Strathclyde Institute of Pharmacy and Biomedical Sciences

Approaches for rejuvenating the natural product discovery process from *Streptomyces*

Darren James Scobie

Thesis presented in fulfillment of the requirement for the degree of Doctor of Philosophy

2023

<u>Declaration</u>

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree. The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis. The figures used in this thesis are a result of work carried out by the author.

Signed:

Date: 15/05/2023

Acknowledgements

I'll try and keep this brief, but I'd like to start (what will likely turn into an absolute rollercoaster/waffle) by thanking the most important beings on the planet: my dogs Lubo and Marnie. It's absolutely wild how much I adore you. I'd do anything for you. You could say my entire life is dedicated to keeping you alive. Help.

Aside from my dogs (love you), the biggest thanks have to go to my family. Michelle, I've written this part over and over again but can't find the right words to describe how much your support, encouragement and kindness have meant to me over the course of this PhD. Any time I've doubted myself – and let's face it, that happened practically every day – you were there to help me through it. It perhaps says a lot about the work that's outlined in this thesis, but the hardest fact within these pages is that I couldn't have got through this without you. To my Mum and my brother Richard, your belief in me is something that pushes me every day to try and do better and I hope I've been able to make you proud. Again, without your support I would never have been able to finish this. To Rosemary and Martin, thank you for being the most caring in-laws I could hope for. You make Michelle and I's lives so much easier every day and you're two of the most inspiring people I've ever known. Thank you to my Gran and Granda. I know they're not around to see me finish this, but they helped shape the person I am today.

An enormous thank you to my supervisor Dr Kate Duncan. I feel like at this point you should be sick of the sight of me, but here I am post-doc-ing away. You've given me the opportunity to make the best of myself by seeing potential in me that I would never have thought existed and I'll forever appreciate that. Thank you to my secondary supervisor Dr Paul Herron for the invaluable advice at pretty much every step of the way during my PhD. With that in mind, thank you to all of the academic and technical staff on the 6th floor. Thank you to the Duncan Group!

Lily, we started our PhDs at the same time and spent pretty much every day together since. Maybe it's healthy co-dependency. Maybe it's Stockholm Syndrome. Whatever you want to call it, you're one of the best people on the planet. Parra (genuinely the worlds nicest, kindest and most thoughtful person, even if you did wave a hot dog in my face), Ally (I feel like most of our time working together and subsequent friendship has consisted of shouting general life complaints at each other but I love you for it), Marla, Laia, Alejandro, Connie, Diana and all the various masters and undergrad students that have helped me along the way – Thank you. To Adriana, Eilidh, Elmira and Molly – along with Lily we all started at the same time, and I genuinely couldn't have asked for a better group of people to go through this whole weird adventure with. I will continue to get emotional about this on nights out together. I cannot be stopped. To Liam, Dave, Charlie, Tom (He), Tom (Ha), Logan, Robyn, Stuart, Holly, Johnny, Lis, Ainsley, Becca, James, Jack, JB, John Munnoch, Dan, Katherine, Angelika, Kaisha, Anna, Anil, Derya, Khedidja – thank you for being amazing and helping to foster such a lovely working atmosphere during some seriously tough times.

An astronomical thank you has to go to the lecturers and staff in the science department at Forth Valley College. I started there a decade ago doing classes at night without so much as a higher to my name, and I owe a huge debt of gratitude to everyone that taught me back then for their patience and kindness. I think it's super important to promote further education pathways for people like me who didn't fare well at school when they were younger. I can't imagine the number of talented scientists that haven't had the chance to develop simply because we aren't really actively encouraged to follow an academic pathway later in life.

Extra-large shout out to all of my mates – Dr James McKean first and foremost, who completed a PhD in Catholicism or abbeys or the reformation or something (anything not microbiology-related is obviously irrelevant) and was basically my human scream pillow for a long time for anything related to how mad the world of academia is. Thanks to Cammi and Thomas, my two

oldest mates – love you! Thanks to Bobby, Danny, Fraser, Sean, Pete, Rob, Traynor, Bob, Dean Loud Man, Taylor, Matt, Kaz and Ben. You're the best.

Huge shout out to all my teammates, coaches, board members and any of the various volunteers at Glasgow Saints Football Club. I signed for the club just after starting my PhD and I'm still there despite the global pandemic, several shredded ankle ligaments, a cracked ankle and a mysterious rare blood disorder. Top down, the club have been amazing. Rest In Power Timo Chiwaula, my teammate who we lost in 2021. Check in on your pals. I would also like to thank Clyde Football Club, the team I've supported all my days, but I won't because you've been absolutely shocking during my whole PhD, and it's just added to the stress.

Serious shout out time. A huge thank you has to go to the NHS. Since the start of my PhD, I have had some… weird medical issues. Firstly, I snapped all my ankle ligaments and cracked the same ankle playing football. I was helped straight away in the hospital and with some physio and some great advice it's now holding up as well as a 34-year-old ankle should. After (stupidly) doing a very long charity walk to raise money for the charities Glasgow Saints FC are associated with, my legs really weren't in a good way (again). I saw my GP who quickly arranged for a blood test, which showed that I had dangerously low platelet levels and I was subsequently diagnosed with a rare blood disorder called Immune Thrombocytopenia Purpura. It was a scary time for me and my family, but all of the various NHS staff and haematologists I dealt with were some of the most kind and supportive humans I've ever interacted with and made things a lot easier to digest. In March 2022, Michelle's mum Rosemary was diagnosed with pancreatic cancer. She was treated swiftly and thanks to the care, attention, and action of NHS staff, she's still with us. The NHS really is incredible, and we shouldn't let it be torn apart.

Finally, I'm going to give myself a pat on the back – Well done, I guess? You probably did fine.

Table of Contents

List of Figures

List of Tables

Chapter 1 – Introduction

1.1 Antimicrobial resistance

1.1.1 The golden age of antimicrobial discovery

In 1935, Sir Alexander Fleming gifted a sample of *Penicillium notatum* to his friend and colleague at St. Mary's Hospital, Douglas Macleod (**Figure 1.1**). This was a generous gift indeed as seven years previously, a chance event led to his discovery of penicillin from this strain (Fleming, 1944) (Gaynes, 2017), heralding the golden age of antimicrobials (Julian Davies, 2006). Despite this era of discovery, rampant misuse and overzealous prescription of these drugs has caused myriad issues (Ramachandran *et al*., 2019). In fact, the United States collectively prescribed 22 doses of antibiotics per person in 2010 (Ventola, 2015). Following the 1960's, the frequency of antimicrobial metabolite discovery fell rapidly as rediscovery of known products from traditional screening methods became ever more frequent (Jones *et al*., 2017). The period between 1983 and 2007 saw a 75% decrease in FDA-approved antimicrobials (Boucher *et al*., 2009) and only two new classes of antimicrobials have reached the market since 1962 (Coates, Halls, & Hu, 2011).

Figure 1.1 – A sample of penicillium mould. Sample of *Penicillium notatum* gifted from Sir Alexander Fleming to his colleague Douglas Macleod from St Mary's Hospital in London in 1935 (SMG, 1997).

1.1.2 Early antimicrobial discovery methods

This golden age of antimicrobial discovery, between the 1940s and 1960s, yielded the vast majority of antimicrobial metabolites through manual screening of natural sources, such as soil Actinomycetota (Valiquette & Laupland, 2015). These discoveries were facilitated by an antimicrobial discovery pipeline pioneered by Selman Waksman, first outlined in the Albert Schatz-led paper describing streptomycin (Schatz, Bugle, & Waksman, 1944). The pipeline was beautifully simple – soil-derived Actinomycetota were screened for antimicrobial activity against susceptible test organisms by measuring zones of inhibition on overlay plates, mimicking the accidental discovery of penicillin in a very deliberate sense (Lewis, 2013) (Fleming, 1944).

The story of the pipeline development and its aftermath is as dramatic as the effect it had on the field. Selman Waksman had success isolating both actinomycin and streptothricin, but the metabolites were too toxic to be used by humans. In 1943 his student, the aforementioned Albert Schatz, observed that two strains of *Streptomyces griseus* (then known as *Actinomyces griseus*) inhibited many Gram-negative organisms, as well as *Mycobacterium tuberculosis*. A patent for streptomycin was then granted to both Schatz and Waksman, however the latter bullied the former into signing over his share to Rutgers University whilst deceiving him by stating that he had already done so. Yet Waksman had not and struck a deal with Rutgers to receive 20% of net royalties. Schatz attempted to sue but Waksman effectively halted his career and diminished Schatz contributions to the discovery of streptomycin and as such, Waksman was the sole recipient of the resultant Nobel Prize (Rawlins, 2012) (Pringle, 2012).

1.1.3 Antimicrobial resistance

In 1940, five years before penicillin was available over the counter in the US, penicillin resistance was identified when it was reported that an *E. coli* strain could cause its inactivation

by producing penicillinase (Abraham & Chain, 1940) (**Figure 1.2**). Antimicrobial Resistance (AMR) and the emergence of Multi-Drug Resistant (MDR) bacteria has since become one of the largest threats to human health (Llor & Bjerrum, 2014) (Klemm, Wong, & Dougan, 2018). In India alone, in excess of 58,000 infant deaths per year are associated with MDR microorganisms (Barker, Brown, Ahsan, Sengupta, & Safdar, 2017). Indeed, the much-quoted O'Neill report published in 2016 attributed 700,000 deaths per year worldwide to AMR (O'Neill, 2016) however a recent study found that in 2019, 1.27 million deaths were directly attributable to bacterial AMR (Murray *et al*., 2022) - a staggering increase of 81% in three years.

AMR has spread via diverse means. In the developing world, conventional treatment processes are only utilised for 8 to 28% of wastewater, resulting in a high level of antibiotic resistance genes making their way into natural bodies of water (Hazra, Joshi, Williams, & Watts, 2022). This environmental issue is compounded by overuse in sectors such as the food industry, with around 80% of all antibiotics sold in the US used in animal agriculture to promote growth and prevent disease (Martin, Thottathil, & Newman, 2015). This is exemplified by *mcr-3*, the gene that confers transferable colistin resistance. It was first identified in China and has now spread globally through the international trading of contaminated pork (Xu *et al*., 2018) (Y. Y. Liu *et al*., 2016). A lack of waste control in the food industry has also been found to cause the spread of AMR genes. A recent study concluded that the displacement of the *cmy-2* gene in *E. coli* by CTX-M-15 (a similar gene which confers cephalosporin resistance) is likely occurring *ex vivo* in the excreted urine of Washington State dairy cattle (Avillan *et al*., 2022).

In terms of human health, Methicillin-resistant *Staphylococcus aureus* (MRSA) has emerged as a major World Health Organisation (WHO) target pathogen. MRSA was identified in 1961 (**Figure 1.2**) and has been the cause of 11,285 deaths per annum in the US alone - more than AIDS, Parkinson's disease, emphysema and homicide combined (Llor & Bjerrum, 2014) (Ventola, 2015). The Actinomycetota-derived last-resort antibiotic vancomycin, first isolated from *Amycolatopsis orientalis* (McCormick, McGuire, Pittenger, Pittenger, & Stark, 1955), is

the last defence against multi-drug resistant MRSA (Ventola, 2015) but vancomycin resistance in *enterococci* was observed in the mid-80's and the first vancomycin-resistant *S. aureus* isolate was detected in 2002 (Chang *et al*., 2003) (**Figure 1.2**). Misuse of therapy with last resort antibiotics has led to reduced effectiveness of these drugs (Woudt *et al*., 2017), such as the development of carbapenem resistance in several species of the pneumonia-causing *Acinetobacter* (Ventola, 2015). Increases in population displacement and refugee movement has also contributed to the increase in AMR. Incidences of tuberculosis occurring in North America and western Europe, where cases of the disease are minimal, is directly correlated to the influx of displaced populations from tuberculosis-hyperendemic regions (MacPherson *et al*., 2009). Transmission of these MDR strains among refugee populations is oftentimes facilitated by substandard housing, hygiene and healthcare infrastructures in origin countries as these groups are often displaced as a result of civil war, as well as factors such as extremely poor hygiene on trips to their destination (Maltezou, Theodoridou, & Daikos, 2017). For example, in an international study of AMR strains of *Salmonella typhae*, it was found that in South Asia, *S. typhae* with high level fluoroquinolone resistance have been frequently displacing strains with fewer mutations before spreading globally (da Silva *et al*., 2022). Interestingly, with increasingly fewer ways to stem the tide of AMR and MDR organisms, we must turn to another bacterial phylum - Actinomycetota.
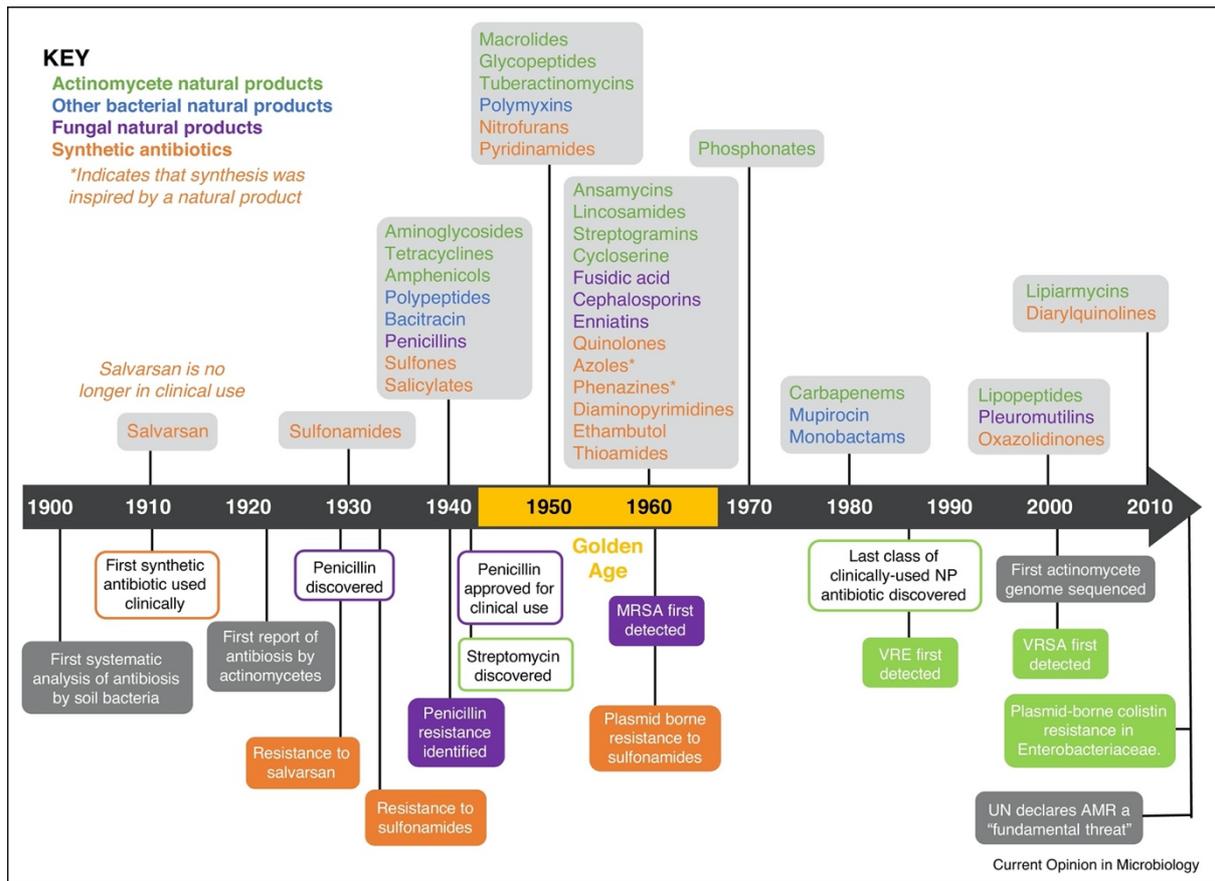
Figure 1.2 – Timeline by decade of clinical implementation of antibiotic classes.

This includes the date of the first reports of drug resistant strains of methicillin-resistant *S. aureus* (MRSA), vancomycin-resistant *enterococci* (VRE), vancomycin-resistant *S. aureus* (VRSA) and plasmid-borne colistin resistance in *Enterobacteriaceae* (Matthew I. Hutchings, Andrew W. Truman, & Barrie Wilkinson, 2019).

