K., Pund, A. I., Lele, U., . . . Singh, R. (2020). Phylogenetic diversity and activity screening of cultivable actinobacteria isolated from marine sponges and associated environments from the western coast of India. *bioRxiv*.

Bakir, M. A., Kudo, T., & Benno, Y. (2008). Microbacterium hatanonis sp. nov., isolated as a contaminant of hairspray. *International Journal of Systematic and Evolutionary Microbiology, 58*(3), 654-658.

Baltz, R. H. (2010). Streptomyces and Saccharopolyspora hosts for heterologous expression of secondary metabolite gene clusters. *Journal of Industrial Microbiology and Biotechnology, 37*(8), 759-772.

Baltz, R. H. (2019). Natural product drug discovery in the genomic era: realities, conjectures, misconceptions, and opportunities. *Journal of Industrial Microbiology and Biotechnology, 46*(3-4), 281-299.

Bano, N., Siddiqui, S., Amir, M., Zia, Q., Banawas, S., & Iqbal, D. (2022). Bioprospecting of the novel isolate Microbacterium proteolyticum LA2 (R) from the rhizosphere of Rauwolfia serpentina. *Saudi Journal of Biological Sciences, 29*(3), 1858-1868.

Barka, E. A., Vatsa, P., Sanchez, L., Gaveau-Vaillant, N., Jacquard, C., Klenk, H.-P., . . . van Wezel, G. P. (2016). Taxonomy, physiology, and natural products of Actinobacteria. *Microbiology and Molecular Biology Reviews, 80*(1), 1-43.

Barona-Gómez, F., Wong, U., Giannakopulos, A. E., Derrick, P. J., & Challis, G. L. (2004). Identification of a Cluster of Genes that Directs Desferrioxamine Biosynthesis in Streptomyces c oelicolor M145. *Journal of the American Chemical Society, 126*(50), 16282-16283.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Bauer, A. (1966). Antibiotic susceptibility testing by a standardized single disc method. *American journal of clinical pathology, 45,*, 149-158.

Bauer, J. D., King, R. W., & Brady, S. F. (2010). Utahmycins A and B, azaquinones produced by an environmental DNA clone. *Journal of Natural Products, 73*(5), 976-979.

Bauman, K. D., Li, J., Murata, K., Mantovani, S. M., Dahesh, S., Nizet, V., . . . Moore, B. S. (2019). Refactoring the cryptic streptophenazine biosynthetic gene cluster unites phenazine, polyketide, and nonribosomal peptide biochemistry. *Cell chemical biology, 26*(5), 724-736. e727.

Bavestrello, G., Arillo, A., Calcinai, B., Cattaneo-Vietti, R., Cerrano, C., Gaino, E., . . . Sara, M. (2000). Parasitic diatoms inside Antarctic sponges. *Biology, 198*(1), 29-33.

Bentley, S. D., Chater, K. F., Cerdeño-Tárraga, A.-M., Challis, G. L., Thomson, N., James, K. D., . . . Harper, D. (2002). Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3 (2). *Nature, 417*(6885), 141-147.

Berger, J. (1949). Borrelidin, a new antibiotic with anti-borrelia activity and penicillin enhancement properties. *Archives of biochemistry and biophysics, 22*, 476-478.

Bibb, M. J. (2005). Regulation of secondary metabolism in streptomycetes. *Current Opinion in Microbiology, 8*(2), 208-215.

Bierman, M., Logan, R., O'brien, K., Seno, E., Rao, R. N., & Schoner, B. (1992). Plasmid cloning vectors for the conjugal transfer of DNA from Escherichia coli to Streptomyces spp. *Gene, 116*(1), 43-49.

Bitok, J. K., Lemetre, C., Ternei, M. A., & Brady, S. F. (2017). Identification of biosynthetic gene clusters from metagenomic libraries using PPTase complementation in a Streptomyces host. *FEMS Microbiology Letters, 364*(16).

Biver, S., & Vandenbol, M. (2013). Characterization of three new carboxylic ester hydrolases isolated by functional screening of a forest soil metagenomic library. *Journal of Industrial Microbiology and Biotechnology, 40*(2), 191-200.

Blin, K., Shaw, S., Kloosterman, A. M., Charlop-Powers, Z., van Wezel, G. P., Medema, M. H., & Tilmann, W. (2021). antiSMASH 6.0. *Nucleic Acids Research*.

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S. Y., . . . Weber, T. (2019). antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Research, 47*(W1), W81-W87.

Blodgett, J. A., Oh, D.-C., Cao, S., Currie, C. R., Kolter, R., & Clardy, J. (2010). Common biosynthetic origins for polycyclic tetramate macrolactams from phylogenetically diverse bacteria. *Proceedings of the National Academy of Sciences, 107*(26), 11692-11697.

Bodor, A., Bounedjoum, N., Vincze, G. E., Kis, Á. E., Laczi, K., Bende, G., . . . Rákhely, G. (2020). Challenges of unculturable bacteria: environmental perspectives. *Reviews in Environmental Science and Bio/Technology, 19*(1), 1-22.

Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D., & Pirovano, W. (2011). Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics, 27*(4), 578-579.

Boetzer, M., & Pirovano, W. (2012). Toward almost closed genomes with GapFiller. *Genome Biology, 13*(6), 1-9.

Boyer, H. W., & Roulland-Dussoix, D. (1969). A complementation analysis of the restriction and modification of DNA in Escherichia coli. *Journal of Molecular Biology, 41*(3), 459-472.

Braga, R. M., Dourado, M. N., & Araújo, W. L. (2016). Microbial interactions: ecology in a molecular perspective. *Brazilian Journal of Microbiology, 47*, 86-98.

Bredholt, H., Fjærvik, E., Johnsen, G., & Zotchev, S. B. (2008). Actinomycetes from sediments in the Trondheim fjord, Norway: diversity and biological activity. *Marine Drugs, 6*(1), 12-24.

Brown, S. D., Palumbo, A. V., Panikov, N., Ariyawansa, T., Klingeman, D. M., Johnson, C. M., . . . Epstein, S. S. (2012). Draft genome sequence for Microbacterium laevaniformans strain OR221, a bacterium tolerant to metals, nitrate, and low pH. In: Am Soc Microbiol.

Bruns, H., Crüsemann, M., Letzel, A.-C., Alanjary, M., McInerney, J. O., Jensen, P. R., . . . Ziemert, N. (2018). Function-related replacement of bacterial siderophore pathways. *The ISME journal, 12*(2), 320-329.

Bulens, S. N., Sarah, H. Y., Walters, M. S., Jacob, J. T., Bower, C., Reno, J., . . . Janelle, S. J. (2018). Carbapenem-nonsusceptible Acinetobacter baumannii, 8 US metropolitan areas, 2012–2015. *Emerging Infectious Diseases, 24*(4), 727.

Cacho, R. A., Tang, Y., & Chooi, Y.-H. (2015). Next-generation sequencing approach for connecting secondary metabolites to biosynthetic gene clusters in fungi. *Frontiers in microbiology, 5*, 774.

Caesar, L. K., & Cech, N. B. (2019). Synergy and antagonism in natural product extracts: when 1+ 1 does not equal 2. *Natural product reports, 36*(6), 869-888.

Carini, P. (2019). A "cultural" renaissance: genomics breathes new life into an old craft. *Msystems, 4*(3), e00092-00019.

Carlson, J. C., Li, S., Burr, D. A., & Sherman, D. H. (2009). Isolation and characterization of tirandamycins from a marine-derived Streptomyces sp. *Journal of Natural Products, 72*(11), 2076-2079.

Chakraburtty, R., & Bibb, M. (1997). The ppGpp synthetase gene (relA) of *Streptomyces coelicolor* A3 (2) plays a conditional role in antibiotic production and morphological differentiation. *Journal of Bacteriology, 179*(18), 5854-5861.

Charles, T. C., Liles, M. R., & Sessitsch, A. (2017). *Functional metagenomics: Tools and applications*: Springer.

Charlop-Powers, Z., Milshteyn, A., & Brady, S. F. (2014). Metagenomic small molecule discovery methods. *Current Opinion in Microbiology, 19*, 70-75.

Charlop-Powers, Z., Pregitzer, C. C., Lemetre, C., Ternei, M. A., Maniko, J., Hover, B. M., . . . Forgione, H. M. (2016). Urban park soil microbiomes are a rich reservoir of natural product biosynthetic diversity. *Proceedings of the national academy of sciences, 113*(51), 14811-14816.

Chase, A. B., Sweeney, D., Muskat, M. N., Guillén-Matus, D. G., & Jensen, P. R. (2021). Vertical Inheritance Facilitates Interspecies Diversification in Biosynthetic Gene Clusters and Specialized Metabolites. *Mbio, 12*(6), e02700-02721.

Chater, K. F. (2016). Recent advances in understanding *Streptomyces*. *F1000Research, 5*.

Chen, H., Olson, A. S., Su, W., Dussault, P. H., & Du, L. (2015). Fatty acyl incorporation in the biosynthesis of WAP-8294A, a group of potent anti-MRSA cyclic lipodepsipeptides. *RSC advances, 5*(128), 105753-105759.

Chen, I.-M. A., Chu, K., Palaniappan, K., Ratner, A., Huang, J., Huntemann, M., . . . Seshadri, R. (2021). The IMG/M data management and analysis system v. 6.0: new tools and advanced capabilities. *Nucleic Acids Research, 49*(D1), D751-D763.

Chen, R., Wong, H. L., & Burns, B. P. (2019). New approaches to detect biosynthetic gene clusters in the environment. *Medicines, 6*(1), 32.

Chevrette, M. G., Carlos-Shanley, C., Louie, K. B., Bowen, B. P., Northen, T. R., & Currie, C. R. (2019). Taxonomic and metabolic incongruence in the ancient genus *Streptomyces*. *Frontiers in microbiology, 10*, 2170.

Chevrette, M. G., Carlson, C. M., Ortega, H. E., Thomas, C., Ananiev, G. E., Barns, K. J., . . . Flanigan, W. (2019). The antimicrobial potential of Streptomyces from insect microbiomes. *Nature communications, 10*(1), 1-11.

Chistoserdova, L. (2009). Functional metagenomics: recent advances and future challenges. *Biotechnology and Genetic Engineering Reviews, 26*(1), 335-352.

Cho, J. Y., & Kim, M. S. (2012). Induction of antifouling diterpene production by Streptomyces cinnabarinus PK209 in co-culture with marine-derived Alteromonas sp. KNS-16. *Bioscience, Biotechnology, and Biochemistry, 76*(10), 1849-1854.

Christensen, W. B. (1946). Urea decomposition as a means of differentiating *Proteus* and paracolon cultures from each other and from *Salmonella* and *Shigella* types. *Journal of Bacteriology, 52*(4), 461.

Cimermancic, P., Medema, M. H., Claesen, J., Kurita, K., Brown, L. C. W., Mavrommatis, K., . . . Clardy, J. (2014). Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell, 158*(2), 412-421.

Claessen, D., De Jong, W., Dijkhuizen, L., & Wösten, H. A. (2006). Regulation of *Streptomyces* development: reach for the sky! *Trends in Microbiology, 14*(7), 313-319.

Claessen, D., Rink, R., de Jong, W., Siebring, J., de Vreugd, P., Boersma, F. H., . . . Wösten, H. A. (2003). A novel class of secreted hydrophobic proteins is involved in aerial hyphae formation in *Streptomyces coelicolor* by forming amyloid-like fibrils. *Genes & Development, 17*(14), 1714-1726.

Claessen, D., Stokroos, I., Deelstra, H. J., Penninga, N. A., Bormann, C., Salas, J. A., . . . Wösten, H. A. (2004). The formation of the rodlet layer of streptomycetes is the result of the interplay between rodlins and chaplins. *Molecular Microbiology, 53*(2), 433-443.

Cleary, D. F., Becking, L. E., Voogd, N. J. d., Pires, A. C., Polónia, A. R., Egas, C., & Gomes, N. C. (2013). Habitat-and host-related variation in sponge bacterial symbiont communities in Indonesian waters. *FEMS Microbiology Ecology, 85*(3), 465-482.

CLSI. (2015). M100-S25 performance standards for antimicrobial susceptibility testing. *Twenty-Fifth Informational Supplement, 35*, 1-240.

Colin, P.-Y., Kintses, B., Gielen, F., Miton, C. M., Fischer, G., Mohamed, M. F., . . . Hollfelder, F. (2015). Ultrahigh-throughput discovery of promiscuous enzymes by picodroplet functional metagenomics. *Nature communications, 6*(1), 1-12.

Collins, C. H. (1967). Microbiological methods. *Journal of Microbiological Methods*(2nd Edition).

Control, C. f. D., & Prevention. (2019). *Antibiotic resistance threats in the United States, 2019*: US Department of Health and Human Services, Centres for Disease Control and ….

Corretto, E., Antonielli, L., Sessitsch, A., Höfer, C., Puschenreiter, M., Widhalm, S., . . . Brader, G. (2020). Comparative genomics of Microbacterium species to reveal diversity, potential for secondary metabolites and heavy metal resistance. *Frontiers in microbiology*, 1869.

Costa, M., Zúñiga, P., Peñalver, A. M., Thorsteinsdottir, M., Pérez, M., Cañedo, L. M., & Cuevas, C. (2017). New fluvirucinins C1 and C2 produced by a marine derived actinomycete. *Natural product communications, 12*(5), 122-173.

Craig, J. W., Chang, F.-Y., Kim, J. H., Obiajulu, S. C., & Brady, S. F. (2010). Expanding small-molecule functional metagenomics through parallel screening of broad-host-range cosmid environmental DNA libraries in diverse proteobacteria. *Applied and Environmental Microbiology, 76*(5), 1633-1641.

Craney, A., Ahmed, S., & Nodwell, J. (2013). Towards a new science of secondary metabolism. *The Journal of antibiotics, 66*(7), 387-400.

Cribari-Neto, F., & Zeileis, A. (2010). Beta regression in R. *Journal of statistical software, 34*, 1-24.

D'Costa, V. M., Griffiths, E., & Wright, G. D. (2007). Expanding the soil antibiotic resistome: exploring environmental diversity. *Current Opinion in Microbiology, 10*(5), 481-489.

Darling, A. C., Mau, B., Blattner, F. R., & Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research, 14*(7), 1394-1403.

de Jesus Sousa, J. A., & Olivares, F. L. (2016). Plant growth promotion by streptomycetes: ecophysiology, mechanisms and applications. *Chemical and Biological Technologies in Agriculture, 3*(1), 1-12.

de Lima Procópio, R. E., da Silva, I. R., Martins, M. K., de Azevedo, J. L., & de Araújo, J. M. (2012). Antibiotics produced by *Streptomyces*. *The Brazilian journal of infectious diseases, 16*(5), 466-471.

Demain, A. L., & Fang, A. (2000). The natural functions of secondary metabolites. *History of modern biotechnology I*, 1-39.

Devine, R., Hutchings, M. I., & Holmes, N. A. (2017). Future directions for the discovery of antibiotics from actinomycete bacteria. *Emerging Topics in Life Sciences*, ETLS20160014.

Dharmaraj, S. (2010). Marine *Streptomyces* as a novel source of bioactive substances. *World Journal of Microbiology & Biotechnology, 26*(12), 2123-2139.

Dias, D. A., Urban, S., & Roessner, U. (2012). A historical overview of natural products in drug discovery. *Metabolites, 2*(2), 303-336.

Dickinson, L., Griffiths, A., Mason, C., & Mills, R. (1965). Anti-viral activity of two antibiotics isolated from a species of *Streptomyces*. *Nature, 206*, 265-268.

Doroghazi, J. R., Albright, J. C., Goering, A. W., Ju, K.-S., Haines, R. R., Tchalukov, K. A., . . . Metcalf, W. W. (2014). A roadmap for natural product discovery based on large-scale genomics and metabolomics. *Nature Chemical Biology, 10*(11), 963-968.

Du, L., Sánchez, C., & Shen, B. (2001). Hybrid peptide–polyketide natural products: biosynthesis and prospects toward engineering novel molecules. *Metabolic Engineering, 3*(1), 78-95.

Du, L., & Shen, B. (2001). Biosynthesis of hybrid peptide-polyketide natural products. *Current opinion in drug discovery & development, 4*(2), 215-228.

Dutescu, I. A., & Hillier, S. A. (2021). Encouraging the Development of New Antibiotics: Are Financial Incentives the Right Way Forward? A Systematic Review and Case Study. *Infection and Drug Resistance, 14*, 415.

Dyson, P. (2009). Streptomyces, in encyclopedia of microbiology. In M. Schaechter (Ed.), (pp. 318-322): Oxford: Academic Press.

Dzhavakhiya, V., Glagoleva, E., Voinova, T., Savelyeva, V., Ovchinnikov, A., & Matsyura, A. (2020). Effect of adsorbing resins on the rapamycin biosynthesis by Streptomyces hygroscopicus VKM Ac-2737D. *Ukrainian Journal of Ecology, 10*(6), 230-235.

Elliot, M., Damji, F., Passantino, R., Chater, K., & Leskiw, B. (1998). The bldD gene of *Streptomyces coelicolor* A3 (2): a regulatory gene involved in morphogenesis and antibiotic production. *Journal of Bacteriology, 180*(6), 1549-1555.

Erich, N. (2014). RColorBrewer: ColorBrewer Palettes. *R package version*, 1.1-2.

Farris, J. S. (1972). Estimating phylogenetic trees from distance matrices. *The American naturalist, 106*(951), 645-668.

Farris, M. H., Duffy, C., Findlay, R. H., & Olson, J. B. (2011). *Streptomyces scopuliridis* sp. nov., a bacteriocin-producing soil streptomycete. *International Journal of Systematic and Evolutionary Microbiology, 61*(Pt 9), 2112.

Fazal, A., Thankachan, D., Harris, E., & Seipke, R. F. (2020). A chromatogram-simplified Streptomyces albus host for heterologous production of natural products. *Antonie Van Leeuwenhoek, 113*(4), 511-520.

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution, 17*(6), 368-376.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution, 39*(4), 783-791.

Feng, Z., Kallifidas, D., & Brady, S. F. (2011). Functional analysis of environmental DNA-derived type II polyketide synthases reveals structurally diverse secondary metabolites. *Proceedings of the National Academy of Sciences, 108*(31), 12629-12634.

Figurski, D. H., & Helinski, D. R. (1979). Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans. *Proceedings of the National Academy of Sciences, 76*(4), 1648-1652.

Fisch, K. M. (2013). Biosynthesis of natural products by microbial iterative hybrid PKS–NRPS. *RSC advances, 3*(40), 18228-18247.

Fischbach, M. A., & Walsh, C. T. (2006). Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. *Chemical reviews, 106*(8), 3468-3496.

Fischbach, M. A., Walsh, C. T., & Clardy, J. (2008). The evolution of gene collectives: How natural selection drives chemical innovation. *Proceedings of the National Academy of Sciences, 105*(12), 4601-4608.

Flärdh, K. (2003). Growth polarity and cell division in *Streptomyces*. *Current Opinion in Microbiology, 6*(6), 564-571.

Flärdh, K., & Buttner, M. J. (2009). *Streptomyces* morphogenetics: dissecting differentiation in a filamentous bacterium. *Nature Reviews Microbiology, 7*(1), 36-49.

Flärdh, K., Richards, D. M., Hempel, A. M., Howard, M., & Buttner, M. J. (2012). Regulation of apical growth and hyphal branching in *Streptomyces*. *Current Opinion in Microbiology, 15*(6), 737-743.

Fleming, A. (1929). On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of *B. influenzae*. *British journal of experimental pathology, 10*(3), 226.

Florez, A. B., Alvarez, S., Zabala, D., Brana, A. F., Salas, J. A., & Mendez, C. (2015). Transcriptional regulation of mithramycin biosynthesis in Streptomyces argillaceus: dual role as activator and repressor of the PadR-like regulator MtrY. *Microbiology, 161*(2), 272-284.

Freire-Moran, L., Aronsson, B., Manz, C., Gyssens, I. C., So, A. D., Monnet, D. L., . . . group, E.-E. w. (2011). Critical shortage of new antibiotics in development against multidrug-resistant bacteria—Time to react is now. *Drug resistance updates, 14*(2), 118-124.

Gabor, E. M., Alkema, W. B., & Janssen, D. B. (2004). Quantifying the accessibility of the metagenome by random expression cloning techniques. *Environmental Microbiology, 6*(9), 879-886.

Gadd, G. M., & Sariaslani, S. (2016). *Advances in Applied Microbiology*: Academic Press.

Gavriilidou, A., Kautsar, S. A., Zaburannyi, N., Krug, D., Müller, R., Medema, M. H., & Ziemert, N. (2022). Compendium of specialized metabolite biosynthetic diversity encoded in bacterial genomes. *Nature Microbiology, 7*(5), 726-735.

Genilloud, O. (2017). Actinomycetes: still a source of novel antibiotics. *Natural product reports, 34*(10), 1203-1232.

Gillespie, D. E., Brady, S. F., Bettermann, A. D., Cianciotto, N. P., Liles, M. R., Rondon, M. R., . . . Handelsman, J. (2002). Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. *Applied and Environmental Microbiology, 68*(9), 4301-4306.

Gomez-Escribano, J. P., & Bibb, M. J. (2011). Engineering Streptomyces coelicolor for heterologous expression of secondary metabolite gene clusters. *Microbial biotechnology, 4*(2), 207-215.