1.2 Actinomycetota

1.2.1 Actinomycetota

Actinomycetota is a one of the largest and most diverse phyla within the Bacteria domain, comprising of over 250 genera including *Streptomyces*, *Micromonospora, Bifidobacterium* and *Pseudonocardia* (Ludwig *et al*., 2012). The phylum consists of Gram-positive bacteria with high G+C DNA content which ranges from 51% in corynebacteria genomes to upwards of 70% in the genomes of the plant commensal *Frankia* sp. (Ventura *et al*., 2007). Actinomycetota also exhibit diverse morphological properties, with *Arthrobacter* sp. able to undergo a change from rods, which predominate in young cultures, to cocci in late-stage batch cultures (Mulder & Antheunisse, 1963). The name '*Actinomycete'* is derived from the Greek word meaning 'ray fungus' due to their ability to form a mycelium consisting of hyphae that are morphologically similar to filamentous fungi, which may yield spores (Procopio, Silva, Martins, Azevedo, & Araujo, 2012) (Wiliams, 1990). Actinomycetota genomes encode a high level of biosynthetic diversity, with over 17,000 Gene Cluster Families (GCFs) detected in a study of 1,185,995 biosynthetic gene clusters (Gavriilidou *et al*., 2022). As a consequence of this biosynthetic diversity, Actinomycetota are prolific producers of antimicrobial specialised metabolites. Members of this phylum produce 64% of all known natural product antibiotic classes (Matthew I. Hutchings *et al*., 2019) with over 80% of all known antibiotics originating from Actinomycetota species (Barka *et al*., 2016). Of particular interest in this regard is the genus *Streptomyces.*

1.2.2 *Streptomyces*

Streptomycetaceae, a family of Gram-positive bacteria within the Actinomycetota phylum, was initially described upon the isolation of *Streptomyces griseus* (A. Schatz & Waksman, 1945). *Streptomyces* is the largest genus in terms of number of species within the family *Streptomycetaceae.* A recent study used a molecular clock to estimate that *Streptomyces* is

around 380 million years old – as ancient as the early land vertebrates (McDonald & Currie, 2017). Regarding clinical medicine, this genus is a key source of important metabolites, accounting for 39% of all microbial metabolites (Bérdy, 2012) and two thirds of all clinically relevant specialised metabolites (Yagüe, López-García, Rioseras, Sánchez, & Manteca, 2013). The first three antibiotics isolated from Actinomycetota were in fact all isolated from strains of *Streptomyces*: actinomycin from the aptly named *Streptomyces antibioticus* (Selman A. Waksman & Woodruff, 1940), streptothricin from *Streptomyces lavendulae* (S. A. Waksman & Woodruff, 1942) and streptomycin from *Streptomyces griseus* (Albert Schatz & Waksman, 1944). Perhaps these strains are so well studied in part due to their ability to survive in diverse and often hostile ecosystems, which is testament to their unique growth cycle.

1.2.3 *Streptomyces* life cycle

The *Streptomyces* life cycle is unusual for bacteria as it facilitates cell survival through the formation of spores (Chater, 2016). The formation of aerial hyphae and early cell differentiation events are tightly controlled by *bld* genes (Elliot, Bibb, Buttner, & Leskiw, 2001), called as such because their deletion results in a bald mutant devoid of aerial hyphae (Pope, Green, & Westpheling, 1996). On the other hand, another class of life cycle regulatory genes, *whi*, control the maturation of spores from spore chains and as a result, *whi* mutants lack spore pigment (Kelemen *et al*., 1998). Interestingly, in recent studies cyclic-di-GMP has been observed interacting with both the transcription factors of *bld* and *whi* genes and therefore controlling spore development and differentiation (McLean *et al*., 2019). The cycle begins with the germination of spores, which develop a compartmentalised mycelium (Manteca & Yague, 2018). The germination process varies between species of *Streptomyces* with a few undergoing rapid and almost universal germination, others germinate slowly with more complex germination behaviour and a higher number of spores that do not germinate at all (Bobek, Smidova, & Cihak, 2017). Some of the cells in the early substrate mycelium undergo a round of programmed cell death. The vegetative hyphal stage follows, with occasional cross-

walling dividing the hyphae, making *Streptomyces* a rare multi-cellular prokaryote (van Wezel & McDowall, 2011). *Streptomyces* vegetative hyphae grow by tip extension, where cell wall components are added to the tips of apical cells as opposed to the system employed by most bacteria where growth is achieved via extension of the lateral cell wall (Jakimowicz & van Wezel, 2012). When essential nutrients are depleted, the remaining segments begin to grow as a multinucleated mycelium known as late substrate mycelium. This form then undergoes a second round of programmed cell death, differentiating into aerial hyphae where spore chains which become dispersed, thus beginning a new cycle (**Figure 1.3**) (Barka *et al*., 2016) (Manteca & Yague, 2018). It is worth noting that the cycle differs when *Streptomyces* sp. are grown in liquid cultures. After early-stage mycelial growth, pellets form, with programmed cell death taking place in the core of the pellets and late-stage mycelium forming on the periphery (**Figure 1.4**) (Manteca & Yague, 2018).

Actinomycetota spore morphology is diverse, and they can form chains or single cells on the substrate mycelium, aerial mycelium, or both. For example, *Micromonospora* and *Thermoactinomyces* spores form on the substrate mycelium directly. In contrast, *Streptomyces* spores grow from aerial mycelium. In some instances, sporangia or flagella are formed. Naming conventions can also be incredibly literal, as in *Spirilospora* which has spores that develop in a spiral confirmation, or *Ampullariella* with its spores developing in ampules connected to aerial hyphae (**Figure 1.5**) (Barka *et al*., 2016). Spore dispersal can be facilitated in many ways. Odours containing geosmin and 2-MIB, metabolites emitted by *Streptomyces* with synthesis under direct control of sporulation specific transcription factors, attract the model springtail *Folsomia candida* by inducing an electrophysiological response in its antennae. The *Streptomyces* colonies are then consumed by *F. candida* which spreads spores via faecal pellets (Becher *et al*., 2020). *Streptomyces* spores are also capable of utilizing the motility machinery of other soil bacteria for transportation to favoured niches such as plant roots, as evidenced by their adherence to the flagella of the motile *Bacillus subtilis* (Alise R. Muok, Dennis Claessen, & Ariane Briegel, 2021).
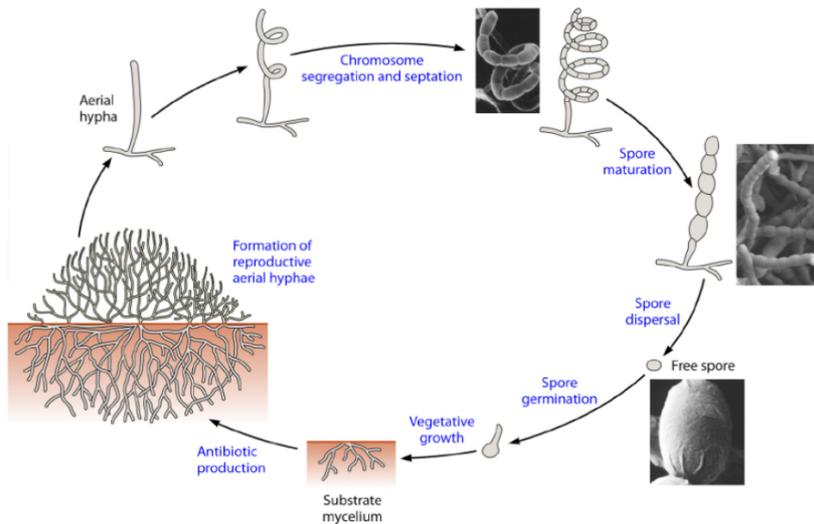
Figure 1.3 – The life cycle of sporulating Actinomycetota. Schematic showing the life cycle of sporulating Actinomycetota through each stage in growth as depicted in (Barka *et al*., 2016).
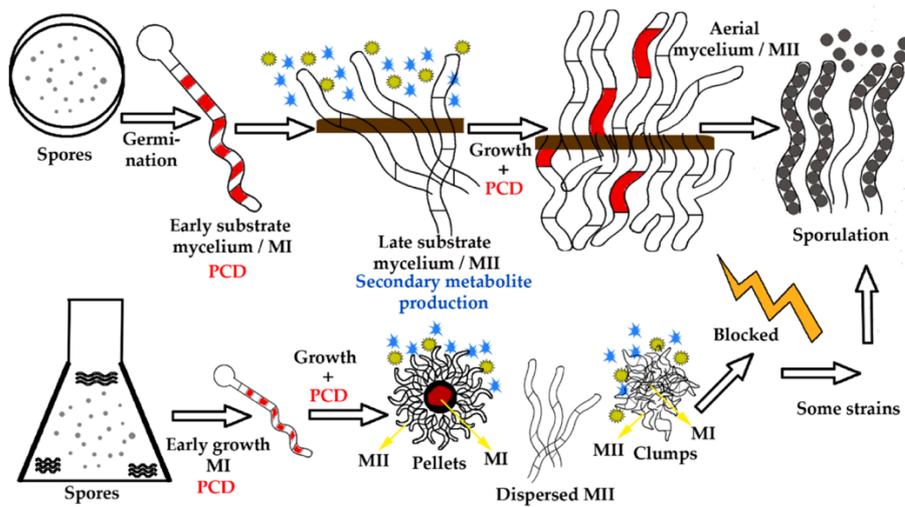
Figure 1.4 – Schematic representing the life cycle of *Streptomyces* in liquid media.

The life cycle in liquid media is shown compared to the life cycle of a similar strain of *Streptomyces* grown on solid media. In liquid cultures, hyphae form clumps and pellets, with sporulation blocked in many strains (Manteca & Yague, 2018).
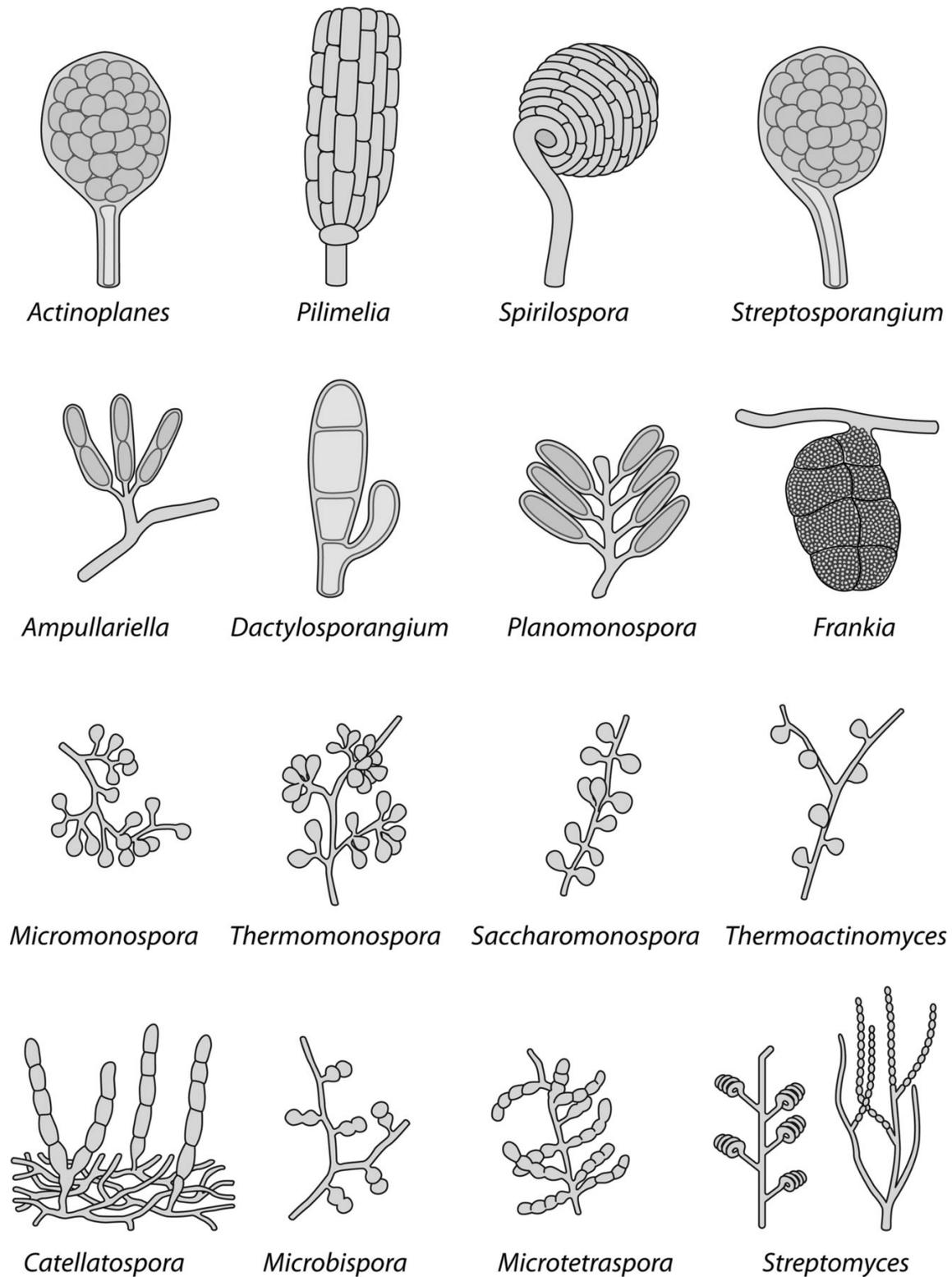
Figure 1.5 – Spore chain variation in Actinomycetota. Schematic drawings of diversity in Actinomycetota spore chain morphology (Barka *et al*., 2016).

1.3 *Streptomyces* specialised metabolism

1.3.1 *Streptomyces* primary metabolism

Bacterial natural products can be categorised as primary or specialised metabolites (Nwokeji, Enodiana, Ezenweani, Osasere, & Abiola, 2016), with a recent shift in nomenclature altering 'secondary' to 'specialised' to emphasise the biological and ecological importance of their products (Chevrette *et al*., 2020). Primary metabolism is essential for survival in organisms such as bacteria, and encompasses processes such as respiration and aspects of carbon and nitrogen metabolism (Singh, Kumar, Mittal, & Mehta, 2017). Primary metabolites include monomers of macromolecules such as amino acids and nucleotides, vitamins and acids such as citric acid and acetic acid (Sanchez & Demain, 2008). However, from a drug discovery perspective, the more complex specialised metabolism is more compelling.

1.3.2 *Streptomyces* specialised metabolism

Specialised metabolism involves the production of metabolites which, whilst not essential for growth, confer an adaptational advantage by enabling interactions with the producer's ecological niche (Demain & Fang, 2000) (van Wezel & McDowall, 2011). It is believed that the majority of specialised metabolites, such as actinorhodin from *S. coelicolor* A3(2) (Gramajo, Takano, & Bibb, 1993), are produced during the stationary phase of growth as a response to nutrient depletion in media (Čihák *et al*., 2017) or upon the formation of pellets within liquid culture (Yagüe *et al*., 2013). On the unusual linear chromosome of *Streptomyces*, genes considered to be required for viability, i.e., primary metabolism, tend to be clustered in the chromosome core, close to the origin of replication (oriC) whilst the diverse and more species-specific genes relating to specialised metabolism are often located on flanking chromosomal arms (Bentley *et al*., 2002). In *S. coelicolor* the core region containing essential genes was

approximately 4.9 Mbp in length, whilst the 'left arm' was 1.5 Mbp and the 'right arm' was 2.3 Mbp (Bentley *et al*., 2002). In a recent study, it was observed that in 320,263 genes laterally acquired by *Streptomyces*, specialised metabolite gene clusters are overrepresented in lateral gene transfer events (McDonald & Currie, 2017), further enforcing that these sections of the genome are disposable.

1.3.3 Biosynthetic gene clusters

Genes involved in the biosynthesis of specialised metabolites are often clustered in a single, coregulated genomic region known as a Biosynthetic Gene Cluster (BGC) (Cimermancic *et al*., 2014), which can be described as a physical grouping of genes which encode enzymes responsible for specialised metabolite biosynthesis (Medema *et al*., 2015). BGC expression is often subject to stringent regulation and in some cases, as with the regulation of actinorhodin in *S. coelicolor* by ActII-ORF4, a single regulatory gene controls the production of the antibiotic (Gramajo *et al*., 1993). In terms of size, *Streptomyces* BGCs vary widely. Towards the lower end of the scale are the likes of the cremeomycin BGC in *Streptomyces cremeus* NRRL3241 at only 18 kilobase pairs (kbp) (Waldman, Pechersky, Wang, Wang, & Balskus, 2015) and the goadsporin BGC in *Streptomyces* sp. TP-A0584, at only 14 kbp (Haginaka *et al*., 2014). On the other side of the scale, the chaxamycin BGC in *Streptomyces leewenhoekii* heterologously expressed in *S. coelicolor* at 80.2 kbp (Castro *et al*., 2015). Regardless of size, there are some common architectural features.