González-Menéndez, V., Asensio, F., Moreno, C., de Pedro, N., Monteiro, M. C., de la Cruz, M., . . . Genilloud, O. (2014). Assessing the effects of adsorptive polymeric resin additions on fungal secondary metabolite chemical diversity. *Mycology, 5*(3), 179-191.

Goodfellow, M. (2012). Phylum XXVI. Actinobacteria phyl. nov. In *Bergey's Manual® of Systematic Bacteriology* (pp. 33-2028): Springer.

Gould, K. (2016). Antibiotics: from prehistory to the present day. *Journal of Antimicrobial Chemotherapy, 71*(3), 572-575.

Graça, A. P., Bondoso, J., Gaspar, H., Xavier, J. R., Monteiro, M. C., de la Cruz, M., . . . Lage, O. M. (2013). Antimicrobial activity of heterotrophic bacterial communities from the marine sponge Erylus discophorus (Astrophorida, Geodiidae). *PLoS One, 8*(11), e78992.

Gregson, J. M., Chen, J.-L., Patterson, G. M., & Moore, R. E. (1992). Structures of puwainaphycins A–E. *Tetrahedron, 48*(18), 3727-3734.

Greule, A., Marolt, M., Deubel, D., Peintner, I., Zhang, S., Jessen-Trefzer, C., . . . Friedrich, T. (2017). Wide Distribution of Foxicin Biosynthetic Gene Clusters in Streptomyces Strains–

An Unusual Secondary Metabolite with Various Properties. *Frontiers in microbiology, 8*, 221.

Guerrero-Garzón, J. F., Zehl, M., Schneider, O., Rückert, C., Busche, T., Kalinowski, J., . . . Zotchev, S. B. (2020). *Streptomyces* spp. from the marine sponge *Antho dichotoma*: analyses of secondary metabolite biosynthesis gene clusters and some of their products. *Frontiers in microbiology, 11*, 437.

Guo, Z., Ma, S., Khan, S., Zhu, H., Zhang, B., Zhang, S., & Jiao, R. (2021). Zhaoshumycins A and B, Two Unprecedented Antimycin-Type Depsipeptides Produced by the Marine-Derived Streptomyces sp. ITBB-ZKa6. *Marine Drugs, 19*(11), 624.

Gustafsson, C., Govindarajan, S., & Minshull, J. (2004). Codon bias and heterologous protein expression. *Trends in Biotechnology, 22*(7), 346-353.

Hamm, P. S., Caimi, N. A., Northup, D. E., Valdez, E. W., Buecher, D. C., Dunlap, C. A., . . . Porras-Alfaro, A. (2017). Western bats as a reservoir of novel Streptomyces species with antifungal activity. *Applied and Environmental Microbiology, 83*(5), e03057-03016.

Handelsman, J. (2004). Metagenomics: application of genomics to uncultured microorganisms. *Microbiology and Molecular Biology Reviews, 68*(4), 669-685.

Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., & Goodman, R. M. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry & Biology, 5*(10), R245-R249.

Hannigan, G. D., Prihoda, D., Palicka, A., Soukup, J., Klempir, O., Rampula, L., . . . Chang, D. (2019). A deep learning genome-mining strategy for biosynthetic gene cluster prediction. *Nucleic Acids Research, 47*(18), e110-e110.

Hanson, J. R. (2003). The classes of natural product and their isolation. In *Natural Products* (pp. 1-34).

Healy, F., Ray, R., Aldrich, H., Wilkie, A., Ingram, L., & Shanmugam, K. (1995). Direct isolation of functional genes encoding cellulases from the microbial consortia in a thermophilic, anaerobic digester maintained on lignocellulose. *Applied Microbiology and Biotechnology, 43*(4), 667-674.

Heiberger, R. M., Heiberger, R. M., & Burt Holland, B. H. (2015). *Statistical Analysis and Data Display An Intermediate Course with Examples in R*: Springer.

Hentschel, U., Schmid, M., Wagner, M., Fieseler, L., Gernert, C., & Hacker, J. (2001). Isolation and phylogenetic analysis of bacteria with antimicrobial activities from the Mediterranean sponges *Aplysina aerophoba* and *Aplysina cavernicola*. *FEMS Microbiology Ecology, 35*(3), 305-312.

Herai, S., Hashimoto, Y., Higashibata, H., Maseda, H., Ikeda, H., Ōmura, S., & Kobayashi, M. (2004). Hyper-inducible expression system for streptomycetes. *Proceedings of the National Academy of Sciences, 101*(39), 14031-14035.

Hesketh, A., Sun, J., & Bibb, M. (2001). Induction of ppGpp synthesis in *Streptomyces coelicolor* A3 (2) grown under conditions of nutritional sufficiency elicits actII-ORF4 transcription and actinorhodin biosynthesis. *Molecular Microbiology, 39*(1), 136-144.

Hillerich, B., & Westpheling, J. (2006). A new GntR family transcriptional regulator in Streptomyces coelicolor is required for morphogenesis and antibiotic production and controls transcription of an ABC transporter in response to carbon source. *Journal of Bacteriology, 188*(21), 7477-7487.

Hofer, U. (2018). The majority is uncultured. *Nature Reviews Microbiology, 16*(12), 716-717.

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

Hover, B. M., Kim, S.-H., Katz, M., Charlop-Powers, Z., Owen, J. G., Ternei, M. A., . . . Park, S. (2018). Culture-independent discovery of the malacidins as calcium-dependent antibiotics with activity against multidrug-resistant Gram-positive pathogens. *Nature microbiology, 3*(4), 415-422.

Hu, G.-P., Yuan, J., Sun, L., She, Z.-G., Wu, J.-H., Lan, X.-J., . . . Chen, S.-P. (2011). Statistical research on marine natural products based on data obtained between 1985 and 2008. *Marine Drugs, 9*(4), 514-525.

Huang, F., Tang, J., He, L., Ding, X., Huang, S., Zhang, Y., . . . Xia, L. (2018). Heterologous expression and antitumor activity analysis of syringolin from Pseudomonas syringae pv. syringae B728a. *Microbial Cell Factories, 17*(1), 1-12.

Huo, L., Hug, J. J., Fu, C., Bian, X., Zhang, Y., & Müller, R. (2019). Heterologous expression of bacterial natural product biosynthetic pathways. *Natural product reports, 36*(10), 1412-1436.

Hur, G. H., Vickery, C. R., & Burkart, M. D. (2012). Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. *Natural product reports, 29*(10), 1074-1098.

Hutchings, M. I., Truman, A. W., & Wilkinson, B. (2019). Antibiotics: past, present and future. *Current Opinion in Microbiology, 51*, 72-80.

Ian, E., Malko, D. B., Sekurova, O. N., Bredholt, H., Rückert, C., Borisova, M. E., . . . Zotchev, S. B. (2014). Genomics of sponge-associated *Streptomyces* spp. closely related to *Streptomyces albus* J1074: insights into marine adaptation and secondary metabolite biosynthesis potential. *PLoS One, 9*(5), e96719.

Imamura, N., Nishijima, M., Adachi, K., & Sano, H. (1993). Novel antimycin antibiotics, urauchimycins A and B, produced by marine actinomycete. *Journal of antibiotics research, 46*(2), 241-246.

Iqbal, H. A., Craig, J. W., & Brady, S. F. (2014). Antibacterial enzymes from the functional screening of metagenomic libraries hosted in Ralstonia metallidurans. *FEMS Microbiology Letters, 354*(1), 19-26.

Iqbal, H. A., Low-Beinart, L., Obiajulu, J. U., & Brady, S. F. (2016). Natural product discovery through improved functional metagenomics in Streptomyces. *Journal of the American Chemical Society, 138*(30), 9341-9344.

Ivanova, V., Lyutskanova, D., Kolarova, M., Aleksieva, K., Raykovska, V., & Stoilova-Disheva, M. (2010). Structural elucidation of a bioactive metabolites produced by Streptomyces Avidinii SB9 strain, isolated from permafrost soil in Spitsbergen, Arctic. *Biotechnology & Biotechnological Equipment, 24*(4), 2092-2095.

Jackson, N., Czaplewski, L., & Piddock, L. J. (2018). Discovery and development of new antibacterial drugs: learning from experience? *Journal of Antimicrobial Chemotherapy, 73*(6), 1452-1459.

Johnston, C. W., Badran, A. H., & Collins, J. J. (2020). Continuous bioactivity-dependent evolution of an antibiotic biosynthetic pathway. *Nature communications, 11*(1), 1-8.

Jones, A. C., Gust, B., Kulik, A., Heide, L., Buttner, M. J., & Bibb, M. J. (2013). Phage p1-derived artificial chromosomes facilitate heterologous expression of the FK506 gene cluster. *PloS one, 8*(7), e69319.

Jones, K. L. (1949). Fresh isolates of actinomycetes in which the presence of sporogenous aerial mycelia is a fluctuating characteristic. *Journal of Bacteriology, 57*(2), 141.

Jones, S. E., Ho, L., Rees, C. A., Hill, J. E., Nodwell, J. R., & Elliot, M. A. (2017). *Streptomyces* exploration is triggered by fungal interactions and volatile signals. *Elife, 6*, e21738.

Kakirde, K. S., Parsley, L. C., & Liles, M. R. (2010). Size does matter: application-driven approaches for soil metagenomics. *Soil Biology and Biochemistry, 42*(11), 1911-1923.

Kalinovskaya, N. I., Romanenko, L. A., Irisawa, T., Ermakova, S. P., & Kalinovsky, A. I. (2011). Marine isolate Citricoccus sp. KMM 3890 as a source of a cyclic siderophore nocardamine with antitumor activity. *Microbiological Research, 166*(8), 654-661.

Kallen, A. J., Mu, Y., Bulens, S., Reingold, A., Petit, S., Gershman, K., . . . Dumyati, G. (2010). Health care–associated invasive MRSA infections, 2005-2008. *Jama, 304*(6), 641-647.

Kanagasabhapathy, M., Sasaki, H., & Nagata, S. (2008). Phylogenetic identification of epibiotic bacteria possessing antimicrobial activities isolated from red algal species of Japan. *World Journal of Microbiology and Biotechnology, 24*(10), 2315-2321.

Kassambara, A. (2018). ggpubr:"ggplot2" based publication ready plots. *R package version 0.1, 7*.

Kassambara, A. (2020). ggpubr:"ggplot2" based publication ready plots. *R package version 0.4. 0, 438*.

Katz, L., & Baltz, R. H. (2016). Natural product discovery: past, present, and future. *Journal of Industrial Microbiology and Biotechnology, 43*(2-3), 155-176.

Katz, M., Hover, B. M., & Brady, S. F. (2016). Culture-independent discovery of natural products from soil metagenomes. *Journal of Industrial Microbiology and Biotechnology, 43*(2-3), 129-141.

Kautsar, S. A., Blin, K., Shaw, S., Navarro-Muñoz, J. C., Terlouw, B. R., van der Hooft, J. J., . . . Pascal Andreu, V. (2020). MIBiG 2.0: a repository for biosynthetic gene clusters of known function. *Nucleic Acids Research, 48*(D1), D454-D458.

Kaweewan, I., Komaki, H., Hemmi, H., Hoshino, K., Hosaka, T., Isokawa, G., . . . Kodani, S. (2019). Isolation and structure determination of a new cytotoxic peptide, curacozole, from Streptomyces curacoi based on genome mining. *The Journal of antibiotics, 72*(1), 1-7.

Kaysser, L., Bernhardt, P., Nam, S.-J., Loesgen, S., Ruby, J. G., Skewes-Cox, P., . . . Moore, B. S. (2012). Merochlorins A–D, cyclic meroterpenoid antibiotics biosynthesized in divergent pathways with vanadium-dependent chloroperoxidases. *Journal of the American Chemical Society, 134*(29), 11988-11991.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., . . . Duran, C. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics, 28*(12), 1647-1649.

Kelly, K. L., & Judd, D. B. (1965). ISCC-NBS color-name charts illustrated with centroid colors.

Kemung, H. M., Tan, L. T.-H., Khan, T. M., Chan, K.-G., Pusparajah, P., Goh, B.-H., & Lee, L.-H. (2018). *Streptomyces* as a prominent resource of future anti-MRSA drugs. *Frontiers in microbiology, 9*, 2221.

Kennedy, J., Baker, P., Piper, C., Cotter, P. D., Walsh, M., Mooij, M. J., . . . Ross, R. P. (2009). Isolation and analysis of bacteria with antimicrobial activities from the marine sponge *Haliclona simulans* collected from Irish waters. *Marine Biotechnology, 11*(3), 384-396.

Kepplinger, B., Morton-Laing, S., Seistrup, K. H., Marrs, E. C. L., Hopkins, A. P., Perry, J. D., . . . Allenby, N. E. E. (2017). Mode of action and heterologous expression of the natural product antibiotic vancoresmycin. *ACS chemical biology, 13*(1), 207-214.

Khater, S., Anand, S., & Mohanty, D. (2016). In silico methods for linking genes and secondary metabolites: the way forward. *Synthetic and systems biotechnology, 1*(2), 80-88.

Kim, J. H., Lee, N., Hwang, S., Kim, W., Lee, Y., Cho, S., . . . Cho, B.-K. (2021). Discovery of novel secondary metabolites encoded in actinomycete genomes through coculture. *Journal of Industrial Microbiology and Biotechnology, 48*(3-4), kuaa001.

Kluge, A. G., & Farris, J. S. (1969). Quantitative phyletics and the evolution of anurans. *Systematic Biology, 18*(1), 1-32.

Kohli, R. M., Takagi, J., & Walsh, C. T. (2002). The thioesterase domain from a nonribosomal peptide synthetase as a cyclization catalyst for integrin binding peptides. *Proceedings of the National Academy of Sciences, 99*(3), 1247-1252.

Komaki, H., Sakurai, K., Hosoyama, A., Kimura, A., Igarashi, Y., & Tamura, T. (2018). Diversity of nonribosomal peptide synthetase and polyketide synthase gene clusters among taxonomically close *Streptomyces* strains. *Scientific reports, 8*(1), 1-11.

Koo, B.-J., Kim, J.-M., Byun, S.-M., & Hong, S.-K. (1999). Optimal production conditions of *Streptomyces griseus* trypsin (SGT) in *Streptomyces lividans*. *BMB Reports, 32*(1), 86-91.

Korner-Nievergelt, F., Roth, T., Von Felten, S., Guélat, J., Almasi, B., & Korner-Nievergelt, P. (2015). *Bayesian data analysis in ecology using linear models with R, BUGS, and Stan*: Academic Press.

Kroiss, J., Kaltenpoth, M., Schneider, B., Schwinger, M.-G., Hertweck, C., Maddula, R. K., . . . Svatoš, A. (2010). Symbiotic streptomycetes provide antibiotic combination prophylaxis for wasp offspring. *Nature Chemical Biology, 6*(4), 261-263.

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution, 35*(6), 1547-1549.

Kurosawa, K., Ghiviriga, I., Sambandan, T., Lessard, P. A., Barbara, J. E., Rha, C., & Sinskey, A. J. (2008). Rhodostreptomycins, antibiotics biosynthesized following horizontal gene transfer from Streptomyces padanus to Rhodococcus fascians. *Journal of the American Chemical Society, 130*(4), 1126-1127.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017a). lmerTest package: tests in linear mixed effects models. *Journal of statistical software, 82*(13), 1-26.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017b). lmerTest package: tests in linear mixed effects models. *Journal of statistical software, 82*, 1-26.

Labeda, D. P. (2011). Multilocus sequence analysis of phytopathogenic species of the genus *Streptomyces*. *International Journal of Systematic and Evolutionary Microbiology, 61*(10), 2525-2531.

Labeda, D. P., Doroghazi, J., Ju, K.-S., & Metcalf, W. (2014). Taxonomic evaluation of *Streptomyces albus* and related species using multilocus sequence analysis and proposals to emend the description of *Streptomyces albus* and describe *Streptomyces pathocidini* sp. nov. *International Journal of Systematic and Evolutionary Microbiology, 64*(3), 894.

Lanyi, B. (1988). Classical and rapid identification methods for medically important bacteria. *Journal of Microbiological Methods, 19*, 1-67.

Lapaz, M. I., Cisneros, E. J., Pianzzola, M. J., & Francis, I. M. (2019). Exploring the Exceptional Properties of *Streptomyces*: A Hands-On Discovery of Natural Products. *The American Biology Teacher, 81*(9), 658-664.

Lee, H.-S., Shin, H. J., Jang, K. H., Kim, T. S., Oh, K.-B., & Shin, J. (2005). Cyclic peptides of the nocardamine class from a marine-derived bacterium of the genus Streptomyces. *Journal of Natural Products, 68*(4), 623-625.

Lee, I., Kim, Y. O., Park, S.-C., & Chun, J. (2016). OrthoANI: an improved algorithm and software for calculating average nucleotide identity. *International Journal of Systematic and Evolutionary Microbiology, 66*(2), 1100-1103.

Lee, J. Y., Lee, J. Y., Jung, H. W., & Hwang, B. K. (2005). *Streptomyces koyangensis* sp. nov., a novel actinomycete that produces 4-phenyl-3-butenoic acid. *International Journal of Systematic and Evolutionary Microbiology, 55*(1), 257-262.

Lee, N., Hwang, S., Kim, J., Cho, S., Palsson, B., & Cho, B.-K. (2020). Mini review: genome mining approaches for the identification of secondary metabolite biosynthetic gene clusters in Streptomyces. *Computational and Structural Biotechnology Journal, 18*, 1548-1556.

Lee, N., Hwang, S., Lee, Y., Cho, S., Palsson, B., & Cho, B.-K. (2019). Synthetic biology tools for novel secondary metabolite discovery in Streptomyces.

Lee, Y., Lee, N., Hwang, S., Kim, K., Kim, W., Kim, J., . . . Cho, B.-K. (2020). System-level understanding of gene expression and regulation for engineering secondary metabolite production in *Streptomyces*. *Journal of Industrial Microbiology & Biotechnology: Official Journal of the Society for Industrial Microbiology and Biotechnology, 47*(9-10), 739-752.

Lefort, V., Desper, R., & Gascuel, O. (2015). FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program. *Molecular Biology and Evolution, 32*(10), 2798-2800.

Lemon, J. (2006). Plotrix: a package in the red light district of R. R-News 6: 8–12. In.

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). Emmeans: Estimated marginal means, aka least-squares means. *R package version, 1*(1), 3.

Letzel, A. C., Li, J., Amos, G. C., Millán-Aguiñaga, N., Ginigini, J., Abdelmohsen, U. R., . . . Jensen, P. R. (2017). Genomic insights into specialized metabolism in the marine actinomycete Salinispora. *Environmental Microbiology, 19*(9), 3660-3673.

Li, J. W.-H., & Vederas, J. C. (2009). Drug discovery and natural products: end of an era or an endless frontier? *Science, 325*(5937), 161-165.

Li, W.-J., Zhang, L.-P., Xu, P., Cui, X.-L., Lu, Z.-T., Xu, L.-H., & Jiang, C.-L. (2002). *Streptomyces beijiangensis* sp. nov., a psychrotolerant actinomycete isolated from soil in China. *International Journal of Systematic and Evolutionary Microbiology, 52*(5), 1695-1699.

Li, X., Zhang, L., Huang, F., Zhao, J., Wang, H., Jiao, Y., . . . Xiang, W. (2021). Microbacterium helvum sp. nov., a novel actinobacterium isolated from cow dung. *Archives of Microbiology, 203*(6), 3287-3294.

Liles, M. R., Williamson, L. L., Rodbumrer, J., Torsvik, V., Goodman, R. M., & Handelsman, J. (2008). Recovery, purification, and cloning of high-molecular-weight DNA from soil microorganisms. *Applied and Environmental Microbiology, 74*(10), 3302-3305.

Linares-Otoya, L., Linares-Otoya, V., Armas-Mantilla, L., Blanco-Olano, C., Crüsemann, M., Ganoza-Yupanqui, M. L., . . . Schäberle, T. F. (2017). Identification and heterologous expression of the kocurin biosynthetic gene cluster. *Microbiology, 163*(10), 1409-1414.

Liria, C. W., & Kilikian, B. V. (1997). Kinetics of a recombinant protein production by *E. coli* BL21. *Nature reviews microbiology*, 172-178.

Liu, G., Chater, K. F., Chandra, G., Niu, G., & Tan, H. (2013). Molecular regulation of antibiotic biosynthesis in *Streptomyces*. *Microbiology and Molecular Biology Reviews, 77*(1), 112-143.

Liu, Q., Shen, Q., Bian, X., Chen, H., Fu, J., Wang, H., . . . Li, D. (2016). Simple and rapid direct cloning and heterologous expression of natural product biosynthetic gene cluster in Bacillus subtilis via Red/ET recombineering. *Scientific reports, 6*(1), 1-10.

Locey, K. J., & Lennon, J. T. (2016). Scaling laws predict global microbial diversity. *Proceedings of the National Academy of Sciences, 113*(21), 5970-5975.

Lodhi, A. F., Zhang, Y., Adil, M., & Deng, Y. (2018). Antibiotic discovery: combining isolation chip (iChip) technology and co-culture technique. *Applied Microbiology and Biotechnology, 102*(17), 7333-7341.