BGCs often have a highly modular structure (Del Carratore *et al*., 2019) with observations involving larger BGCs, such as everninomicin, indicating that they evolve via smaller sub-clusters merging together (Medema, Cimermancic, Sali, Takano, & Fischbach, 2014). As well as enzymes responsible for specialised metabolite production, BGCs may contain pathway-specific genes as the entire cluster is often regulated by pathway-specific transcription factors where the coding gene may be found within the cluster itself (Keller, Turner, & Bennett, 2005).

One important clue which may tell us about metabolite production is the presence of resistance genes within the cluster to prevent suicide (P. N. Tran, M. R. Yen, C. Y. Chiang, H. C. Lin, & P. Y. Chen, 2019) and modification enzymes and transporters to facilitate the produced specialised metabolite leaving the cell (Graham-Taylor, Kamphuis, & Derbyshire, 2020). The diversity of gene function within BGCs is illustrated in the minimal formicamycin BGC, which contains 24 genes, including genes for three methyltransferases, multidrug resistance proteins and ion exchangers (Devine *et al*., 2021). If we are to isolate and study the products of BGCs, activity must be elicited.

1.3.4 Elicitation of cryptic or silent BGCs

The advent of genome sequencing technology has illustrated that on average *Streptomyces* species possess large genomes, encoding between 20 and 50 specialised metabolite BGCs whilst it is estimated that only around 3% of the specialised metabolites encoded within bacterial genomes have been characterised experimentally (Gavriilidou *et al*., 2022). This indicates that the majority of BGCs are either cryptic or silent (Z. Liu, Zhao, Huang, & Luo, 2021). These terms are often incorrectly assumed to be interchangeable. If a BGC is expressed but the product cannot be observed, the product is cryptic; if the BGC remains unexpressed it is considered to be silent; if the BGC is unidentified and its product can be observed, the biosynthesis of the product is cryptic; a truly cryptic natural product is unobserved and produced from an unidentified BGC (Hoskisson & Seipke, 2020). This issue must be overcome using elicitation.

Strategies have been employed to elicit specialised metabolite production from these cryptic or silent BGCs. One of the more common methods is to employ the One Strain Many Compounds (OSMAC) approach, building on the idea that one bacterial strain is capable of producing many structurally diverse metabolites depending on culture conditions (Bode, Bethe, Höfs, & Zeeck, 2002). As such, controlling culture conditions makes it possible to

target, or optimise the production of, specific specialised metabolites. OSMAC has been used to prioritise the production of many unknown natural products from Actinomycetota. After identifying homologues of various lassopeptide precursor and immunity proteins using genome mining in *Burkholderia thailandensis* E264, a novel lassopeptide named capistruin (**Figure 1.6**) was successfully isolated using OSMAC to tailor the target media and optimise its production (Knappe *et al.*, 2008). OSMAC was also recently used to maximise chemical production from *Streptomyces* 26D9-414, resulting in the identification of the discovery of two novel cytotoxic cyclodipeptides (D. Zhang *et al.*, 2022). However, OSMAC is not the only method of eliciting specialised metabolite production.
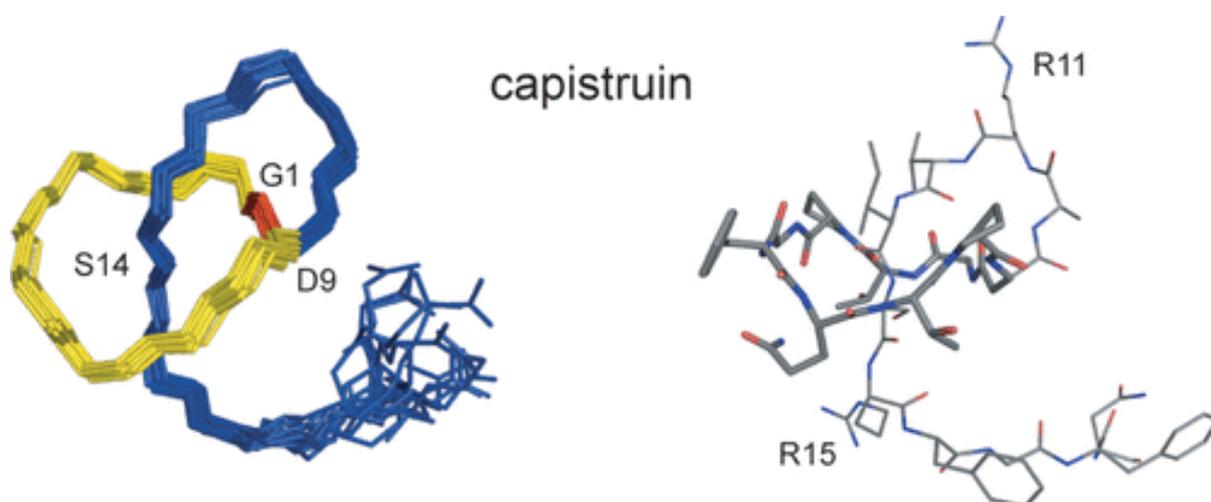
Figure 1.6 – Capistruin structure. The structure of the novel lassopeptide capistruin in solution shown in stereoview (left) and coloured by elements with the macrolactam ring highlighted (right) (Knappe *et al.*, 2008).

A further method for elicitation is to attempt to imitate the complex natural environment through coculture techniques (Zarins-Tutt *et al.*, 2016). Diffusible signalling molecules of low molecular mass are secreted by one cell and received by another, resulting in chemical elicitation (Baral, Akhgari, & Metsa-Ketela, 2018) but such interactions do not occur in standard laboratory conditions where strains are grown in isolation. In coculture with *Streptomyces coelicolor*

A3(2), production of the siderophore myxochelin was enhanced in *Myxococcus xanthus* leading to *M. xanthus* dominating iron scavenging and triggering iron restriction in *S. coelicolor* A3(2). These conditions caused the activation of the actinorhodin pathway and a novel actinorhodin export system in *S. coelicolor* A3(2) (Lee *et al*., 2020). Similar iron restriction conditions increased the expression of 21 specialised metabolite BGCs in other *Streptomyces* species (Lee *et al*., 2020). In a study linking a bioactivity-targeted approach coupled with coculture, antibacterial activity against *Bacillus subtilis* was induced in two *Streptomyces* isolates from soil when cocultured with the fungus *Schizophyllum commune* (Nicault *et al*., 2021).

1.4 Specialised metabolite classes

1.4.1 Polyketide Synthases (PKS)

Streptomyces sp. produce many classes of medically and industrially important natural products. Polyketide Synthases (PKS) are one of the most important BGC classes and polyketide natural products boast a remarkable range of structural and functional diversity (Staunton & Weissman, 2001). PKSs biosynthesise polyketides from short acyl-CoA units (Salo *et al*., 2016) (H. Chen & Du, 2016) and most can be grouped into three types based on the architecture of their biosynthetic machinery. These are Type I, Type II and Type III (B. Wang, Guo, Huang, & Zhao, 2020) with each type harbouring additional variants, such as the noniterative Type I PKS where each module lacks the acyltransferase domain (Cheng, Coughlin, Lim, & Shen, 2009).

Type I PKSs (T1PKS) are minimally comprised of the three domains required to catalyse one chain extension cycle - acetyltransferase (AT), keto synthase (KS) and acyl carrier protein (ACP) domains - as well as a subset of reducing domains working as a modular assembly line (Sabatini *et al*., 2018). Each individual module is grouped into one of two classes - *cis*-AT, where each module contains all three of the essential PKS domains (AT, KS, ACP) and *trans*-

AT, where a free-standing AT (often shared with other modules) transacylates an extender unit onto the ACP domain (Helfrich & Piel, 2016). The peptide modules are then joined via intermodular linkers (Gokhale, Tsuji, Cane, & Khosla, 1999). In cases where the PKS spans several polypeptides, noncovalent interactions link docking domains between modules (Broadhurst, Nietlispach, Wheatcroft, Leadlay, & Weissman, 2003). In the specific case of genes involved in the production of the erythromycin precursor 6-deoxyerythronolide B (6-dEB), the cluster is split into six modules across three Open Reading Frames (ORFs). Each ORF consists of two modules and codes for a large multienzyme polypeptide of around 350 kDa (Staunton & Weissman, 2001).

Interestingly, Actinomycetota are the only known group of organisms that utilise Type II PKS systems for the biosynthesis of polyketides (Hertweck, Luzhetskyy, Rebets, & Bechthold, 2007). Type II PKSs (T2PKS) differ from T1PKS with each protein existing separately, transiently interacting as opposed to forming a large megaenzyme and primarily coding for the production of aromatic polyketides (Sattely, Fischbach, & Walsh, 2008). A set of three enzymes has been found in all characterised T2PKS systems so far. The minimal T2PKS consists of two ketosynthase units (KS_α and KS_β) and an ACP unit to anchor the chain of polyketides and generally these genes are architecturally grouped together in a KS_α/KS_β/ACP formation (Shen, 2000). Condensation of an acyl starter unit with malonyl-CoA extender units is catalysed by the minimal T2PKS unit. The KS_α subunit catalyses the formation of C-C bonds, whilst the KS_β subunit is involved in loading malonyl-CoA and generating acetyl KS from the decarboxylation of ACP (Bisang *et al.*, 1999). The KS_β subunit is also determines of the length of the final carbon chain and as such it has been named CLF, or Chain Length Factor (R McDaniel, Ebert-Khosla, Fu, Hopwood, & Khosla, 1994).

Type III PKS (T3PKS) are structurally simpler than T1PKS and T2PKS, consisting of homodimers formed by self-contained enzymes (D. Yu, Xu, Zeng, & Zhan, 2012). In bacteria,

T3PKS catalyse the decarboxylation and condensation of malonyl-CoA molecules to form an intermediate poly-β-ketomethylene intermediate molecule, which in turn undergoes subsequent decarboxylation, cyclisation and dehydration (Kalaitzis, Hamano, Nilsen, & Moore, 2003). When Bentley *et al.* sequenced the *S. coelicolor* A3(2) genome, three ORFs coding for putative T3PKSs were observed. Two of these T3PKS were not linked to any known *S. coelicolor* A3(2) metabolites (Bentley *et al.*, 2002). It has since been reported that one of those T3PKS, known as Gcs, is required for germicidin biosynthesis (Song *et al.*, 2006).

Polyketides are a diverse group of specialised metabolites spanning useful antibiotics to harmful carcinogens. One such antibiotic is geldanamycin (**Figure 1.7 a**), which was initially purified from *Streptomyces hygroscopicus* broth in 1970 (DeBoer, Meulman, Wnuk, & Peterson, 1970) and binds to members of the protein family Hsp90 (Ochel, Eichhorn, & Gademann, 2001). The Hsp90 family are chaperone proteins that are overproduced in many human cancers, which has heightened interest in geldanamycin and its analogues for potential anti-cancer treatment (Neckers, 2002) (Rascher *et al.*, 2003) (Shin *et al.*, 2008). The blue-pigmented actinorhodin is a further example of a PKS antibiotic produced by *Streptomyces coelicolor* A3(2) and first linked to a T2PKS in 1984 (Malpartida & Hopwood, 1984). PKSs are not exclusive to Actinomycetota, but were also described amongst fungal species such as *Aspergillus* secreting a host of polyketides (Sarma, Bhetaria, Devi, & Varma, 2017), such as the carcinogenic Aflatoxin B1 produced by *Aspergillus* sp. fungi (**Figure 1.7 b**) (Rushing & Selim, 2019) (Ehrlich & Cotty, 2002).
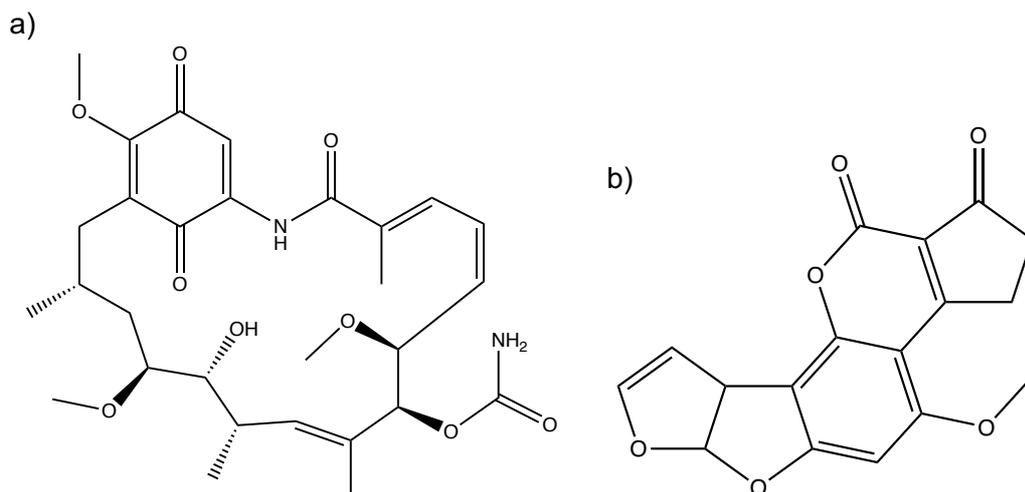
Figure 1.7 – Polyketide structure and biosynthesis. (a) and (b) show the structures of geldanamycin and aflatoxin B1 respectively.

1.4.2 Nonribosomal Peptide Synthetases (NRPS)

Nonribosomal Peptide Synthetases (NRPS) are a further example of an important BGC group, which encode nonribosomal peptides. NRPS are incredibly diverse, which can be attributed to both its ability to incorporate a wide range of monomers and extensive peptide modifications which occur during and post-chain assembly (Izoré *et al*., 2021). As well as the prevalence of NRPS in bacteria, they have also been reported in archaea and eukaryotes (Martínez-Núñez & López, 2016). Each single module harbours several catalytic domains, with each one responsible for the incorporation of a single residue. The activated amino acid is attached covalently to an integrated carrier protein domain and the substrate and intermediate molecules are transported to catalytic domains for peptide bond formation or modification. The biosynthesized peptide is transported to a terminal thioesterase domain which catalyses the release of the product (Miller & Gulick, 2016).

Actinomycin D (**Figure 1.8 a**) is an important example of an NRPS, as it was the first antibiotic isolated from Actinomycetota by Salman Waksman and H. Boyd Woodruff in 1940 (Selman A. Waksman & Woodruff, 1940) and the first antibiotic with anti-cancer activity (Hollstein,

1974). Actinomycin D was shown to inhibit RNA synthesis and induce apoptosis in cell lines such as PANC-1 (Kleeff, Kornmann, Sawhney, & Korc, 2000). Actinomycin D has also been shown to inhibit cell proliferation and induce apoptosis in MG63 human osteosarcoma cells, which cause malignant bone tumours (Lu *et al*., 2015). Vancomycin further illustrates the importance of NRPS antibiotics (**Figure 1.8 b**), and was first isolated from sediment collected from Borneo (Geraci, Heilman, Nichols, Wellman, & Ross, 1956). Initially overlooked because of its toxic effects, vancomycin is now considered to be a 'last resort' antibiotic with the rise of pseudomembranous enterocolitis and methicillin-resistant *Staphylococcus aureus* prompting a resurgence in its use (Levine, 2006). The vancomycin mechanism of action differs from most antibiotics as it binds to the cell envelope rather than a protein target, and thus it was considered unsusceptible to resistance. However, complex resistance mechanisms have now emerged and are found widely in pathogenic bacteria (Stogios & Savchenko, 2020) potentially creating serious downstream issues and adding further fuel to the global AMR crisis.
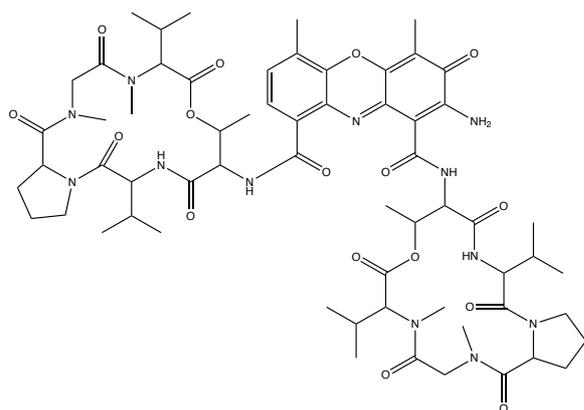
a)　　　　　　　　　　　　　　　　　　　b)
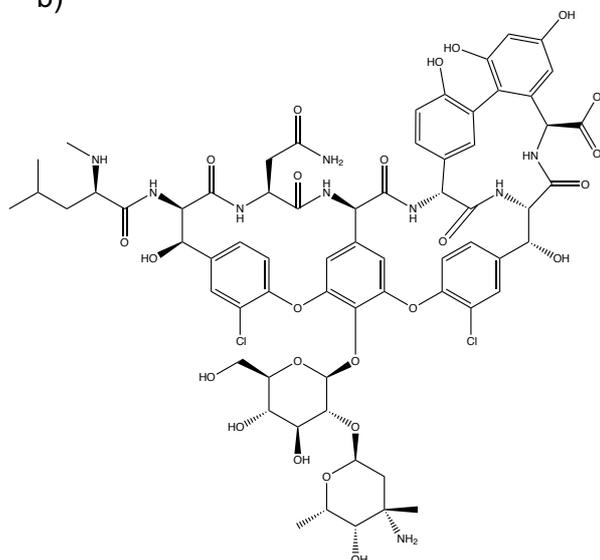
Figure 1.8 – NRP structures. (a) the structure of actinomycin D, the first antibiotic isolated from Actinomycetota in 1940 (Selman A. Waksman & Woodruff, 1940) and (b) vancomycin, commonly acknowledged as a last resort antibiotic.