Lundy, T. A., Mori, S., & Garneau-Tsodikova, S. (2020). Lessons learned in engineering interrupted adenylation domains when attempting to create trifunctional enzymes from three independent monofunctional ones. *RSC Advances, 10*(56), 34299-34307.

Luo, Y., Cobb, R. E., & Zhao, H. (2014). Recent advances in natural product discovery. *Current Opinion in Biotechnology, 30*, 230-237.

Ma, M., Welch, R. D., & Garza, A. G. (2021). The σ54 system directly regulates bacterial natural product genes. *Scientific reports, 11*(1), 1-11.

Manteca, Á., & Yagüe, P. (2019). *Streptomyces* as a source of antimicrobials: novel approaches to activate cryptic secondary metabolite pathways. *Antimicrobials, Antibiotic Resistance, Antibiofilm Strategies and Activity Methods*, 119.

Martinez, A., Kolvek, S. J., Yip, C. L. T., Hopke, J., Brown, K. A., MacNeil, I. A., & Osburne, M. S. (2004). Genetically modified bacterial strains and novel bacterial artificial chromosome shuttle vectors for constructing environmental libraries and detecting heterologous natural products in multiple expression hosts. *Applied and Environmental Microbiology, 70*(4), 2452-2463.

Matsumoto, M., Hashimoto, Y., Saitoh, Y., Kumano, T., & Kobayashi, M. (2016). Development of nitrilase promoter-derived inducible vectors for Streptomyces. *Bioscience, Biotechnology, and Biochemistry, 80*(6), 1230-1237.

McCormick, J. R., & Flärdh, K. (2012). Signals and regulators that govern *Streptomyces* development. *FEMS Microbiology Reviews, 36*(1), 206-231.

McDonald, B. R., & Currie, C. R. (2017). Lateral gene transfer dynamics in the ancient bacterial genus *Streptomyces*. *MBio, 8*(3), e00644-00617.

McFall-Ngai, M., Hadfield, M. G., Bosch, T. C., Carey, H. V., Domazet-Lošo, T., Douglas, A. E., . . . Gilbert, S. F. (2013). Animals in a bacterial world, a new imperative for the life sciences. *Proceedings of the National Academy of Sciences, 110*(9), 3229-3236.

McLean, T. C., Wilkinson, B., Hutchings, M. I., & Devine, R. (2019). Dissolution of the disparate: co-ordinate regulation in antibiotic biosynthesis. *Antibiotics, 8*(2), 83.

McMahon, M. D., Guan, C., Handelsman, J., & Thomas, M. G. (2012). Metagenomic analysis of Streptomyces lividans reveals host-dependent functional expression. *Applied and Environmental Microbiology, 78*(10), 3622-3629.

Meier-Kolthoff, J. P., Auch, A. F., Klenk, H.-P., & Göker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics, 14*(1), 1-14.

Meier-Kolthoff, J. P., & Göker, M. (2019). TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nature communications, 10*(1), 1-10.

Meier-Kolthoff, J. P., Klenk, H.-P., & Göker, M. (2014). Taxonomic use of DNA G+ C content and DNA–DNA hybridization in the genomic age. *International Journal of Systematic and Evolutionary Microbiology, 64*(2), 352-356.

Meleshko, D., Mohimani, H., Tracanna, V., Hajirasouliha, I., Medema, M. H., Korobeynikov, A., & Pevzner, P. A. (2019). BiosyntheticSPAdes: reconstructing biosynthetic gene clusters from assembly graphs. *Genome Research, 29*(8), 1352-1362.

Michael, C. A., Dominey-Howes, D., & Labbate, M. (2014). The antimicrobial resistance crisis: causes, consequences, and management. *Frontiers in public health, 2*, 145.

Miller, B. R., & Gulick, A. M. (2016). Structural biology of nonribosomal peptide synthetases. In *Nonribosomal Peptide and Polyketide Biosynthesis* (pp. 3-29): Springer.

Miller, B. R., Sundlov, J. A., Drake, E. J., Makin, T. A., & Gulick, A. M. (2014). Analysis of the linker region joining the adenylation and carrier protein domains of the modular nonribosomal peptide synthetases. *Proteins: Structure, Function, and Bioinformatics, 82*(10), 2691-2702.

Milshteyn, A., Schneider, J. S., & Brady, S. F. (2014). Mining the metabiome: identifying novel natural products from microbial communities. *Chemistry & Biology, 21*(9), 1211-1223.

Minnikin, D., O'donnell, A., Goodfellow, M., Alderson, G., Athalye, M., Schaal, A., & Parlett, J. (1984). An integrated procedure for the extraction of bacterial isoprenoid quinones and polar lipids. *Journal of Microbiological Methods, 2*(5), 233-241.

Mizuno, C. M., Kimes, N. E., Lopez-Perez, M., Auso, E., Rodriguez-Valera, F., & Ghai, R. (2013). A hybrid NRPS-PKS gene cluster related to the bleomycin family of antitumor antibiotics in Alteromonas macleodii strains. *PLoS One, 8*(9), e76021.

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., & Kanehisa, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research, 35*(suppl_2), W182-W185.

Murli, S., Kennedy, J., Dayem, L. C., Carney, J. R., & Kealey, J. T. (2003). Metabolic engineering of Escherichia coli for improved 6-deoxyerythronolide B production. *Journal of Industrial Microbiology and Biotechnology, 30*(8), 500-509.

Myronovskyi, M., & Luzhetskyy, A. (2019). Heterologous production of small molecules in the optimized Streptomyces hosts. *Natural product reports, 36*(9), 1281-1294.

Myronovskyi, M., Rosenkränzer, B., Nadmid, S., Pujic, P., Normand, P., & Luzhetskyy, A. (2018). Generation of a cluster-free Streptomyces albus chassis strains for improved heterologous expression of secondary metabolite clusters. *Metabolic Engineering, 49*, 316-324.

Nair, A. G., Selvakumar, D., & Dhevendaran, K. (2011). Occurrence of sponges associated *Streptomyces* and its antimicrobial activity. *World journal of fish and marine sciences, 3*, 151-158.

Nakashima, Y., Mori, T., Nakamura, H., Awakawa, T., Hoshino, S., Senda, M., . . . Abe, I. (2018). Structure function and engineering of multifunctional non-heme iron dependent oxygenases in fungal meroterpenoid biosynthesis. *Nature communications, 9*(1), 1-10.

Nasrin, S., Ganji, S., Kakirde, K. S., Jacob, M. R., Wang, M., Ravu, R. R., . . . Mead, D. A. (2018). Chloramphenicol derivatives with antibacterial activity identified by functional metagenomics. *Journal of Natural Products, 81*(6), 1321-1332.

Nepal, K. K., & Wang, G. (2019). Streptomycetes: surrogate hosts for the genetic manipulation of biosynthetic gene clusters and production of natural products. *Biotechnology Advances, 37*(1), 1-20.

Neuwirth, E. (2014). RColorBrewer: ColorBrewer palettes. R package version 1.1-2. In.

Newman, D. J., & Cragg, G. M. (2012). Natural products as sources of new drugs over the 30 years from 1981 to 2010. *Journal of Natural Products, 75*(3), 311-335.

Nguyen, C. T., Dhakal, D., Pham, V. T. T., Nguyen, H. T., & Sohng, J.-K. (2020). Recent advances in strategies for activation and discovery/characterization of cryptic biosynthetic gene clusters in *Streptomyces*. *Microorganisms, 8*(4), 616.

Nikolouli, K., & Mossialos, D. (2012). Bioactive compounds synthesized by non-ribosomal peptide synthetases and type-I polyketide synthases discovered through genome-mining and metagenomics. *Biotechnology Letters, 34*(8), 1393-1403.

Niu, G., Li, L., Wei, J., & Tan, H. (2013). Cloning, heterologous expression, and characterization of the gene cluster required for gougerotin biosynthesis. *Chemistry & Biology, 20*(1), 34-44.

Nodwell, J. R., Yang, M., Kuo, D., & Losick, R. (1999). Extracellular complementation and the identification of additional genes involved in aerial mycelium formation in *Streptomyces coelicolor*. *Genetics, 151*(2), 569-584.

O'Brien, J., & Wright, G. D. (2011). An ecological perspective of microbial secondary metabolism. *Current Opinion in Biotechnology, 22*(4), 552-558.

Oh, D.-C., Scott, J. J., Currie, C. R., & Clardy, J. (2009). Mycangimycin, a polyene peroxide from a mutualist Streptomyces sp. *Organic letters, 11*(3), 633-636.

Olano, C., Méndez, C., & Salas, J. A. (2009). Antitumor compounds from actinomycetes: from gene clusters to new derivatives by combinatorial biosynthesis. *Natural product reports, 26*(5), 628-660.

Olano, C., Wilkinson, B., Sánchez, C., Moss, S. J., Sheridan, R., Math, V., . . . Oliynyk, M. (2004). Biosynthesis of the angiogenesis inhibitor borrelidin by *Streptomyces parvulus* Tü4055: cluster analysis and assignment of functions. *Chemistry & Biology, 11*(1), 87-97.

Ondov, B. D., Treangen, T. J., Melsted, P., Mallonee, A. B., Bergman, N. H., Koren, S., & Phillippy, A. M. (2016). Mash: fast genome and metagenome distance estimation using MinHash. *Genome biology, 17*(1), 1-14.

Pace, N. R. (1985). Analyzing natural microbial populations by rRNA sequences. *ASM news, 51*, 4-12.

Park, J.-H., Ahn, M.-Y., Kim, Y.-C., Kim, S.-A., Moon, Y.-H., Ahn, S.-G., & Yoon, J.-H. (2012). In vitro and in vivo antimicrobial effect of photodynamic therapy using a highly pure chlorin e6 against *Staphylococcus aureus* Xen29. *Biological and pharmaceutical bulletin, 35*(4), 509-514.

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research, 25*(7), 1043-1055.

Parsley, L. C., Linneman, J., Goode, A. M., Becklund, K., George, I., Goodman, R. M., . . . Liles, M. R. (2011). Polyketide synthase pathways identified from a metagenomic library are derived from soil Acidobacteria. *FEMS Microbiology Ecology, 78*(1), 176-187.

Parte, A. C. (2014). LPSN—list of prokaryotic names with standing in nomenclature. *Nucleic Acids Research, 42*(D1), D613-D616.

Peng, Q., Gao, G., Lü, J., Long, Q., Chen, X., Zhang, F., . . . Deng, Z. (2018). Engineered Streptomyces lividans strains for optimal identification and expression of cryptic biosynthetic gene clusters. *Frontiers in microbiology, 9*, 3042.

Pereira, F. (2019). Metagenomics: A gateway to drug discovery. In *Advances in Biological Science Research* (pp. 453-468): Elsevier.

Petersen, L.-E., Kellermann, M. Y., & Schupp, P. J. (2020). *Secondary metabolites of marine microbes: From natural products chemistry to chemical ecology.* Paper presented at the

YOUMARES 9-The Oceans: Our Research, Our Future: Proceedings of the 2018 conference for YOUng MArine RESearcher in Oldenburg, Germany.

Pettit, R. K. (2009). Mixed fermentation for natural product drug discovery. *Applied Microbiology and Biotechnology, 83*(1), 19-25.

Pham, J. V., Yilma, M. A., Feliz, A., Majid, M. T., Maffetone, N., Walker, J. R., . . . Song, M. C. (2019). A review of the microbial production of bioactive natural products and biologics. *Frontiers in microbiology, 10*, 1404.

Pimentel-Elardo, S. M., Buback, V., Gulder, T. A., Bugni, T. S., Reppart, J., Bringmann, G., . . . Hentschel, U. (2011). New tetromycin derivatives with anti-trypanosomal and protease inhibitory activities. *Marine Drugs, 9*(10), 1682-1697.

Procópio, R. E. d. L., Silva, I. R. d., Martins, M. K., Azevedo, J. L. d., & Araújo, J. M. d. (2012). Antibiotics produced by *Streptomyces*. *Brazilian Journal of Infectious Diseases, 16*(5), 466-471.

Qian, Z., Bruhn, T., D'Agostino, P. M., Herrmann, A., Haslbeck, M., Antal, N., . . . Gulder, T. A. (2019). Discovery of the streptoketides by direct cloning and rapid heterologous expression of a cryptic PKS II gene cluster from Streptomyces sp. Tu 6314. *The Journal of organic chemistry, 85*(2), 664-673.

Qin, Z., Munnoch, J. T., Devine, R., Holmes, N. A., Seipke, R. F., Wilkinson, K. A., . . . Hutchings, M. I. (2017). Formicamycins, antibacterial polyketides produced by Streptomyces formicae isolated from African Tetraponera plant-ants. *Chemical science, 8*(4), 3218-3227.

Rashad, F. M., Fathy, H. M., El-Zayat, A. S., & Elghonaimy, A. M. (2015). Isolation and characterization of multifunctional *Streptomyces* species with antimicrobial, nematicidal and phytohormone activities from marine environments in Egypt. *Microbiological Research, 175*, 34-47.

Rateb, M. E., Houssen, W. E., Arnold, M., Abdelrahman, M. H., Deng, H., Harrison, W. T., . . . Ferguson, G. (2011). Chaxamycins A–D, bioactive ansamycins from a hyper-arid desert Streptomyces sp. *Journal of Natural Products, 74*(6), 1491-1499.

Rateb, M. E., Houssen, W. E., Harrison, W. T., Deng, H., Okoro, C. K., Asenjo, J. A., . . . Ebel, R. (2011). Diverse metabolic profiles of a Streptomyces strain isolated from a hyper-arid environment. *Journal of Natural Products, 74*(9), 1965-1971.

Rebets, Y., Kormanec, J., Luzhetskyy, A., Bernaerts, K., & Anné, J. (2017). Cloning and expression of metagenomic DNA in Streptomyces lividans and subsequent fermentation for optimized production. In *Metagenomics* (pp. 99-144): Springer.

Reddy, B. V. B., Milshteyn, A., Charlop-Powers, Z., & Brady, S. F. (2014). eSNaPD: a versatile, web-based bioinformatics platform for surveying and mining natural product biosynthetic diversity from metagenomes. *Chemistry & Biology, 21*(8), 1023-1033.

Richter, M., & Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the national academy of sciences, 106*(45), 19126-19131.

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

Rigali, S., Titgemeyer, F., Barends, S., Mulder, S., Thomae, A. W., Hopwood, D. A., & Van Wezel, G. P. (2008). Feast or famine: the global regulator DasR links nutrient stress to antibiotic production by *Streptomyces*. *EMBO reports, 9*(7), 670-675.

Robinson, S. L., Piel, J., & Sunagawa, S. (2021). A roadmap for metagenomic enzyme discovery. *Natural Product Reports*.

Sabarathnam, B., Manilal, A., Sujith, S., Kiran, G. S., Selvin, J., Thomas, A., & Ravji, R. (2010). Role of sponge associated actinomycetes in the marine phosphorous biogeochemical cycles. *American-Eurasian journal of agriculture and environmental sciences, 8*, 253-256.

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution, 4*(4), 406-425.

Salam, N., Jiao, J.-Y., Zhang, X.-T., & Li, W.-J. (2020). Update on the classification of higher ranks in the phylum Actinobacteria. *International Journal of Systematic and Evolutionary Microbiology, 70*(2), 1331-1355.

Santana-Pereira, A. L., Sandoval-Powers, M., Monsma, S., Zhou, J., Santos, S. R., Mead, D. A., & Liles, M. R. (2020). Discovery of novel biosynthetic gene cluster diversity from a soil metagenomic library. *Frontiers in microbiology, 11*.

Sara, M. (1971). Ultrastructural aspects of the symbiosis between two species of the genus *Aphanocapsa* (*Cyanophyceae*) and *Ircinia variabilis* (*Demospongiae*). *Marine Biology, 11*(3), 214-221.

Sasser, M. (1990). *Identification of bacteria by gas chromatography of cellular fatty acids*: MIDI technical note 101. Newark, DE: MIDI inc.

Schatz, A., Bugle, E., & Waksman, S. A. (1944). Streptomycin, a Substance Exhibiting Antibiotic Activity Against Gram-Positive and Gram-Negative Bacteria.∗. *Proceedings of the Society for Experimental Biology and Medicine, 55*(1), 66-69.

Scherlach, K., & Hertweck, C. (2009). Triggering cryptic natural product biosynthesis in microorganisms. *Organic & biomolecular chemistry, 7*(9), 1753-1760.

Schipper, C., Hornung, C., Bijtenhoorn, P., Quitschau, M., Grond, S., & Streit, W. (2009). Metagenome-derived clones encoding two novel lactonase family proteins involved in biofilm inhibition in Pseudomonas aeruginosa. *Applied and Environmental Microbiology, 75*(1), 224-233.

Schmidt, E., Obraztsova, A., Davidson, S., Faulkner, D., & Haygood, M. (2000). Identification of the antifungal peptide-containing symbiont of the marine sponge *Theonella swinhoei* as a novel δ-proteobacterium,"Candidatus *Entotheonella palauensis*". *Marine Biology, 136*(6), 969-977.

Schmidt, T. M., DeLong, E. F., & Pace, N. R. (1991). Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *Journal of Bacteriology, 173*(14), 4371-4378.

Schneemann, I., Kajahn, I., Ohlendorf, B., Zinecker, H., Erhard, A., Nagel, K., . . . Imhoff, J. F. (2010). Mayamycin, a cytotoxic polyketide from a Streptomyces strain isolated from the marine sponge Halichondria panicea. *Journal of Natural Products, 73*(7), 1309-1312.

Schneider, Y. K. (2021). Bacterial Natural Product Drug Discovery for New Antibiotics: Strategies for Tackling the Problem of Antibiotic Resistance by Efficient Bioprospecting. *Antibiotics, 10*(7), 842.

Schumann, P. (2011). Peptidoglycan structure. In *Methods in Microbiology* (Vol. 38, pp. 101-129): Elsevier.

Science, R. A. (2009). DIG application manual for filter hybridization. *Roche Applied Science, Indianapolis, IN.*

Seipke, R. F., Kaltenpoth, M., & Hutchings, M. I. (2012). *Streptomyces* as symbionts: an emerging and widespread theme? *FEMS Microbiology Reviews, 36*(4), 862-876.

Seyedsayamdost, M. R. (2019). Toward a global picture of bacterial secondary metabolism. *Journal of Industrial Microbiology and Biotechnology, 46*(3-4), 301-311.

Sharrar, A. M., Crits-Christoph, A., Méheust, R., Diamond, S., Starr, E. P., & Banfield, J. F. (2020). Bacterial secondary metabolite biosynthetic potential in soil varies with phylum, depth, and vegetation type. *Mbio, 11*(3), e00416-00420.

Shen, B. (2015). A new golden age of natural products drug discovery. *Cell, 163*(6), 1297-1300.

Shepherd, M. D., Kharel, M. K., Bosserman, M. A., & Rohr, J. (2010). Laboratory maintenance of *Streptomyces* species. *Current protocols in microbiology, 18*(1), 10E-11.

Shin, D., Byun, W. S., Moon, K., Kwon, Y., Bae, M., Um, S., . . . Oh, D.-C. (2018). Coculture of marine Streptomyces sp. with Bacillus sp. produces a new piperazic acid-bearing cyclic peptide. *Frontiers in chemistry, 6*, 498.

Shirling, E. T., & Gottlieb, D. (1966). Methods for characterization of *Streptomyces* species. *International Journal of Systematic Bacteriology, 16*(3), 313-340.

Shivlata, L., & Tulasi, S. (2015). Thermophilic and alkaliphilic Actinobacteria: biology and potential applications. *Frontiers in microbiology, 6*, 1014.

Shrivastava, S. R., Shrivastava, P. S., & Ramasamy, J. (2018). World health organization releases global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics. *Journal of Medical Society, 32*(1), 76.

Singh, S., Gurusiddaiah, S., & Whalen, J. (1985). Treponemycin, a nitrile antibiotic active against *Treponema hyodysenteriae*. *Antimicrobial agents and chemotherapy, 27*(2), 239-245.

Sivalingam, P., Hong, K., Pote, J., & Prabakar, K. (2019). Extreme environment Streptomyces: potential sources for new antibacterial and anticancer drug leads? *International Journal of Microbiology, 2019*.

Skinnider, M. A., Merwin, N. J., Johnston, C. W., & Magarvey, N. A. (2017). PRISM 3: expanded prediction of natural product chemical structures from microbial genomes. *Nucleic Acids Research, 45*(W1), W49-W54.

Smibert, R., Krieg, N., Gerhardt, P., Murray, R., & Wood, W. (1994). *Methods for general and molecular bacteriology*. Washington, DC: American Society for Microbiology.

Song, Y., Li, Q., Liu, X., Chen, Y., Zhang, Y., Sun, A., . . . Ju, J. (2014). Cyclic hexapeptides from the deep South China Sea-derived Streptomyces scopuliridis SCSIO ZJ46 active against pathogenic Gram-positive bacteria. *Journal of Natural Products, 77*(8), 1937-1941.

Stackebrandt, E., & Goebel, B. M. (1994). Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *International Journal of Systematic and Evolutionary Microbiology, 44*(4), 846-849.

Staley, J. T., & Konopka, A. (1985). Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annual Review of Microbiology, 39*(1), 321-346.

Stankey, R. J., Stephens, H. E., Sandoval-Powers, M., Liles, M. R., & Mead, D. A. (2022). Dual-inducible heterologous expression of cloned biosynthetic gene clusters in Actinobacteria. *[manuscript in preparation]*.