1.4.3 Terpenes

Terpenes are a class of natural product commonly produced by fungi, plants and other eukaryotic organisms, fulfilling diverse roles such as attracting pollinators or repelling predators (Cimmino, Andolfi, & Evidente, 2014). Terpene biosynthesis follows a different logic to that of polyketide or non-ribosomal peptide synthesis {Helfrich, 2019 #504}. The main building blocks for terpene biosynthesis are made up of the subunits α, β, γ, δ, ε and ζ and are: Trans-isoprenoid diphosphates produced from C_5 isoprene precursors via αα and αδ head-to-tail trans-prenyl transferases; these diphosphates are then converted into the tri- or tetra-terpene sterol precursors by ε head-to-head prenyl transferases; the α, αβ and αβγ terpene synthases for plant terpene production; βγ di- and tri-terpene synthases, and finally the ζ head-to-tail cis-prenyl transferases which produce the cis-isoprenoid diphosphates for cell wall biosynthesis in bacteria {Oldfield, 2012 #503}.

A number of bacterial terpene synthase BGCs have been well characterised (Y. Yamada *et al*., 2015). Based on a training set of 140 previously identified bacterial terpene synthase BGCs, mining of 8.7 million bacterial proteins from public databases has revealed 262 putative terpene synthase BGCs. This illustrates that terpene synthases are widely distributed in bacteria. The group were also able to isolate and determine the complete structures of 13 novel terpenes through heterologous expression, using an engineered biosynthetically inert *S. avermitilis* as host (**Figure 1.9**) (Y. Yamada *et al*., 2015). It has been reported that fungi acquired terpene synthase genes, such as *Metarhizium* sp. and *Ophiocordyceps* sp., a common phylogenetic ancestor was shared with *Streptomyces* sp., *Pseudomonas* sp. and *Burkholderia* sp. (Jia *et al*., 2019), from bacteria via horizontal gene transfer.

Geosmin (**Figure 1.10 a**) is a terpene commonly produced by *Streptomyces* species. First isolated from a *Streptomyces griseus* fermentation broth, it is responsible for the earthy odour that emanates from soil (Gerber & Lechevalier, 1965). In *S. coelicolor* A3(2), the geosmin

metabolite is generated from farnesyl diphosphate by an enzyme encoded by the gene SCO6073 (Jiang, He, & Cane, 2006) with the N-terminal half of the protein catalysing Mg^{2+}-dependent cyclisation of farnesyl diphosphate to germacradienol and the C-terminal half catalysing the Mg^{2+}-dependent conversion of the germacradienol molecule to geosmin (Jiang, He, & Cane, 2007). Sesquiterpenes are a subclass of terpene that consists of three isoprene units which tend to contains C15 in their molecular structures (F. Yu & Utsumi, 2009) and counts albaflavenone (**Figure 1.10 b**) in its numbers, initially isolated from the highly odorous *Streptomyces albidoflavus* (Gürtler *et al*., 1994).
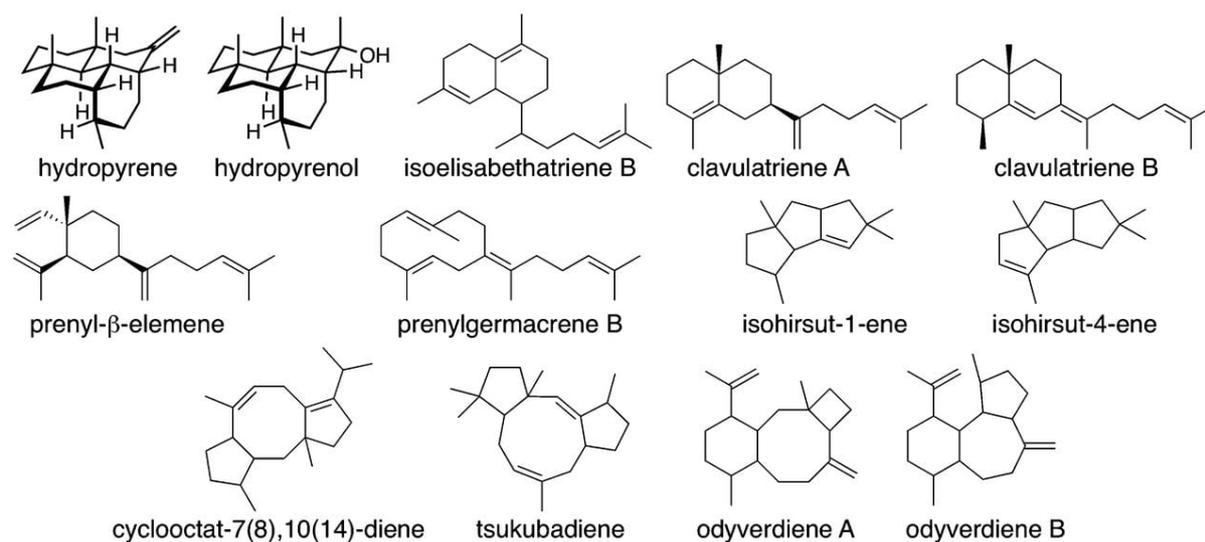
Figure 1.9 – Novel terpenes isolated through heterologous expression in *S. avermitilis*. The structures of 13 newly identified sesquiterpenes and diterpenes derived from *Streptomyces* sp. and heterologously expressed in *S. avermitilis* SUKA22 (Y. Yamada *et al*., 2015).
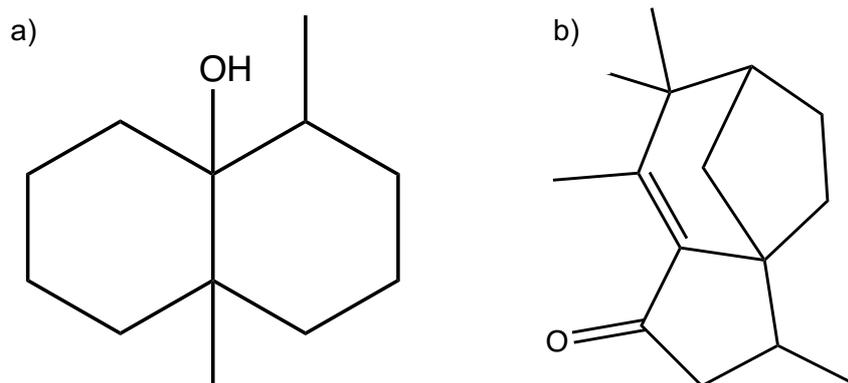
Figure 1.10 – Terpene structure. (a) Geosmin structure, the metabolite responsible for the earthy odour of soil and (b) albaflavenone, a sesquiterpene metabolite.

1.4.4 Ribosomally synthesised and Post-translationally modified Peptides (RiPPs)

Ribosomally synthesised and Post-translationally modified Peptides (RiPPs) are a diverse and structurally complex class of BGCs. RiPP biosynthesis involves ribosomal synthesis of a precursor peptide (consisting of a core peptide and leader peptide) which is modified by a series of RiPP-tailoring enzymes post-translation, with a biologically active final product the result of a final cleavage of the core peptide (Arnison *et al*., 2013). The peptide antibiotic thioviridamide **(Figure 1.11)** (Izawa, Kawasaki, & Hayakawa, 2013) is one such example of a RiPP, which was isolated from *Streptomyces olivoriridis* and its cryptic cytotoxic counterpart neothioviridamide (Kawahara *et al*., 2018)

Due to the aforementioned complex nature of RiPP biosynthesis with component genes not co-localised, it can often be difficult to characterise RiPP BGCs by way of commonly used bioinformatic genome mining tools such as antiSMASH (Kloosterman, Medema, & van Wezel, 2021). With this in mind, there are specialist RiPP genome mining tools, such as RiPP

Precursor Peptide Enhanced Recognition (RIPPER) which identifies RiPP precursor peptides in close proximity to YcaO-domain proteins (Santos-Aberturas *et al*., 2019). RIPPER has been successfully utilised to identify novel metabolites, such as the amidine-containing streptamidine from *S. albidoflavus* J1074 (Russell & Truman, 2020). The discovery of this novel antibiotic from *S. albidoflavus* J1074, a relatively well studied strain, highlights the unexplored chemical space in which many RiPPs occupy.
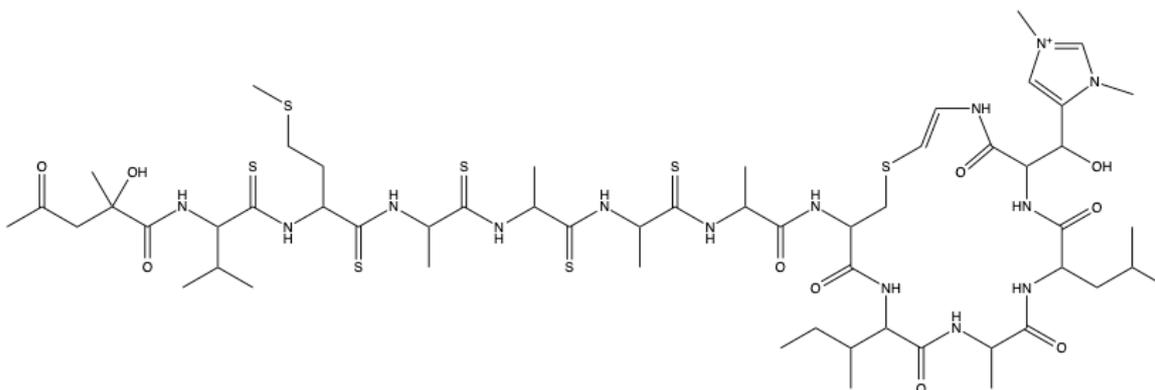
Figure 1.11 – Thioviridamide structure. The chemical structure of the RiPP thioviridamide isolated from *S. olivoriridis* (Izawa *et al*., 2013).

1.4.5 Other classes of specialised metabolite produced by *Streptomyces*

Specialised metabolites biosynthesised by *Streptomyces* are of course not limited to the four natural product chemical classes described (Alam *et al*., 2022). Thiopeptides, such as geninthiocin A isolated from *Streptomyces* sp. RSF18, have shown potent antibiotic activity against Gram positive bacteria (S. Li *et al*., 2019). Ectoine does not fall into any of the outlined classes but is widely conserved among *Streptomyces* sp. as an osmolyte commonly expressed under salt stress conditions (Galinski, Pfeiffer, & Trüper, 1985) (Bursy *et al*., 2008).

One of the most important classes beyond those already outlined are the siderophores, iron-chelators produced by microorganisms such as *Streptomyces* mostly growing under iron-

deficient conditions. The expressed and excreted siderophore scavenges environmental iron, forming a siderophore-iron complex which is taken up by the producer cell using a membrane-associated ATP-dependant transport mechanism with high substrate specificity (Köster, 2001). Iron acquisition in these conditions results in a gain for the chelating bacterium and ultimately deprives neighbouring competitors of an important nutrient (Terra, Ratcliffe, Castro, Vicente, & Dyson, 2021). Desferrioxamine E is one such siderophore which is highly conserved within *Streptomyces* sp. and has been identified as the major desferrioxamine siderophore produced by the model organism *S. coelicolor* M145 (Barona-Gómez, Wong, Giannakopulos, Derrick, & Challis, 2004). Siderophores such as desferrioxamine E may also contribute to interactions between bacterial species. When produced by *Streptomyces griseus,* desferrioxamine E was shown to stimulate growth and differentiation in *Streptomyces tanashiensis* as well as stimulating specialised metabolite production and morphological alterations in other Actinomycetota, activity which was abolished when the desferrioxamine E BGC was disrupted in *S. coelicolor* A3(2) (Yamanaka *et al*., 2005). These results highlight the importance of siderophores when it comes to many aspects of complex *Streptomyces* sp. physiology.
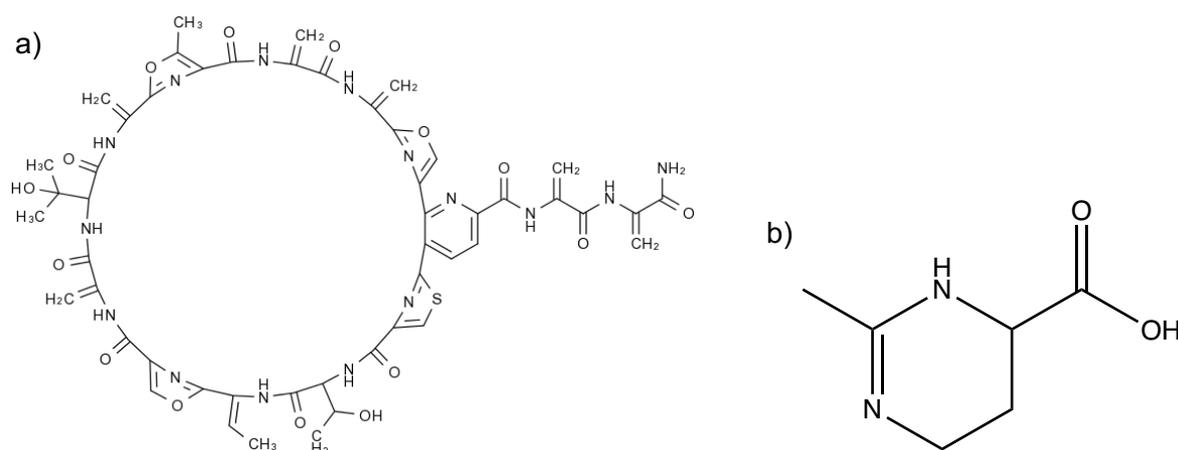
Figure 1.12 – Structures of geninthiocin A and ectoine. The chemical structures of a) the thiopeptide geninthiocin A {Li, 2019 #437} and b) the highly conserved osmolyte ectoine {Galinski, 1985 #438}.

1.5 Actinomycetota ecology

1.5.1 Actinomycetota in the environment

Streptomyces are extremely well adapted to the soil habitat, where they exist ubiquitously and in concert with cohabiting organisms (Tarkka & Hampp, 2008) and represent between 1 and 20% of the total viable cell count (R. Kumar & Jadeja, 2016). In this environment, *Streptomyces* sp. develop substrate mycelia consisting of numerous hyphae which grow via tip extension and explore surroundings for nutrients by branching through the soil (Seipke, Kaltenpoth, & Hutchings, 2012). The role of specialised metabolites within the natural environment is thought to include chemical weaponry as one of potentially myriad functions, yet is poorly understood (J. Davies & Davies, 2010).

Indeed in nature, Actinomycetota such as *Streptomyces* sp. are well known to form symbiotic relationships with insects and plants. One of the most well-characterised are leaf-cutter ants (*Acromyrmex*) (Kaltenpoth, 2009), which cultivate a fungus garden (*Leucoagaricus gonglyophorus*) as a food source. The ants house antimicrobial-producing Actinomycetota on their cuticles to protect the gardens from the invasive fungus *Escovopsis* sp. (Currie, Scott, Summerbell, & Malloch, 1999). Furthermore, *Streptomyces* isolated from leaf-cutter ant cuticles were shown to produce of the antibiotic valinomycin, which was distributed directly on the integuments of *Acromyrmex* workers, strongly suggesting that this metabolite was actively utilised by the ants as protection from bacterial pathogens (Schoenian *et al*., 2011). Similarly, Haeder *et al*. observed that *Streptomyces* sp. Ao10, again isolated from leaf-cutter ants, proved to be active against the *Escovopsis* fungal pathogen. The group were able to identify candidacin macrolides and showed that they killed *Escovopsis* but did not inhibit the fungal symbiont (Haeder, Wirth, Herz, & Spiteller, 2009). These results show how Actinomycetota have co-evolved with environmental neighbours, forming integral symbioses which influence many aspects of the leaf-cutter ant ecosystem.
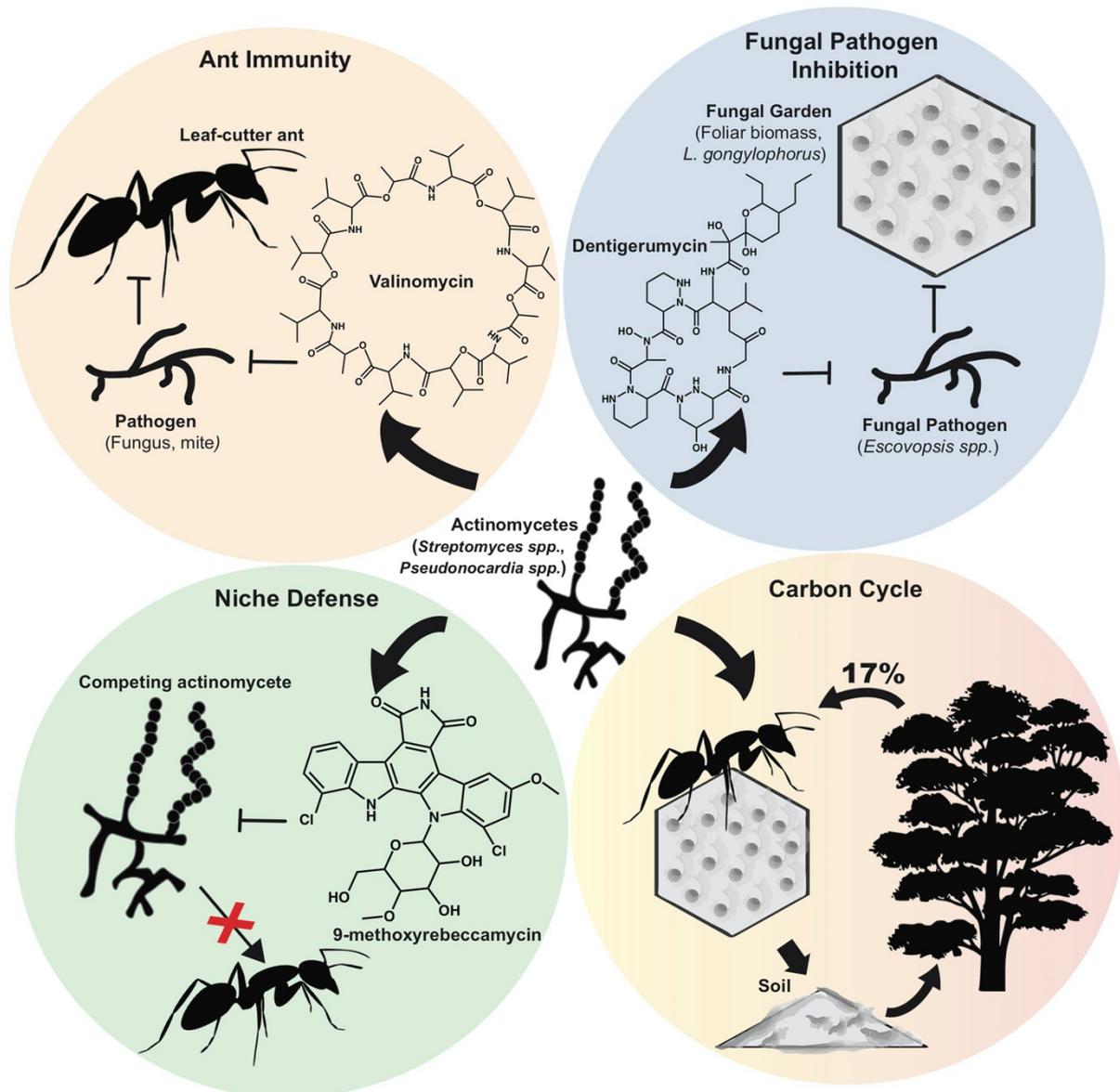
Figure 1.12 – Diverse roles of Actinomycetota natural products in the leaf-cutter ant ecosystem. Actinomycetota such as *Streptomyces* sp. and *Pseudonocardia* sp. are involved in many aspects of the leaf-cutter ant ecosystem, such as producing antibiotics to confer immunity, antifungals to protect food sources, offering niche defence and by preserving the ant community, the symbiosis in turn helps the larger global ecosystem (Behie, Bonet, Zacharia, McClung, & Traxler, 2017).

1.5.2 Subinhibitory concentrations of antibiotics in nature

At subinhibitory concentrations, antibiotics have been found to elicit transcription activation, suggesting roles beyond inhibition within the natural environment (Julian Davies, Spiegelman, & Yim, 2006). The behaviour of *Streptomyces* sp. either producing or in response to subinhibitory concentrations of antibiotics is the focus of ongoing study. In response to a subinhibitory concentration of 8-*O*-methyltetrangomycin, produced by *Streptomyces* SBRK2, the ability of *Staphylococcus aureus* ATCC 25923 to form biofilms was reduced and the cell surface hydrophobicity index was increased (Jabila Mary, Kannan, Iniyan, Ramachandran, & Prakash Vincent, 2021). At the opposite end of this scale, the *Streptomyces*-derived RiPP antibiotic thiostrepton was found to stimulate biofilm formation in *Pseudomonas aeruginosa* PAO1 and PA14, which possessed plasmids conferring resistance via *tsr* resistance genes (Ranieri *et al.*, 2019). It is clear that further investigations into the behaviour of Actinomycetota *in situ* are necessary to elucidate the roles of antibiotics at subinhibitory concentrations.

1.5.3 Soil microcosm systems

Despite *Streptomyces* prevalence in soil, investigations into the genus' behaviour *in situ* are limited. It has been shown that a soil microcosm system is a viable alternative method of cell culture. In one study, *S. lividans* and *S. violaceolatus* were inoculated to both sterile and non-sterile soil microcosms. Transconjugants of both strains containing the self-transmissible, high-copy number plasmid pIJ673 were recovered indicating that the strains not only survived, but were metabolically active (Wellington, Cresswell, & Saunders, 1990). The amendment of soil using various carbon sources and/or chemical elicitors within a microcosm has been used to analyse the inhibitory profiles of *Streptomyces* strains. *Streptomyces* were isolated from both natural and carbon-amended soil and were cultured together in competition assays. Isolates from natural soil primarily inhibited isolates from carbon-amended soil and in contrast, isolates from carbon-amended soils primarily inhibited isolates from natural soils (Dundore-

Arias, Felice, Dill-Macky, & Kinkel, 2019). The same team conducted a study evaluating the relationship between carbon amendments within soil microcosms and *Streptomyces* community characteristics. It was found that an increase in carbon amendments resulted in significantly reduced soil pH which in turn was gave a decrease in total *Streptomyces* densities in soil with both high and low levels of organic matter (Dundore-Arias, Castle, Felice, Dill-Macky, & Kinkel, 2019). A further study of amended soil was conducted to measure the effects of glucose and lignin amendments on the inhibitory behaviour of *Streptomyces* isolated from prairie soil. It was found that isolates from soil microcosm communities amended with high levels of glucose and lignin (250 g C/m^2) were altogether more inhibitory to other *Streptomyces* strains than isolates from communities amended with lower levels of both metabolites (100 g C/m^2) (Schlatter *et al*., 2009). The potential use of soil microcosm systems in laboratory culture is clear and will be all the more fruitful when combined with multi-omics methods to interpret the resultant data.

1.6 The dawn of microbial genomics

1.6.1 Introduction to genomics and its rise in microbiology

The founding principle of '-omics' approaches is that complex systems, such as the genome, metabolome, and transcriptome, will be more thoroughly understood when individual datasets are treated as part of a whole (R. Yamada, Okada, Wang, Basak, & Koyama, 2021). Genomics was the first of these disciplines to become prevalent, and differs from genetics which tends to encompass individual variants or single genes (Hasin, Seldin, & Lusis, 2017). Genomics can be defined as being the study of the structure, function, evolution, and mapping of the full repertoire of genes encoded for by the genome (Vailati-Riboni, Palombo, & Loor, 2017). The term 'genomics' in relation to the field it is currently used to describe was suggested to McKusick and colleagues by T.H. Roderick of the Jackson Laboratory in Maine as a name for

their new journal detailing this developing discipline of mapping and sequencing. The word itself is a hybrid of the word's 'gene' and 'chromosome', with the parentage of both being Greek (McKusick & Ruddle, 1987). The seed of the concept was planted early in the 20th century, when Wilhelm Johannsen coined the term 'gene' to describe the physical unit that ties with the genetic determinant of any inheritable trait in a given organism whilst at the same time distinguishing the 'genotype' – the hereditary disposition of an organism, from the 'phenotype' – an organism's physical characteristics that manifest as a consequence of said dispositions (Peirson, 2012). From here, in 1920, Hans Winkler suggested that the complete genetic make-up of any given organism should be referred to as the 'genome'. However, it took decades to determine both that DNA was the physical hereditary material and a further decade to elucidate the double-helix structure in three dimensions, a point developed by Watson and Crick based on facts regarding the structure determined by Rosalind Franklin, is regarded as the event that accelerated the molecular biology era after its genesis in the 1930s (Weissenbach, 2016).

1.6.2 Genome sequencing

RNA sequencing became possible as technology developed in the 1960's, initially focusing on ribosomal and transfer RNAs with the first results from sequencing protein coding genes coming from the protein coat of the MS2 bacteriophage. Results fell into line with expectation – the genes were flanked by start and stop codons and the amino acids matched the established code (Min Jou, Haegeman, Ysebaert, & Fiers, 1972). In 1977, Maxam and Gilbert published a paper detailing a new method for sequencing DNA, to be known as Maxam-Gilbert sequencing. In this method, the denatured, single-stranded DNA is radiolabelled at the 5' end with ^{32}P. The DNA is then cleaved at specific points before being run on an agarose gel, from which the fragments can be visualised through the plutonium tag {Maxam, 1977 #509}. Also in 1977, Fred Sanger and colleagues published a paper detailing the first instance of

sequencing a viral DNA genome, from the bacteriophage ΦX174 (Sanger, Air, *et al*., 1977). Later that year Sanger *et al.* published work detailing the comparison of a new method where the sequencing of bacteriophage ΦX174 was compared to the outcome of the sequencing using the previously described plus and minus method. The technique would come to be known as 'Sanger Sequencing', and for around four decades was the most widely used DNA sequencing technique (Sanger, Nicklen, & Coulson, 1977). In 1993, the full nucleotide sequence of a particularly virulent strain of variola virus (the causative agent of smallpox), designated as Bangladesh-1975, was the first example of a DNA genome being sequenced and assembled via a fully automated process (Massung *et al*., 1993). This work also led into the use of whole genome sequencing, where a shotgun sequencing strategy was employed to rapidly reassemble genomes accurately whilst also keeping costs at a minimum. This whole genome shotgun sequencing eliminated the need for initial mapping, as genome maps may have been unavailable for certain strains (Fleischmann *et al*., 1995). One major goal of these whole genome sequencing projects was to create a library of publicly available genomes, as well as providing a completely novel representation of the physical nature of the genomes of both prokaryotic and eukaryotic organisms.

1.6.3 Automated genome mining platforms

The advancement of sequencing technology and readily available bioinformatic pipelines has facilitated the mining of BGCs (Palazzotto & Weber, 2018). The vast majority of microbial diversity is thought to be unculturable – this majority is often referred to as microbial dark matter - so an enormous amount of this previously unreachable specialised metabolism data can now be elucidated using these modern tools (R. Chen, Wong, & Burns, 2019). Laboratory-based research cannot progress at the same rate as genomic discovery, the main reason being that characterisation of gene clusters experimentally is a heavily labour-intensive process. The manual annotation of gene clusters is also a laborious task, and as such results

can suffer and annotations remain incomplete. This renders the successful mining and annotation of genomes via effective *in silico* methods of identification essential (Medema *et al.*, 2011). Published in 2008, ClustScan was the first of several *in silico* platforms to facilitate automated analysis of specialized metabolism in bacterial genomes that have been published, detecting and annotating PKS and NRPS gene clusters (Starcevic *et al.*, 2008). Other platforms such as NP. searcher (M. H. Li, Ung, Zajkowski, Garneau-Tsodikova, & Sherman, 2009) and SBSPKS (Anand *et al.*, 2010) followed but the tools were generally limited to analysis of core genes involved in the synthesis of type I polyketides and non-ribosomal peptides.

The gold standard for annotation and analysis of BGCs is antiSMASH (Medema *et al.*, 2011). The software can detect all known classes of specialised metabolite BGCs as well as recognising similarities in the evolution of the query cluster and other detected gene clusters, resulting in the prediction of gene functionality (Medema *et al.*, 2011). antiSMASH features the integration of independently developed cluster identification databases for comprehensive analysis and annotation – v4 included the ClusterFinder method for the prediction of gene cluster boundaries (Kai Blin *et al.*, 2017) and v6 (K. Blin *et al.*, 2021), the most recent iteration, facilitated the scanning of protein family definitions with the integration of the TIGRFAMs database (Haft *et al.*, 2013).

A prediction of BGC identity is facilitated by comparison to experimentally verified clusters found within the Minimum Information about a Biosynthetic Gene Cluster (MIBiG) database – an open-source library of BGCs of known function (Kautsar *et al.*, 2020). The antiSMASH platform can be used in conjunction with other tools, such as MetaBAT – an automated open-source platform which is used for the accurate reconstruction of single whole genomes from large, complex microbial communities or synthetic metagenome datasets (Kang, Froula, Egan, & Wang, 2015). This was demonstrated to great effect by Cuadrat *et al*, who created draft genomes from aquatic environmental samples obtained from Lake Stechlin in north-east

Germany using MetaBAT and screened the recovered genomes for specialised metabolite BGCs using antiSMASH. From the 121 genomes recovered from the samples, the team were able to identify 243 known BGCs including 19 PKS's, 18 NRPS's and 3 PKS/NRPS hybrids (Cuadrat, Ionescu, Davila, & Grossart, 2018).

The genome mining process has been the subject of unprecedented scale-up, with it now possible to perform analysis on strain collections and entire microbiomes, using data mined from whole genera. A recent comparative genomics study of 121 *Streptomyces* genome sequences identified a total of 5,289 BGCs at an average of 43.7 per genome (C. Caicedo-Montoya, M. Manzo-Ruiz, & R. Ríos-Estepa, 2021). Draft or complete genome sequences of a further 292 *Streptomyces* sp. produced a monophyletic clade with exceptionally large genomes, ranging between 10.7 and 12.7 Mb. In this clade the lowest number of BGCs per genome was 45, and the highest 55. A genetic network of BGC diversity revealed 89 unique BGCs across 11 genome sequences, showing an average of 16% of each BGC suite to be exclusive to the host genome (Chung *et al*., 2021). With this increase in the size and complexity of these generated datasets, the need for the development of a bioinformatic framework capable of performing analysis to this level became clear. This came led to the development of the congruent platforms BiGSCAPE (Biosynthetic gene similarity clustering and prospecting engine) and CORASON (Core analysis of syntenic orthologues to prioritise natural product gene clusters) (Navarro-Munoz *et al*., 2020). BiGSCAPE provides rapid and highly interactive sequence similarity network analysis of specialised metabolite BGCs within genomes that have been predicted using tools such as antiSMASH, and groups similar BGCs together into Gene Cluster Families (GCFs). These families consist of clusters from the genomes of many organisms. The GCFs can then be connected via feature-based linking to molecular families within mass spectrometry data. The CORASON platform then phylogenetically evaluates the relationships within these GCFs with the output portrayed as a phylogenetic tree.

Attempts have been made by teams to map this somewhat daunting level of biosynthetic diversity. Methods used previous to the advent of BiGSCAPE/CORASON faced many obstacles, such as failing to recognise evolutionary relationships within GCFs and an inability to measure similarity between complete and fragmented gene clusters (Cimermancic *et al*., 2014). Previous attempts also fell short with the requirement of large-scale computing facilities paired with very long processing times. The streamlined BiGSCAPE/CORASON platform utilizes both the gene cluster prediction capabilities of antiSMASH, and the biosynthetic data comparison provided by MIBiG, working rapidly in comparison to previous models and processing hundreds of genomes in a matter of minutes without the need for any computing hardware more than a conventional laptop.

In a study presenting the first comparative genomic study of *Streptomyces* involving a large amount of available complete, high-quality genomes, BiGSCAPE/CORASON was utilised to great effect. The platform helped to reveal the *Streptomyces* sp. pan-genome in terms of protein sequence similarity and phylogenetic relationships within the analysed genomes. Using BiGSCAPE/CORASON to group BGCs which produce similar metabolites into GCFs from 121 complete, high-quality genomes, 2,359 nodes and 12,969 edges were shown, underlining the high level of variability within *Streptomyces* sp. BGC suites (Carlos Caicedo-Montoya, Monserrat Manzo-Ruiz, & Rigoberto Ríos-Estepa, 2021).