Starcevic, A., Zucko, J., Simunkovic, J., Long, P. F., Cullum, J., & Hranueli, D. (2008). ClustScan: an integrated program package for the semi-automatic annotation of modular biosynthetic gene clusters and in silico prediction of novel chemical structures. *Nucleic Acids Research, 36*(21), 6882-6892.

Stefani, F. O., Bell, T. H., Marchand, C., de la Providencia, I. E., El Yassimi, A., St-Arnaud, M., & Hijri, M. (2015). Culture-dependent and-independent methods capture different microbial community fractions in hydrocarbon-contaminated soils. *PloS one, 10*(6), e0128272.

Stein, J. L., Marsh, T. L., Wu, K. Y., Shizuya, H., & DeLong, E. F. (1996). Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *Journal of Bacteriology, 178*(3), 591-599.

Stevens, D. C., Conway, K. R., Pearce, N., Villegas-Peñaranda, L. R., Garza, A. G., & Boddy, C. N. (2013). Alternative sigma factor over-expression enables heterologous expression of a type II polyketide biosynthetic pathway in Escherichia coli. *PloS one, 8*(5), e64858.

Stevens, D. C., Hari, T. P., & Boddy, C. N. (2013). The role of transcription in heterologous expression of polyketides in bacterial hosts. *Natural Product Reports, 30*(11), 1391-1411.

Stevenson, L. J., Bracegirdle, J., Liu, L., Sharrock, A. V., Ackerley, D. F., Keyzers, R. A., & Owen, J. G. (2021). Metathramycin, a new bioactive aureolic acid discovered by heterologous expression of a metagenome derived biosynthetic pathway. *RSC Chemical Biology, 2*(2), 556-567.

Stewart, E. J. (2012). Growing unculturable bacteria. *Journal of Bacteriology, 194*(16), 4151-4160.

Stratigopoulos, G., & Cundliffe, E. (2002). Expression analysis of the tylosin-biosynthetic gene cluster: pivotal regulatory role of the tylQ product. *Chemistry & Biology, 9*(1), 71-78.

Subramani, R., & Aalbersberg, W. (2012). Marine actinomycetes: an ongoing source of novel bioactive metabolites. *Microbiological Research, 167*(10), 571-580.

Subramani, R., & Aalbersberg, W. (2013). Culturable rare Actinomycetes: diversity, isolation and marine natural product discovery. *Applied Microbiology and Biotechnology, 97*(21), 9291-9321.

Sujarit, K., Kudo, T., Ohkuma, M., Pathom-Aree, W., & Lumyong, S. (2016). *Streptomyces palmae* sp. nov., isolated from oil palm (*Elaeis guineensis*) rhizosphere soil. *International Journal of Systematic and Evolutionary Microbiology, 66*(10), 3983-3988.

Sulavik, M. C., Houseweart, C., Cramer, C., Jiwani, N., Murgolo, N., Greene, J., . . . Hare, R. (2001). Antibiotic susceptibility profiles of Escherichia coli strains lacking multidrug efflux pump genes. *Antimicrobial agents and chemotherapy, 45*(4), 1126-1136.

Sung, A. A., Gromek, S. M., & Balunas, M. J. (2017). Upregulation and identification of antibiotic activity of a marine-derived Streptomyces sp. via co-cultures with human pathogens. *Marine Drugs, 15*(8), 250.

Taitt, C. R., Leski, T. A., Stockelman, M. G., Craft, D. W., Zurawski, D. V., Kirkup, B. C., & Vora, G. J. (2014). Antimicrobial resistance determinants in Acinetobacter baumannii isolates taken from military treatment facilities. *Antimicrobial agents and chemotherapy, 58*(2), 767-781.

Talbot, G. H., Jezek, A., Murray, B. E., Jones, R. N., Ebright, R. H., Nau, G. J., . . . America, I. D. S. o. (2019). The Infectious Diseases Society of America's 10×'20 initiative (10 new systemic antibacterial agents US Food and Drug Administration approved by 2020): is 20×'20 a possibility? *Clinical Infectious Diseases, 69*(1), 1-11.

Tanwar, J., Das, S., Fatima, Z., & Hameed, S. (2014). Multidrug resistance: an emerging crisis. *Interdisciplinary perspectives on infectious diseases, 2014*.

Taylor, M. W., Radax, R., Steger, D., & Wagner, M. (2007). Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. *Microbiology and Molecular Biology Reviews, 71*(2), 295-347.

Taylor, M. W., Schupp, P. J., Dahllöf, I., Kjelleberg, S., & Steinberg, P. D. (2004). Host specificity in marine sponge-associated bacteria, and potential implications for marine microbial diversity. *Environmental Microbiology, 6*(2), 121-130.

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

Tianero, M. D., Balaich, J. N., & Donia, M. S. (2019). Localized production of defence chemicals by intracellular symbionts of Haliclona sponges. *Nature microbiology, 4*(7), 1149-1159.

Tindall, B. (1990a). A comparative study of the lipid composition of *Halobacterium saccharovorum* from various sources. *Systematic and Applied Microbiology, 13*(2), 128-130.

Tindall, B. (1990b). Lipid composition of *Halobacterium lacusprofundi*. *FEMS Microbiology Letters, 66*(1-3), 199-202.

Toyohara, M., Nagata, A., Hayano, K., & Abe, J. (1969). Study on the antituberculous activity of tuberactinomycin, a new antimicrobial drug. *American Review of Respiratory Disease, 100*(2), 228-230.

Traxler, M. F., Watrous, J. D., Alexandrov, T., Dorrestein, P. C., & Kolter, R. (2013). Interspecies interactions stimulate diversification of the Streptomyces coelicolor secreted metabolome. *MBio, 4*(4), e00459-00413.

Trindade, M., van Zyl, L. J., Navarro-Fernández, J., & Abd Elrazak, A. (2015). Targeted metagenomics as a tool to tap into marine natural product diversity for the discovery and production of drug candidates. *Frontiers in microbiology, 6*, 890.

Tyc, O., Song, C., Dickschat, J. S., Vos, M., & Garbeva, P. (2017). The ecological role of volatile and soluble secondary metabolites produced by soil bacteria. *Trends in Microbiology, 25*(4), 280-292.

Ulvatne, H., Haukland, H. H., Samuelsen, Ø., Krämer, M., & Vorland, L. H. (2002). Proteases in Escherichia coli and Staphylococcus aureus confer reduced susceptibility to lactoferricin B. *Journal of Antimicrobial Chemotherapy, 50*(4), 461-467.

Um, S., Choi, T. J., Kim, H., Kim, B. Y., Kim, S.-H., Lee, S. K., . . . Oh, D.-C. (2013). Ohmyungsamycins A and B: cytotoxic and antimicrobial cyclic peptides produced by

Streptomyces sp. from a volcanic island. *The Journal of organic chemistry, 78*(24), 12321-12329.

Unson, M. D., Holland, N., & Faulkner, D. J. (1994). A brominated secondary metabolite synthesized by the cyanobacterial symbiont of a marine sponge and accumulation of the crystalline metabolite in the sponge tissue. *Marine Biology, 119*(1), 1-11.

van Bergeijk, D. A., Terlouw, B. R., Medema, M. H., & van Wezel, G. P. (2020). Ecology and genomics of Actinobacteria: new concepts for natural product discovery. *Nature Reviews Microbiology, 18*(10), 546-558.

Van der Meij, A., Worsley, S. F., Hutchings, M. I., & van Wezel, G. P. (2017). Chemical ecology of antibiotic production by actinomycetes. *FEMS Microbiology Reviews, 41*(3), 392-416.

Vandamme, P., Pot, B., Gillis, M., De Vos, P., Kersters, K., & Swings, J. (1996). Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiological reviews, 60*(2), 407-438.

Vartoukian, S. R., Palmer, R. M., & Wade, W. G. (2010). Strategies for culture of 'unculturable'bacteria. *FEMS Microbiology Letters, 309*(1), 1-7.

Vecherskii, M., Semenov, M., Lisenkova, A., & Stepankov, A. (2021). Metagenomics: A New Direction in Ecology. *Biology Bulletin, 48*(3), S107-S117.

Ventura, M., Canchaya, C., Tauch, A., Chandra, G., Fitzgerald, G. F., Chater, K. F., & van Sinderen, D. (2007). Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. *Microbiology and Molecular Biology Reviews, 71*(3), 495-548.

Vimberg, V., Zieglerová, L., Buriánková, K., Branny, P., & Balíková Novotná, G. (2020). VanZ reduces the binding of lipoglycopeptide antibiotics to Staphylococcus aureus and Streptococcus pneumoniae cells. *Frontiers in microbiology, 11*, 566.

Wakabayashi, T., Kageyama, R., Naruse, N., Tsukahara, N., Funahashi, Y., Kitoh, K., & Watanabet, Y. (1997). Borrelidin is an angiogenesis inhibitor; disruption of angiogenic capillary vessels in a rat aorta matrix culture model. *The journal of antibiotics, 50*(8), 671-676.

Waksman, S. A. (1953). Streptomycin: background, isolation, properties, and utilization. *Science, 118*(3062), 259-266.

Waksman, S. A., & Henrici, A. T. (1943). The nomenclature and classification of the actinomycetes. *Journal of Bacteriology, 46*(4), 337.

Waksman, S. A., & Woodruff, H. B. (1940). Bacteriostatic and bactericidal substances produced by a soil actinomyces. *Proceedings of the Society for Experimental Biology and Medicine, 45*(2), 609-614.

Waksman, S. A., & Woodruff, H. B. (1942). Streptothricin, a new selective bacteriostatic and bactericidal agent, particularly active against Gram-negative bacteria. *Proceedings of the Society for Experimental Biology and Medicine, 49*(2), 207-210.

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., . . . Young, S. K. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS one, 9*(11), e112963.

Walsh, C. T. (2004). Polyketide and nonribosomal peptide antibiotics: modularity and versatility. *Science, 303*(5665), 1805-1810.

Wan, X.-F., Xu, D., Kleinhofs, A., & Zhou, J. (2004). Quantitative relationship between synonymous codon usage bias and GC composition across unicellular genomes. *BMC Evolutionary Biology, 4*(1), 1-11.

Wang, G.-Y.-S., Graziani, E., Waters, B., Pan, W., Li, X., McDermott, J., . . . Davies, J. (2000). Novel natural products from soil DNA libraries in a streptomycete host. *Organic letters, 2*(16), 2401-2404.

Wang, H., Fewer, D. P., Holm, L., Rouhiainen, L., & Sivonen, K. (2014). Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. *Proceedings of the National Academy of Sciences, 111*(25), 9259-9264.

Wang, M., Carver, J. J., Phelan, V. V., Sanchez, L. M., Garg, N., Peng, Y., . . . Luzzatto-Knaan, T. (2016). Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nature Biotechnology, 34*(8), 828-837.