1.7 Metabolomics

1.7.1 Early mass spectrometry

In 1910, the physicist Sir Joseph John Thomson constructed an instrument for the measurement of mass-to-charge (m/z) ratios of ionised atoms which would yield a mass spectrum, from which he identified H_2, N_2, O_2 and CO_2 (Thomson, 1913). Earlier in his career, Thomson won the Nobel Prize in Physics for his discovery of the electron. Really quite the

contribution. Thomson's assistant, Francis W. Aston, made several improvements on the technology and he himself won a Nobel Prize in Chemistry in 1922 for identifying isotopes using a mass spectrograph. During the early mass spectrometry experiments, the most notable mass spectra identified were that of neon (Ne). The spectrum clearly showed isotopes with the masses 20 and 22 in a ratio of 10:1, which aligns with Ne's atomic mass of 20.2 and revealed the ability of Mass Spectrometry (MS) to determine exact atomic weights and quantitatively analyse individual isotopes (Aston, 1920). This ability was applied to devastating effect in World War II, when the U-235 used in uranium bombs were separated using MS with an instrument known as a calutron (Yergey & Yergey, 1997).

Until around 1960, the dominant use of molecular MS technologies was for quantitative analysis of low-boiling hydrocarbons, however, structure determination at this time was unreliable which led to poor acceptance of the technology (McLafferty, 2011). The development of Gas Chromatography Mass Spectrometry (GC-MS) in the 1950s expanded the applications of qualitative MS, and the initial GC experiments performed by Anthony T. James and Archer J. P. Martin won them a Nobel Prize in Chemistry. In GC, a mixture is vaporised and eluted by an inert gas through a column filled with small particles. The components of the mixture interact with these particles at different strengths causing them to elute at different speeds through the column towards a detector, resulting in an exit signal (James & Martin, 1952). Fred McLafferty and Roland Gohlke were responsible for the first coupling of GC with MS in 1956, producing spectra of benzene, acetone, toluene and carbon tetrachloride from a mixture of these four components (Gohlke, 1959).

1.7.2 Recent advances in microbial mass spectrometry

For many years, two technologies have presented themselves as the best options for performing microbial mass spectrometry analyses. The aforementioned GC-MS for the analysis of volatile metabolites (Kannaste, Copolovici, & Niinemets, 2014) and Liquid

Chromatography Mass Spectrometry (LC-MS) which has become the gold standard platform for high-throughput metabolomics analysis (Pitt, 2009) (Perez *et al.*, 2016). LC-MS collects signals from the microbially-sourced metabolites indicated by their mass ion peaks along with their fragments and isotopes. Generally, modern MS techniques use Electrospray Ionisation (ESI) as the ionization method of choice (Ho *et al.*, 2003). ESI is preferred because it is a 'soft' ionisation method that causes little fragmentation, allowing simple downstream spectrum interpretation and metabolite identification, whereas 'hard' fragmentation methods tend to yield a larger number of fragments of lower mass {Wang, 2018 #505}. Ionization can be performed in positive, negative or a switch mode which measures both positive and negative ions in one single analysis. Positive ionisation mode results in the peaks of a certain *m/z* in a spectrum representing the protonated form of the metabolite (e.g. $[M+H]^+$). Choosing the optimal ionization mode is important, as one study showed that negative mode allowed for higher sensitivity for 46% of 33 comparable compounds {Liigand, 2017 #506}. Although the combination of LC-MS and ESI has proved fruitful, this single quadrupole detection method is often overlooked in favour of more comprehensive methods.

In the last 15 years, triple quadrupole – referred to as 'tandem' - MS has superseded LC with a single quadrupole MS detector (Grebe & Singh, 2011). LC-MS is sensitive but may lack selectivity, with LC-MS/MS providing experimental validation (Tribalat, Paisse, Dessalces, & Grenier-Loustalot, 2006). LC-MS/MS has been used to characterise natural products from Actinomycetota. PKS sequence data of marine sponge-derived *Salinispora* indicated that the strains may produce rifamycin-like metabolites. The presence of rifamycin B and rifamycin SV was confirmed with LC-MS/MS analysis, confirming *Salinispora* as a potential new source of rifamycins (T. K. Kim, Hewavitharana, Shaw, & Fuerst, 2006). A further study using LC-MS/MS to analyse marine-derived *Salinispora arenicola* established temporal limitations of specialised metabolite production, indicating that a 14- and 29-day incubation period is optimal for maximum production of rifamycin A and S/W respectively (Bose *et al.*, 2015). A novel 4H-chromen-4-one derivative, isolated from marine *Streptomyces ovatisporus* S4702[T], with highly

potent bioactivity against *B. subtilis* and *M. luteus* was chemically characterised using LC-MS/MS (Kurt-Kızıldoğan *et al.*, 2022). Without a doubt, developments in the field of MS are ongoing rapidly.

MS technologies have been developed to directly identify and characterise microbes. Matrix-Assisted Laser Desorption Ionisation-Time Of Flight (MALDI-TOF) mass spectrometry was first used in the mid-1990s to identify single colonies of bacteria directly from culture (Claydon, Davey, Edwards-Jones, & Gordon, 1996). In 2004, a MALDI-TOF database of mass spectra was launched for rapid screening of pathogenic bacteria (Keys *et al.*, 2004). For MALDI-TOF, the sample is mixed with an organic matrix and the two co-crystallise. The sample is then ionised with the protonated ions accelerated and separated by *m/z* and detected by TOF analysers (Yates III, 1998). A Peptide Mass Fingerprint (PMF) is created based on the TOF information and compared to the PMFs of known organisms within the database allowing rapid identification (Singhal, Kumar, Kanaujia, & Virdi, 2015). The proficiency of MALDI-TOF has led to its application in medical diagnostics, food quality control and biodefence {Croxatto, 2012 #507}. As such, MALDI-TOF MS has been proposed as a cost-effective tool for the rapid identification of *Streptomyces.* One study, using a collection of 48 isolates, performed a blind test and identified 100% of the tested isolates in under 30 minutes and obtained reproducible spectra from each isolate (Loucif, Bendjama, Gacemi-Kirane, & Rolain, 2014). Such methods of identifying previously discovered isolates may prove invaluable to overcoming bottlenecks in antimicrobial discovery.

1.7.3 Comparative metabolomics and dereplication

Metabolomics is defined as the comprehensive analysis of the metabolic products at the end of a physiological process – the metabolome – of a given biological specimen or sample, with the aim of completely mapping the biochemical reactions in a biological system (Alarcon-Barrera, Kostidis, Ondo-Mendez, & Giera, 2022). It allows the exploration of interactions

between upstream environmental input and downstream genomic output (Clish, 2015). Whilst genomics can infer the possibility of an occurrence, metabolomics offers a real-time view of events that may be taking place at any point (Wishart, 2016). Metabolomics methods have been continuously developed to provide the optimum means of measuring specialised metabolite production in the extracted metabolome (Riekeberg & Powers, 2017). However, it should be noted that successful metabolite extraction is essential in creating an accurate metabolite profile (Ser, Liu, Tang, & Locasale, 2015) and any sub-optimal conditions, such as temperature or extraction solvent composition, can have severely limiting downstream effects leading to incomplete metabolite profiles. Therefore, optimisation of extraction method is key (Dettmer *et al*., 2011) (Ivanisevic *et al*., 2013).

A consistent bottleneck in the discovery of novel microbial specialised metabolites is the rediscovery of known metabolites, which fed into the relative collapse of the Waksman discovery pipeline by the 1960s (Jones *et al*., 2017). It is therefore extremely important to develop rapid, reliable methods for distinguishing between known and unknown metabolites. This process is known as 'dereplication' (Reynolds, 2017). The three basic pillars of dereplication are the biological taxonomy of specialised metabolite-producing organisms, the knowledge of specialised metabolite molecular structures, and the availability of specialised metabolite spectroscopic signatures (Lianza *et al*., 2021). The process of dereplication - the identification and removal of previously discovered natural products from datasets - is necessary to combat rediscovery of known specialised metabolites and comparative metabolomics can be utilised to this end (Covington, McLean, & Bachmann, 2017). Many natural product databases exist, including the likes of SuperNatural (Dunkel, Fullbeck, Neumann, & Preissner, 2006) and AntiBase, but there was no single platform that brought standard libraries together until the creation of Global Natural Products Social (GNPS) (M. Wang *et al*., 2016). GNPS is the largest public natural product database of MS/MS spectra and exists as an open-access knowledge database of raw or processed LC-MS/MS data. GNPS's reliance on MS/MS data can also be viewed as a flaw, as it lacks the ability to analyse

valuable single quadrupole MS data and unfortunately, there is no platform which integrates all of these libraries. Ultimately, the success databases such as these relies on the community sharing of data as publicly available natural product databases. As of 2020, it is estimated that the GNPS libraries contain only 2.5% of known natural products (van der Hooft *et al*., 2020), so clearly there is still much work to be done. However, GNPS is set apart from other repositories with the ability to rapidly construct molecular networks for rapid, automated dereplication.

Molecular networks are visual displays of the chemical space present in MS facilitating the detection of groups of spectra from related metabolites. Networks appear as clusters representing molecular families of metabolites based on structural similarity provided by their MS/MS fragmentation pattern and retention time (Yang *et al*., 2013). The absence of the identity of a metabolite within the network can then be used to chemically prioritise these metabolites (M. Wang *et al*., 2016). MS/MS data-guided molecular networking has been utilized for the precise identification of bioactive specialised metabolites from *Streptomyces*. From the Unkeshwar hot springs in India, 86 Actinomycetota strains were isolated with *Streptomyces* sp. GH176 displaying potent antimicrobial activity. Upon dereplication via molecular networking, the strain was found to produce both abyssomycin I and terpentecin (Mehetre *et al*., 2019). Carbon concentration in the environment has been shown to upregulate metabolic pathways for biosynthesis in the strain *Streptomyces* sp. MBT27. The specialised metabolite profiles of the strain underwent extensive fluctuation underpinned by the choice of carbon source used. The use of GNPS molecular networking led to the identification of two novel quinazolinone metabolites, as well as a family of previously known quinazolinone metabolites, which were increased in response to the presence of an elevated concentration of glycerol (Machushynets, Wu, Elsayed, Hankemeier, & van Wezel, 2019). Although extremely valuable for the discovery of specialised metabolites, genomic and metabolomic datasets existing in isolation only paints half a picture. These datasets must therefore be linked.

1.8 Metabologenomics

1.8.1 Manual linking approaches for large-scale natural products -omics datasets

The development of interdisciplinary -omics approaches would accelerate the specialised metabolite discovery pipeline (Subramanian, Verma, Kumar, Jere, & Anamika, 2020). Development of various models for integrated analysis of 'upstream' genomic data and 'downstream' metabolomic data is an active area of research in drug discovery and several approaches have been made in an attempt to bridge the gap between these large datasets (Graham *et al*., 2018). Broadly, these methods are defined as manual or automated (Das *et al*., 2018). One of the first manual approaches used a combination of metabolomic analyses paired with mining whole genome sequencing data to investigate the biosynthetic potential of thirteen strains of *Pseudoalteromonas luteoviolacea* that proved to be phylogenetically close despite being geographically diverse. LC-MS/MS analysis was followed by molecular networking, with *in silico* genome mining via antiSMASH. The group analysed presence and absence patterns across the pan-genome as well as molecular features, which lead to the discovery of the indolmycin BGC (Maansson *et al*., 2016). Marine myxobacteria have proven difficult to culture (Felder *et al*., 2013) but the five species that have been isolated and sequenced display promising levels of novel chemistry and bioactivity (Davila-Cespedes, Hufendiek, Crusemann, Schaberle, & Konig, 2016). Combining genome mining via antiSMASH and GNPS molecular networking, polyhyroxybutyric acid (PHB) was detected in all five sequenced marine myxobacteria strains by observing corresponding mass shifts and its appearance in a large cluster within the molecular network, with this data used to link indolmycin to its production gene cluster (Amiri Moghaddam *et al*., 2018).

Curacomycin and its non-bioactive analogue dechlorocuracomycin were discovered by way of manual linking. A study was launched in strains of *Streptomyces curacoi* and *Streptomyces noursei* to investigate whether the expressed tryptophan halogenase could function to provide

chlorinated tryptophan to NRPS genes close by in sequence. It was hypothesised that NRPS and tryptophan halogenase co-functioned to produce novel halogenated peptides. *S. curacoi* and *S. noursei* are known antibiotic producers (curamycin and nystatin, respectively) and were shown to possess a similar suite of NRPS genes. Genome mining and metabolomic methods such as HPLC, MS and NMR were used to identify, isolate and elucidate the chemical structures of both novel peptides. The study also indicated that focusing on modification enzymes such as tryptophan halogenase may be a useful tool for drug discovery (Kaweewan, Hemmi, Komaki, & Kodani, 2020).

A combination of pattern-based genome mining and comparative metabolomics has been utilized as a method of multi-targeted manual linking to isolate novel bioactive specialised metabolites in Chinese-lichen-associated *Streptomyces* sp. YIM 130001. It was found during bioactivity analysis that the strain displayed antimicrobial activity against *B. subtilis*. antiSMASH predicted Cluster 1 to be a putative thiopeptide, renowned for their antimicrobial effects against Gram positive bacteria. It was assumed that this BGC was the cause of the antimicrobial activity against *B. subtilis.* To confirm this, a vector was inserted into the coding region of the *genB* gene. The knockout strain failed to inhibit *B. subtilis,* and comparative metabolomic analysis of the knockout strain extract showed that two peaks disappeared, suggesting a correlation between these peaks and both the putative thiopeptide and antimicrobial activity. The bioactive metabolite was found to be close in structure to geninthiocin following purification and structural elucidation via NMR thereby elucidating the congener geninthiocin B (Schneider *et al*., 2018).

1.8.2 Feature-based automated linking approaches

Automated linking methods are grouped into two categories: feature-based and correlation-based. In feature-based linking, chemical features are predicted from genomic data (Soldatou, Eldjarn, Huerta-Uribe, Rogers, & Duncan, 2019). There are many tools available to facilitate

the prediction of the structural properties of a molecule based on genomic information to directly detect corresponding features in MS data. Kersten and colleagues coined the term peptidogenomics to refer to this practice and used these techniques to characterise AmfS and stendomycin I and their corresponding BGCs from *S. griseus* IFO 13350 and *Streptomyces hygroscopicus* ATCC 53653, respectively (Kersten *et al.*, 2011). Dereplication tools can also be used to this effect. Mohimani and colleagues developed the DEREPLICATOR tool and ran said tool on LC-MS/MS datasets downloaded from the GNPS platform to identify peptidic natural products in a high-throughput manner, resulting in the identification of surugamide A (Mohimani *et al.*, 2017). Later work from the same group, using an updated version of the tool called DEREPLICATOR+ combined with antiSMASH-based genome mining, identified a chalcomycin-like polyketide antibiotic from *Streptomyces* Mg1 (Mohimani *et al.*, 2018).

1.8.3 Correlation-based automated linking approaches

Correlation-based linking approaches rely on establishing patterns of source strain occurrence where data corresponding to the genomic and metabolomic profiles for a large number of strains are available. In the current most widely used correlation-based linking pipeline, BGCs are grouped into Gene Cluster Families (GCFs), with GCFs then correlated with spectra corresponding to Molecular Families (MFs) based on the occurrence of source strains throughout the dataset (Doroghazi *et al.*, 2014). This method assumes that true links will show a high correlation score (Soldatou *et al.*, 2019).

Once BGCs are clustered into GCFs, the shared source strains can be used as a jumping-off point for the correlation of BGCs and produced metabolites. The large-scale combination of gene clusters and metabolites in this way was used to identify and characterise tambromycin, a novel antimicrobial natural product, by establishing a correlation score from data accrued from 178 different strains of Actinomycetota (Goering *et al.*, 2016). During further study of the same 178 strains of Actinomycetota, the same method of correlation was used to identify the

biosynthetic origin of the rimosamides, a new family of specialised metabolite produced by *S. rimosus* NRRL B-2659. The BGC for a closely related family of natural products, the detoxins, was also identified (McClure *et al*., 2016). Correlation-based metabologenomics has also been combined with GNPS molecular networking. By comparing data garnered from 269 strains, a novel class of six ribosomal peptide natural products known as the tyrobetaines as well as their BGC were identified. The presence of the BGC was then confirmed in *S. rimosus* NRRL B-2659 via heterologous expression in *S. lividans* 66 (Parkinson *et al*., 2018). The aforementioned most common scoring method for computationally linking GCFs and MFs in metabologenomics analysis is outlined below, starting from zero and considering that it is more likely that a strain harbours the BGC than produces the metabolite (Doroghazi *et al*., 2014): 10 is added to the correlation score if the strain produces the metabolite and has a BGC in the GCF; 10 is subtracted from the score if the metabolite is produced by the strain but it lacks a BGC in the GCF; One (1) is added to the score if the metabolite is not produced by the strain and the strain has no BGC in the GCF; The score is left unchanged if the strain has a BGC present in the GCF but does not produce the metabolite.