Wang, W., Yang, T., Li, Y., Li, S., Yin, S., Styles, K., . . . Yang, K. (2016). Development of a synthetic oxytetracycline-inducible expression system for Streptomycetes using de novo characterized genetic parts. *ACS synthetic biology, 5*(7), 765-773.

Wang, W., Zheng, G., & Lu, Y. (2021). Recent advances in strategies for the cloning of natural product biosynthetic gene clusters. *Frontiers in Bioengineering and Biotechnology, 9*.

Weber, T., Rausch, C., Lopez, P., Hoof, I., Gaykova, V., Huson, D., & Wohlleben, W. (2009). CLUSEAN: a computer-based framework for the automated analysis of bacterial

secondary metabolite biosynthetic gene clusters. *Journal of Biotechnology, 140*(1-2), 13-17.

Webster, N. S., Luter, H. M., Soo, R. M., Botté, E. S., Simister, R. L., Abdo, D., & Whalan, S. (2013). Same, same but different: symbiotic bacterial associations in GBR sponges. *Frontiers in microbiology, 3*, 444.

Webster, N. S., Negri, A. P., Webb, R. I., & Hill, R. T. (2002). A spongin-boring α-proteobacterium is the etiological agent of disease in the Great Barrier Reef sponge *Rhopaloeides odorabile*. *Marine Ecology Progress Series, 232*, 305-309.

Webster, N. S., & Taylor, M. W. (2012). Marine sponges and their microbial symbionts: love and other relationships. *Environmental Microbiology, 14*(2), 335-346.

Webster, N. S., & Thomas, T. (2016). The sponge hologenome. *MBio, 7*(2), e00135-00116.

Wei, J., He, L., & Niu, G. (2018). Regulation of antibiotic biosynthesis in actinomycetes: perspectives and challenges. *Synthetic and systems biotechnology, 3*(4), 229-235.

Wei, R.-B., Xi, T., Li, J., Wang, P., Li, F.-C., Lin, Y.-C., & Qin, S. (2011). Lobophorin C and D, new kijanimicin derivatives from a marine sponge-associated actinomycetal strain AZS17. *Marine Drugs, 9*(3), 359-368.

Wei, Y., Zhang, L., Zhou, Z., & Yan, X. (2018). Diversity of gene clusters for polyketide and nonribosomal peptide biosynthesis revealed by metagenomic analysis of the yellow sea sediment. *Frontiers in microbiology, 9*, 295.

Weigel, B. L., & Erwin, P. M. (2017). Effects of reciprocal transplantation on the microbiome and putative nitrogen cycling functions of the intertidal sponge, *Hymeniacidon heliophila*. *Scientific reports, 7*(1), 1-12.

Wickham, H. (2007a). Reshaping data with the reshape package. *Journal of statistical software, 21*(12), 1-20.

Wickham, H. (2007b). Reshaping data with the reshape package. *Journal of statistical software, 21*, 1-20.

Wickham, H. (2016). *ggplot2-Elegant Graphics for Data Analysis. Springer International Publishing*. Cham, Switzerland.

Wickham, H. (2017a). Tidyverse: Easily install and load the 'tidyverse'. *R package version, 1*(1), 2017.

Wickham, H. (2017b). tidyverse: Easily Install and Load the 'Tidyverse'. R package version 1.2. 1.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. A., François, R., . . . Hester, J. (2019). Welcome to the Tidyverse. *Journal of open source software, 4*(43), 1686.

Wickham, H., Francois, R., Henry, L., & Müller, K. (2015). dplyr: A grammar of data manipulation. *R package version 0.4, 3*, p156.

Wilke, C. O. (2016). cowplot: streamlined plot theme and plot annotations for 'ggplot2'. *CRAN Repos, 2*, R2.

Wilke, C. O. (2019). cowplot: streamlined plot theme and plot annotations for 'ggplot2'. *R package version 0.9, 4*.

Wilkins, D., & Kurtz, Z. (2019). gggenes: Draw gene arrow maps in "ggplot2". *R package version 0.4. 0*.

Wilkinson, C., & Fay, P. (1979). Nitrogen fixation in coral reef sponges with symbiotic cyanobacteria. *Nature, 279*(5713), 527-529.

Wilkinson, C. R. (1983). Net primary productivity in coral reef sponges. *Science, 219*(4583), 410-412.

Willemse, J., Borst, J. W., de Waal, E., Bisseling, T., & van Wezel, G. P. (2011). Positive control of cell division: FtsZ is recruited by SsgB during sporulation of *Streptomyces*. *Genes & Development, 25*(1), 89-99.

Williams, G. J. (2013). Engineering polyketide synthases and nonribosomal peptide synthetases. *Current Opinion in Structural Biology, 23*(4), 603-612.

Woese, C. R., & Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceedings of the National Academy of Sciences, 74*(11), 5088-5090.

Wohlleben, W., Mast, Y., Stegmann, E., & Ziemert, N. (2016). Antibiotic drug discovery. *Microbial biotechnology, 9*(5), 541-548.

World Health Organization. (2020). *Antibacterial Agents in Clinical and Preclinical Development: An Overview and Analysis*. Retrieved from

Wright, F., & Bibb, M. J. (1992). Codon usage in the G+ C-rich *Streptomyces* genome. *Gene, 113*(1), 55-65.

Wu, C., Shang, Z., Lemetre, C., Ternei, M. A., & Brady, S. F. (2019). Cadasides, calcium-dependent acidic lipopeptides from the soil metagenome that are active against multidrug-resistant bacteria. *Journal of the American Chemical Society, 141*(9), 3910-3919.

Wu, Q., Zhang, G., Wang, B., Li, X., Yue, S., Chen, J., . . . Wang, H. (2018). Production and identification of Inthomycin B produced by a Deep-Sea sediment-derived Streptomyces sp. YB104 based on cultivation-dependent approach. *Current Microbiology, 75*(7), 942-951.

Xu, F., Nazari, B., Moon, K., Bushin, L. B., & Seyedsayamdost, M. R. (2017). Discovery of a cryptic antifungal compound from Streptomyces albus J1074 using high-throughput elicitor screens. *Journal of the American Chemical Society, 139*(27), 9203-9212.

Xu, M., Wang, Y., Zhao, Z., Gao, G., Huang, S.-X., Kang, Q., . . . Deng, Z. (2016). Functional genome mining for metabolites encoded by large gene clusters through heterologous expression of a whole-genome bacterial artificial chromosome library in Streptomyces spp. *Applied and Environmental Microbiology, 82*(19), 5795-5805.

Yagüe, P., López-García, M. T., Rioseras, B., Sánchez, J., & Manteca, Á. (2013). Pre-sporulation stages of *Streptomyces* differentiation: state-of-the-art and future perspectives. *FEMS Microbiology Letters, 342*(2), 79-88.

Ye, J., Zhu, Y., Hou, B., Wu, H., & Zhang, H. (2019). Characterization of the bagremycin biosynthetic gene cluster in Streptomyces sp. Tü 4128. *Bioscience, Biotechnology, and Biochemistry, 83*(3), 482-489.

Yin, J., Hoffmann, M., Bian, X., Tu, Q., Yan, F., Xia, L., . . . Fu, J. (2015). Direct cloning and heterologous expression of the salinomycin biosynthetic gene cluster from Streptomyces albus DSM41398 in Streptomyces coelicolor A3 (2). *Scientific reports, 5*(1), 1-8.

Yoon, S.-H., Ha, S.-M., Kwon, S., Lim, J., Kim, Y., Seo, H., & Chun, J. (2017). Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *International Journal of Systematic and Evolutionary Microbiology, 67*(5), 1613.

Zerikly, M., & Challis, G. L. (2009). Strategies for the discovery of new natural products by genome mining. *ChemBioChem, 10*(4), 625-633.

Zettler, J., Xia, H., Burkard, N., Kulik, A., Grond, S., Heide, L., & Apel, A. K. (2014). New aminocoumarins from the rare actinomycete Catenulispora acidiphila DSM 44928:

identification, structure elucidation, and heterologous production. *ChemBioChem, 15*(4), 612-621.

Zhang, F., Jonas, L., Lin, H., & Hill, R. T. (2019). Microbially mediated nutrient cycles in marine sponges. *FEMS Microbiology Ecology, 95*(11), fiz155.

Zhang, F., Vicente, J., & Hill, R. T. (2014). Temporal changes in the diazotrophic bacterial communities associated with Caribbean sponges *Ircinia stroblina* and *Mycale laxissima*. *Frontiers in microbiology, 5*, 561.

Zhang, J. J., Tang, X., Zhang, M., Nguyen, D., & Moore, B. S. (2017). Broad-host-range expression reveals native and host regulatory elements that influence heterologous antibiotic production in Gram-negative bacteria. *MBio, 8*(5), e01291-01217.

Zhang, L., Song, T., Wu, J., Zhang, S., Yin, C., Huang, F., . . . Zhang, Y. (2020). Antibacterial and cytotoxic metabolites of termite-associated Streptomyces sp. BYF63. *The Journal of Antibiotics, 73*(11), 766-771.

Zhang, Z., Du, C., de Barsy, F., Liem, M., Liakopoulos, A., van Wezel, G. P., . . . Rozen, D. E. (2020). Antibiotic production in Streptomyces is organized by a division of labor through terminal genomic differentiation. *Science advances, 6*(3), eaay5781.

Ziemert, N., Alanjary, M., & Weber, T. (2016). The evolution of genome mining in microbes–a review. *Natural product reports, 33*(8), 988-1005.

Ziemert, N., Lechner, A., Wietz, M., Millán-Aguiñaga, N., Chavarria, K. L., & Jensen, P. R. (2014). Diversity and evolution of secondary metabolism in the marine actinomycete genus Salinispora. *Proceedings of the National Academy of Sciences, 111*(12), E1130-E1139.

Ziemert, N., Podell, S., Penn, K., Badger, J. H., Allen, E., & Jensen, P. R. (2012). The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PloS one, 7*(3), e34064.

Ziermann, R., & Betlach, M. C. (1999). Recombinant polyketide synthesis in Streptomyces: engineering of improved host strains. *BioTechniques, 26*(1), 106-110.

Zotchev, S. B. (2014). Genomics-based insights into the evolution of secondary metabolite biosynthesis in actinomycete bacteria. In *Evolutionary biology: genome evolution, speciation, coevolution and origin of life* (pp. 35-45): Springer.

Zucchi, T. D., Kim, B.-y., Kshetrimayum, J. D., Weon, H.-Y., Kwon, S.-W., & Goodfellow, M. (2012). *Streptomyces brevispora* sp. nov. and *Streptomyces laculatispora* sp. nov.,

actinomycetes isolated from soil. *International Journal of Systematic and Evolutionary Microbiology, 62*(3), 478-483.