This method is easy to interpret but is heavily dependent on total population size, GCF size and the number of strains that produce the metabolite, making accurate prediction of links challenging (Hjörleifsson Eldjárn *et al*., 2021). To combat this, an open-source software network was developed, known as NP Linker, which accepts both locally hosted datasets from the likes of antiSMASH, BiGSCAPE and GNPS and datasets uploaded to the Paired-Omics Data Platform (Schorn *et al*., 2021) and utilises Input-Output Kernel Regression (IOKR) and multi-level scoring methods for a more accurate inference of links between BGCs and produced metabolites (**Figure 1.14**) (Hjörleifsson Eldjárn *et al*., 2021).
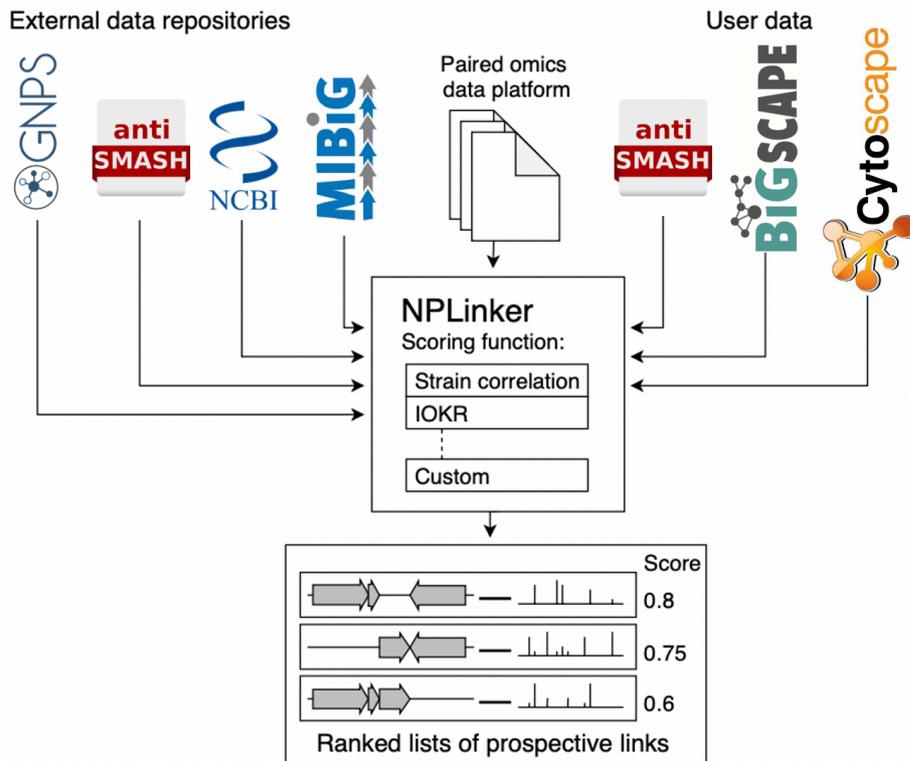
Figure 1.13 – The NP Linker scoring pipeline. NP Linker can use input data from external repositories such as GNPS (M. Wang *et al*., 2016), antiSMASH (Medema *et al*., 2011) and MIBiG (Medema *et al*., 2015), gather data from the Paired Omics Data Platform (Schorn *et al*., 2021), utilize entirely local user data generated via BiGSCAPE (Navarro-Munoz *et al*., 2020) or Cytoscape (Shannon *et al*., 2003) or use a combination of all. The platform then applies multiple scoring methods and ranks prospective links.

1.9 Aims and Objectives

<u>Aim 1</u>

Metabologenomics methods were explored to evaluate and compare the relationships between complex genomic and metabolomic datasets gathered from fully sequenced Actinomycetota across diverse culture conditions. A miniaturised assay was developed to generate high quality data for LC-MS/MS analysis at small scale.

Objective 1 – To link BGCs to metabolites of Actinomycetota using metabologenomics.

Objective 2 – To optimise a miniaturised extraction method to generate >1 mg/mL of extract for LC-MS/MS analysis.

<u>Aim 2</u>

Coculture of *Streptomyces* in natural soil microcosm systems was utilised to elicit the production of potentially novel bioactive specialised metabolites.

Objective 1 – To establish when *Streptomyces* strains enter stationary phase.

Objective 2 – To test the bioactivity of microbial extracts from *Streptomyces* cocultured within natural soil microcosm systems and compare to the chemistry of monocultures.

Objective 3 – To use metabologenomics to target elicited bioactive specialised metabolites for further analysis.

<u>Aim 3</u>

Standardised artificial soil microcosm systems were used to enhance reproducibility of data from *Streptomyces* coculture.

Objective 1 – To develop and optimise artificial soil microcosm systems based on the composition of natural soil.

Objective 2 – To analyse and compare chemistry of *Streptomyces* cocultures between natural and artificial soil microcosm systems.

Chapter 2 – Materials and Methods

2.1 Strains

Table 2.1 – Full list of strains used within study

Strain	Genotype	Reference
Micromonospora echinospora ATCC 15837	Wild type	ATCC
Micromonospora chalcea ATCC 12452	Wild type	ATCC
Nocardia farcinica NBRC 15532	Wild type	NBRC
Streptomyces collinus Tü 365	Wild type	Rückert, 2013
Streptomyces lividans 1326	Wild type	ATCC
Streptomyces lividans TK23	*spc-1*, SLP2–, SLP3–	ATCC
Streptomyces coelicolor M145	Plasmidless (SCP1− SCP2−)	Redenbach, 1996
Streptomyces coelicolor M1146	Δ*act* Δ*red* Δ*cpk* Δ*cda*	Gomez- Escribano and Bibb, 2011
Streptomyces coelicolor M1152	Δ*act* Δ*red* Δ*cpk* Δ*cda* Δ*rpoB*	Gomez- Escribano and Bibb, 2012
Streptomyces coelicolor M1154	Δ*act* Δ*red* Δ*cpk* Δ*cda* Δ*rpoB* Δ*rpsL*	Gomez- Escribano and Bibb, 2012
Streptomyces venezuelae ATCC 10712	Wild type	DSMZ 40230
Streptomyces clavuligerus ATCC 27064	Wild type	ATCC
Streptomyces scabiei 87.22	Wild type	Bignell, 2010
Streptomyces goldiniensis ATCC 21386	Wild type	ATCC
Streptomyces noursei ATCC 11455	Wild type	DSMZ 40635
Streptomyces rimosus ATCC 10970	Wild type	ATCC
Streptomyces fradiae ATCC 10745	Wild type	DSMZ 40063
Streptomyces kanamyceticus ATCC 12853	Wild type	DSMZ 40500

Streptomyces nodosus ATCC 14899	Wild type	DSMZ 40109
Streptomyces spectabilis ATCC 27465	Wild type	DSMZ 40512
Streptomyces vinaceus ATCC 27476	Wild type	DSMZ 40515
Streptomyces platensis ATCC 23948	Wild type	DSMZ 40041
Enterococcus faecium ATCC 51299	Wild type	ATCC
Staphylococcus aureus ATCC 43300	Wild type	ATCC
Klebsiella pneumoniae ATCC 70603	Wild type	ATCC
Acinetobacter baumannii ATCC 19606	Wild type	ATCC
Pseudomonas aeruginosa ATCC 27853	Wild type	ATCC
Escherichia coli ATCC 25922	Wild type	ATCC
Bacillus subtilis ATCC 23857	Wild type	ATCC

Table 2.2 – S. coelicolor deletion mutants used for chemical comparison within analysis.

Strain	Gene alteration	Description
S. coelicolor M145	None	-
S. coelicolor M1146	Δ*act*	Actinorhodin cluster knockout
	Δ*red*	Prodiginine cluster knockout
	Δ*cpk*	CPK cluster knockout
	Δ*cda*	CDA cluster knockout
S. coelicolor M1152	Δ*act*	Actinorhodin cluster knockout
	Δ*red*	Prodiginine cluster knockout
	Δ*cpk*	CPK cluster knockout
	Δ*cda*	CDA cluster knockout
	rpoB [C1298T]	rpoB point mutation
S. coelicolor M1154	Δ*act*	Actinorhodin cluster knockout
	Δ*red*	Prodiginine cluster knockout
	Δ*cpk*	CPK cluster knockout
	Δ*cda*	CDA cluster knockout
	rpoB [C1298T]	rpoB point mutation
	rpsL [A262G]	rpsL point mutation

2.2 Culture media

Table 2.3 – List of culture media recipes used within study

Media	Components (per litre)	pH	Reference
Glucose Yeast Malt Medium (GYM) (DSMZ Medium 65)	10 g malt extract, 4 g glucose, 4 g yeast extract, 2 g CaCO3	7.2	Kieser *et al*., 2000
Soya Flour Mannitol (SFM/MS)	20 g mannitol, 20 g soya flour (Holland & Barrett), 10 mM CaCl	7.1	Hobbs *et al.,* 1989
International *Streptomyces* Project 2 (ISP2) Medium	10 g malt extract, 4 g yeast extract, 4 g dextrose	7.2	Shirling & Gottlieb, 1966
Lysogeny Broth (LB)	10 g tryptone, 5 g NaCl, 5 g yeast extract	7.5	Sambrook, Fritsch & Maniatis, 1989
Difco Nutrient Agar (NA)	4 g Difco nutrient broth powder	6.8	Difco
Gause's No. 1	20 g soluble starch, 1 g KNO_3, 0.5 g NaCl, 0.5 g K_2HPO_4, 0.5 g $MgSO_4$, 0.01 g $FeSO_4$	7.4	DSMZ
2x YT	16 g tryptone, 10 g yeast extract, 5 g NaCl	7	Kieser *et al*., 2000

2.3 Strain selection, genome quality assessment and genome mining of Actinomycetota

Available *Streptomyces* genome sequences were assessed for quality and coverage using the NCBI database. The first set of nine Actinomycetota were selected from the Strathclyde culture collection, and the second set of ten were purchased from the DSMZ culture collection. Strains with coverage below 50x or a genome size less than 7.5 Mbp were discounted. antiSMASH v6 (K. Blin *et al*., 2021) for identification and annotation of BGCs. Strains with less than 20 and/or no uncategorised BGCs were also discounted. The output was visualised using the R geom_ribbon function within the ggplot2 package. BiGSCAPE-CORASON (Navarro-Munoz *et al*., 2020) was used to cluster the BCG's into GCFs and the output was visualised using Cytoscape (Shannon *et al*., 2003).

2.4 Phylogenetic analysis of Actinomycetota strains

Full genome sequences were obtained from the National Centre for Biotechnology Information (NCBI) with accession numbers as follows: *M. echinospora* - LT607413.1*; M. chalcea* - MAGP01000001.1*; N. farcinica* - CP031418.1*; S. collinus* – CP006259.1*; S. lividans* 1326 – NZ_CM001889.1*; S. coelicolor* – AL645882.2*; S. venezuelae* – NC_018750.1*; S. clavuligerus* CP027858.1*; S. scabiei* 87.22 – FN554889.1. A maximum likelihood phylogenetic tree was created using autoMLST (Alanjary, Steinke, & Ziemert, 2019) and was constructed using multi-locus sequence analysis (1000 replicates, IQ-TREE Ultrafast Bootstrap analysis, denovo construction mode). The tree was then visualised and edited with TreeViewer.

Full genome sequences for strains purchased from the DSMZ culture collection were retrieved from NCBI using the following accession numbers: *S. noursei* ATCC 11455 – CP011533.1, S. rimosus ATCC 10970 – CP025551.1, *S. clavuligerus* ATCC 27064 - CP0227858.1, *S. fradiae*

ATCC 10745 - CP023696.1, *S. venezuelae* ATCC 10712 - CP029197.1, *S. kanamyceticus* ATCC 12853 - CP023699.1, *S. nodosus* ATCC 14899 - CP023747.1, *S. spectabilis* ATCC 27465 - CP023690.1, *S. vinaceus* ATCC 27476 - CP023692.1, *S. platensis* ATCC 23948 - CP023691.1. Phylogeny was assessed as above. The output was visualised and annotated using the Interactive Tree of Life (iTOL) (Letunic & Bork, 2019).

2.5 Spore stock preparation of *Streptomyces* strains

Streptomyces strains purchased from the DSMZ culture collection were prepared for use as follows: The tip of the surrounding glass ampoule was heated over the blue flame of a Bunsen burner, after which a few drops of water were placed on the ampoule to crack the glass which was struck off using steel forceps. The insulation material was then removed using the forceps, followed by the inner vial. 0.5 mL liquid GYM media (recommended by DSMZ for pellet rehydration of the *Streptomyces* strains) was placed on the pellet in the inner vial and the pellet was allowed to rehydrate (30 mins). The content was mixed gently with half used to inoculate 5 mL liquid GYM media in universal tubes and the other half streaked in triplicate onto solid GYM media. The cultures were then incubated (liquid – 5 days, 30°C, 250 rpm; solid – 5 days, 30°C) and the plates were checked for morphology and purity. To prepare spore stocks of *Streptomyces* strains purchased from DSMZ, 50 µL liquid culture was used to inoculate a confluent lawn on GYM agar. After 10 days and the formation of spores from all strains, 5 mL ddH$_2$O was added to cover the lawn and the spores. A sterile cotton bud was then used to displace the spores and suspend them in water. The spore suspension was then transferred to a syringe stuffed with sterile cotton wool and passed through to remove mycelial debris. The spores were then pelleted by centrifugation at 4000g for 15 minutes before the water was removed and the pellets resuspended in 1.5 mL 40% glycerol and stored at -80°C.

2.6 Acquisition of mass spectrometry data from Actinomycetota

For the Actinomycetota selected in Chapter 3, LC-MS/MS was performed using the Thermo Scientific Accela LC system coupled to a Thermo Finnigan Orbitrap mass spectrometer with an ESI source. Both bacterial and non-bacterial control metabolite extracts were prepared to 1 mg/mL in ACN before injection onto an ACE 5 (Hichrom) C18 column (5 μm, 75 x 3.0 mm) using a gradient of: 1-5 mins (5% ACN in H_2O), 5-25 mins (5-100% ACN), 25-30 mins (100% ACN). Mass data was collected in positive ion mode using ESI and a mass range of 150-1500 *m/z* (15,000 resolution). Data-dependant MS2 scans were obtained using Collision-Induced Dissociation (CID) using an energy of 35 eV with an activation time of 30,000 ms for the first, second and third most intense peaks.

For chapters 4 and 5, C18 untargeted metabolomics analysis with the Exploris 240 instrument was performed: Extracts were randomized and analysed by high-performance liquid chromatography-electrospray ionization quadrupole obritrap mass spectrometry (HPLC-ESI-HRMS) using a Thermo Vanquish binary LC system coupled to a Thermo Exploris 240 orbitrap mass spectrometer. 5 µL of each extract was injected (15 µL) onto an Accucore C18 HPLC Column (2.6µm, 100 mm×2.1mm I.D. (Thermo)). Mobile phase A consisted of 0.1% formic acid in water (v/v). Mobile phase B consisted of 0.1% formic acid in acetonitrile (v/v). Each sample was subsequently run on the following solvent gradient:

Retention (Min)	Flow (mL/min)	%A	%B	Curve
0	0.3	99	1	5
0.5	0.3	99	1	5
2	0.3	50	50	5
10.5	0.3	1	99	8
11	0.3	1	99	5
11.5	0.3	99	1	7
14.9	0.3	99	1	5
15	0.3	99	1	5

Mass spectrometry analysis was performed under the following conditions: ESI source conditions were set as follows: Ion spray voltage (pos 3700V) sheath gas 40, aux gas 10, sweep gas 1, ion transfer tube temperature 300°C, vaporizer temperature 280°C. The instrument was set to acquire over the *m/z* scan range of 70–1050, at 60k resolution and with and RF lens of 70%.

Other full scan filters included: exclusion override factor = 3, exclusion list peak window extension = 3, inclusion list peak fragmentation threshold = 80, preferred ions =[M+H]+1;[M-H]-1;[M+Na]+1, exclusion duration = 10. MS/MS settings were as follows: resolution =60k, FR Lens(%) = 65, EASY-IC = on, Intensity Filter = 5x1.0E3, Precursor Fit Filter = Fit Threshold (%) 51 and Fit Window (m/z) 0.7, Charge State = 1-2 (including undetermined charge states), Custom Dynamic Exclusion = exclude after n times = 1; exclusion duration = 3s; mass tollerance = ppm, low = 2; high = 2; exclude isotope = on, no target or exclusion masses used, desired apex window (%) = 50. ddMS2 was set at 5 scans, isolation window = 0.7, collision energy = normalized, resolution = 30k, scan range mode = auto.

2.7 Mass spectrometry data processing

MZmine v2.53 freeware (http://mzmine.sourceforge.net/ - accessed on March 14th 2021) was used to process mass spectrometry data, employing peak detection, deconvolution, deisotoping, filtering, alignment and gap-filling to allow comparison of data files. For Chapter 3's data, m/z tolerance was 0.01, minimum peak detection was set at 3.00E3 and minimum timespan/tR tolerance was set at 0.1 mins for the duration of data processing. A centroid mass detector was used for mass detection with the noise level set at 2.00 x 1.0E3. For Chapter 4's soil microcosm system data, mass detection was undertaken via a centroid mass detector set at 1.00E3 for MS level 1 and 1.00E1 for MS level 2. Mass to charge (*m/z*) tolerance was set at 0.01, with peaks detected above a threshold of 3.00E3. Minimum time span was set to 0.1.

tR tolerance was set to 0.1. The same method was then applied to data acquired in Chapter 5. MS/MS data was converted from raw to mzML format using FileZilla v3.52.2 (https://filezilla-project.org/ - accessed on March 31st 2021) with the converted data uploaded to the GNPS server (M. Wang *et al*., 2016).

2.8 GNPS molecular networking of mass spectrometry data

The data acquired in Chapter 3 was submitted to the MassIVE Public GNPS database and assigned the access number MSV000084762. A classic molecular network was created using the online workflow (https://ccms-ucsd.github.io/GNPSDocumentation/) on the GNPS website (http://gnps.ucsd.edu - accessed on March 31st 2021). The data was filtered by removing all MS/MS fragment ions within +/- 17 Da of the precursor m/z. MS/MS spectra were window filtered by choosing only the top 6 fragment ions in the +/- 50Da window throughout the spectrum. The precursor ion mass tolerance was set to 0.01 Da and a MS/MS fragment ion tolerance of 0.5 Da. A network was then created where edges were filtered to have a cosine score above 0.7 and more than 6 matched peaks. Further, edges between two nodes were kept in the network if and only if each of the nodes appeared in each other's respective top 10 most similar nodes. Finally, the maximum size of a molecular family was set to 100, and the lowest scoring edges were removed from molecular families until the molecular family size was below this threshold. The spectra in the network were then searched against GNPS' spectral libraries. The library spectra were filtered in the same manner as the input data. All matches kept between network spectra and library spectra were required to have a score above 0.7 and at least 6 matched peaks. Cytoscape version 3.8 (Shannon *et al*., 2003) was used to visualise the molecular network, with each node corresponding to a consensus spectrum and each edge conforming to a modified cosine similarity score between pairs of nodes. For Chapter 4 and 5 data, no submission to MassIVE was made and the files remained private.

2.9 Computational pattern matching using NP Linker

The platform NPLinker was utilised to computationally prioritise links between BGCs and produced metabolites, by way of two separate scoring methods: Metcalf and Rosetta, with each system outlined in - (Hjörleifsson Eldjárn *et al.,* 2021). Metcalf is a version of a standardised strain correlation score which overcomes the limitations of previous correlation scoring methods which were heavily skewed by the number of strains present in each dataset (Doroghazi *et al.*, 2014). Rosetta is a novel approach developed by the Eldjarn, *et al.* (code and data preparation steps available at the following location: https://github.com/sdrogers/nplinker/tree/master/prototype/rosetta_data_prep) which is based on a set of matches between GNPS library spectra (Wang, et al., 2016) and the MIBiG database (Medema *et al.*, 2015) containing 2960 links, 2069 unique spectra and 249 unique MIBiG IDs. Putatively, the observed spectra and BGCs were matched as follows: spectral similarity between measured MS2 spectra and the relevant subset of the GNPS spectra was computed using the modified cosine score (equivalent to "Analog search" in the GNPS framework). Results from antiSMASH were parsed to extract the known cluster blast results and Rosetta links between spectra and BGCs were generated where the spectra showed similarity to the GNPS spectrum and the MIBiG entry was found in the known cluster blast record for the BGC. When performing analysis via NPLinker, each scoring system can be used individually or in concert, with thresholds defined by the user based upon the desired analytical parameters (Hjörleifsson Eldjárn *et al.*, 2021).

2.10 Scale reduction testing for *Streptomyces* culture and metabolite extraction

S. fradiae and *S. spectabilis* cultures (half ISP2 broth, half solid ISP2, triplicate) were set up for extraction in 24-well plates (Axygen) and 6-well plates (Thermo Scientific) which included ISP2 media blank wells. After three days, the contents of each well of the 6-well plates were

chopped using a scalpel and transferred to a test-tube, frozen (-80°C, overnight) and freeze dried (7 hrs, LABCONCO Freezone 2.5 drier). The contents were then transferred to an Erlenmeyer flask (250ml). 24-well plates were frozen and freeze dried without the contents being removed. Ethyl acetate (10ml (6-well plate), 2ml (24-well plate), Fisher) was added to each flask, and each well of the 24-well plate. The flasks and plates were shaken (120 rpm, 2 hrs, Stuart SSL1 orbital shaker) with the supernatant then removed, dried under N_2 and the extract weight recorded (mg). The process was then repeated using only 24-well plates, with GYM and NMMP *Streptomyces* minimal liquid media. The carbon source was n-acetylglucosamine solution prepared to 10 and 50 mM. Each concentration was set in triplicate, along with cultures at 0 mM and media blanks.

2.11 Gravimetric dry mass analysis of *Streptomyces*

For reduced scale culture, spores of *S. spectabilis, S. vinaceus, S. fradiae, S. vinaceusdrappus, S. noursei, S. nodosus* and *S. kanamyceticus* were pre-germinated in baffled Erlenmeyer flasks containing 50 mL GYM media (30°C, 250 rpm, 48 hrs). 14 µL seed stock was inoculated to 7 mL liquid GYM media in 24-well plates (Axygen) along with media blanks. A sterile glass bead was added to each well to disperse mycelia during shaking. The plates were covered with a gas-permeable membrane (Breathe-Easy® sealing membrane) and incubated (30°C, 250 rpm). An initial 1 mL aliquot was taken 4 hrs post-inoculation, and at 24 hr intervals therein for a total of 144 hrs. 1 mL was removed from each well and transferred to a pre-dried and pre-weighted Whatman-grade microfiber filter paper. The culture was pulled through the filter paper via vacuum and Buchner funnel until only the dry cell mass remained on the paper. The filter papers were dried in an oven (18 hrs, 55°C) and re-weighted. Growth curves were plotted, and stationary phase onset was marked using Microsoft Excel.

For dry mass analysis of 50 mL cultures, spores of seven *Streptomyces* species (*S. spectabilis, S. vinaceus, S. fradiae, S. vinaceusdrappus, S. noursei, S. nodosus, S. kanamyceticus*) were pre-germinated in Erlenmeyer flasks containing 50 mL GYM *Streptomyces* media (30°C, 250 rpm, 48 hrs). 100 μL of the seed stock was inoculated in triplicate to Erlenmeyer flasks, again containing 50 mL GYM *Streptomyces* medium (three flasks per strain, totalling 21 flasks). The protocol was followed as above from this point onwards.

2.12 Soil collection and processing

To ensure soil conditions remained the same throughout all of the natural soil experiments, a 5 kg batch of soil categorised as brown earth was harvested from underneath a tree at Drumpellier Country Park, Coatbridge, North Lanarkshire (grid ref: NS7086265133, Lat 55°51'43"N, Long 004°03'54"W). The soil was firstly separated into smaller 1 kg batches and air dried for two days in a drying drawer, before being transferred to an oven set at 80°C and left overnight. Once fully dried, the soil was passed through a 2 mm pore sieve to remove larger particles and separated into 20 g samples for amending and autoclaving. 1 % wt/wt soluble starch (Fisher Scientific) and chitin (Sigma Life Science) were added to each 20 g sample before being UV sterilized and autoclaved twice.

2.13 Soil composition analysis

To determine the composition of the collected soil, analysis was outsourced to the James Hutton Institute (James Hutton Limited, Craigiebuckler, Aberdeen, Scotland, AB15 8QH). Soil pH testing, loss on ignition, total carbon and nitrogen content, particle size determination and aqua regia digest for elemental composition tests were conducted. The soil was burned at 450°C, which is generally accepted to be above the temperature required to burn away all

organic matter. The pH of the soil was tested with both H_2O and $CaCl_2$, with the latter considered the more accurate due to it being less affected by soil electrolyte concentration (Minasny, McBratney, Brough, & Jacquier, 2011). 3 g of soil was oven-dried (overnight, 80°C) and milled to a fine powder using a mortar and pestle. The sample was packaged in a Falcon tube and posted for analysis.

2.14 *Streptomyces* challenge assays

A set of challenge assays were used to prioritise *Streptomyces* strains for soil microcosm cocultures. In groups of five strains, cultures were set up in which one strain would be inoculated in a cross formation, dividing the plate into four sections. The initial group of five strains were selected as they reached stationary phase in a similar timeframe to one another. Cultures were inoculated on Gause's no. 1 and GYM media as well as both media types supplemented with natural soil as previously described. A score was assigned as follows: if inhibition of < 1 cm occurred – 1 pt; > 1 cm – 2 pts; complete inhibition – 4 pts. The strain interactions that scored highest were prioritised for further analysis.

All strains were precultured and inoculated once they reached stationary phase (5 days, 30°C). The protagonist was inoculated first (50 μL of preculture, 1 cm diameter cross) and incubated (5 days, 30°C). This was followed by the antagonist strains (25 μL of preculture in corner sections) and the plates were again incubated (5 days, 30°C). Each plate was set up in triplicate, with positive controls of the protagonist strain alone, and negative controls with the protagonist strain *in absentia* were set up.

2.15 Soil microcosm bacterial monoculture and coculture

To determine if an inoculated culture was pure and inoculation was effective in soil microcosms, post incubation checks were conducted. Twice-autoclaved amended soil (20 g, Falcon tubes) was inoculated with *S. spectabilis* (due to its undecylprodigiosin pigment production. 10^4 CFU/g in 3 mL ddH$_2$O, to raise the moisture level in the soil to 15% wt/vol). The spore suspension was then added to the soil in triplicate, and incubated (10 days, 30°C). Post-incubation, each microcosm was homogenised mechanically. 1 g of soil was removed and transferred to solid GYM agar. A negative control of uninoculated natural soil was established to ensure soil sterility. The plates were then incubated (5 days, 30°C). For the final cultures, amended soil microcosms were prepared as previously described. Negative controls of soil blanks and each strain grown in monoculture were established, with each culture performed in triplicate. The microcosms were then incubated (10 days, 30°C).

2.16 Microcosm culture metabolite extraction, fractionation, and bioactivity testing

Each microcosm was freeze-dried (Labconco FreeZone 2.5 benchtop freeze-drier), with the soil matrices transferred to Erlenmeyer flasks and extracted (EtOAc, 25mL), shaken (overnight, 180 rpm) and dried under N$_2$ with the dry extract weight recorded. The samples were then fractionated using Solid Phase Extraction (SPE) via the following steps: 2 mL MeOH/H$_2$O (80/20) added to extract; Fraction 1: 10 mL MeOH/H$_2$O (80/20); Fraction 2: 10 mL MeOH; Fraction 3: 10 mL ACN; Fraction 4: 10 mL EtOAc. Each fraction was dried (N$_2$) and the weight was recorded. The samples were resuspended in EtOAc to 1 mg/mL for bioactivity analysis.

For bioassays, overnight cultures of *E. coli* and *S. aureus* were inoculated (5 mL), with the OD$_{600}$ calculated and used to inoculate soft nutrient agar (25 mL). This was then poured as a

top layer to nutrient agar (NA) plates. Extract (30 μL) was added to a paper disc which was transferred to the top layer of SNA. The plates were incubated (overnight, 37°C) and checked for zones of inhibition.

2.17 Isolation, bioactivity screening and light microscopy of Actinomycetota from Drumpellier Park soil

Soil collected from previously described coordinates in Drumpellier Park was heat shocked (80°C, one hour) to preferentially select for Actinomycetota. 1 g of soil was added to 9 mL Ringer's solution and vortexed (10 mins). The solution was then serially diluted to 1/10, 1/100 and 1/1000. Each dilution (inc. undiluted) was plated in triplicate (100 microlitres spread on Gause's no. 1 media with nystatin, nalidixic acid and cycloheximide). Seven distinct colonies with Actinomycetota-like morphologies were aseptically picked and subcultured on Gause's no. 1, MS and GYM media until pure. Isolate 3 would not grow in culture so was excluded from further analysis. Spore suspensions of each isolate were established and stored at -80°C.

Each isolate was screened for bioactivity against *B. subtilis,* a common soil microbe and indicator of Actinomycetota bioactivity. Spore suspension (50 μL) was inoculated to 2xYT medium (5 mL) as preculture and incubated (5 days, 30°C, 250 rpm). Preculture (50 μL) was inoculated to a petri dish of MS medium and spread as a lawn before incubation (5 days, 30°C). Square plates of nutrient agar (NA) media were poured. For *B. subtilis* pre-culture, one colony of an active plate was inoculated to LB broth (5 mL) and incubated (overnight, 37°C, 250 rpm). The OD$_{600}$ was read the following morning and was used to calculate the amount to be added to soft nutrient agar (SNA) (20 mL). The inoculated SNA was added as a top layer to the NA square plates, with each plate divided into sections accommodating a plug from each isolate as well as positive and negative controls. The wide end of a blue 1 mL pipette tip

was used to cut plugs from the lawns of each isolate and each plug was transferred to the appropriate section of the square plate. Spectinomycin was used as a positive control and uninoculated MS media was used as a negative control. In place of DS3, *S. noursei* plugs were used as a positive microbial control due to the strains known activity against *B. subtilis.* The plates were then incubated (overnight, 37°C).

To test the inhibitory profile of the isolates against other Streptomycetes and to inform whether or not the strains will be cocultured within the soil/artificial soil microcosm systems, cross inhibition assays (of the same format used in Chapter 2) were established across various types of media. Precultures of isolates with distinct phylogeny (1 and 5), as well as *S. spectabilis* and *S. noursei,* were set up (5 mL, 2xYT media, 30°C, 5 days). Each strain was inoculated (50 μL) to the middle cross section of the plate, with the other strains inoculated (25 μL) inoculated to each sector around the cross.

For light microscopy, DS1 and DS5 were cultured in 5 mL 2xYT media and incubated (250 rpm, five days, 30°C). In triplicate, 1 mL of culture was added to 1.5 mL Eppendorf tubes and centrifuged to pellet the cells (4000 rpm, 4 mins). The supernatant was discarded, and 500 mL methylene blue solution (1% wt/vol methylene blue/water) was added to the tube. The solution was vortexed (30 secs) to ensure homogenous staining of cells and centrifuged (4000 rpm, 4 mins) to pellet. The cells were washed twice with cold methanol with 10 μL added to a microscope slide and covered with a glass coverslip. The suspension was then observed via an inverted microscope (Nikon Eclipse TE2000-S) under oil immersion.

2.18 Genomic DNA extraction, polymerase chain reaction and phylogeny of soil isolates

All isolates were cultured in 5 mL 2x YT media (30°C, 250 rpm, 5 days). 1 mL of each culture was transferred to a 1.5 mL centrifuge tube and centrifuged (4000 rpm, 10 mins). The

supernatant was discarded, and pellets were resuspended in 111 microlitres Tris (50 mM), 111 microlitres EDTA (20 mM), 75 microlitres lysozyme (20 mg/mL) and 3 microlitres RNAse before incubation (30°C, 30 mins). 50 μL 10% SDS was added to the tubes and mixed thoroughly, followed by 85 μL NaCl and 400 μL phenol/chloroform/isoamyl alcohol before vortexing (30 secs) and further centrifugation (8000 rpm, 10 mins). The aqueous phases were then transferred to new tubes with 500 μL isopropanol added and mixed by inversion before resting at room temperature for 5 minutes. A final centrifugation step was utilized (10,000 rpm, 10 mins) to pellet DNA and the isopropanol was discarded. The pellets were washed with cold 70% EtOH before resuspension in 50 μL nuclease free water before nanodrop testing confirmed the presence of gDNA.

For PCR of 16S rRNA, 50 μL reactions were created by mixing, 25 μL REDTaq® ReadyMix™ PCR Reaction Mix, 2.5 μL each forward (5'- AGA GTT TGG ATC MTG GCT CAG -3') and reverse (5'- CGG TTA CCT TGT TAC GAC TT -3') 16S primers, 2 μL extracted gDNA and 18 μL nuclease free water. The PCR was performed in the Applied Biosystems Veriti 96-well thermal cycler, to the following protocol:

Stage	Temperature (°C)	Time	
Initial Denaturation	95	1 min	
Denaturation	95	15 secs	
Extension	62	15 secs	35 cycles
Annealing	72	30 secs	
Final Extension	72	2 mins	
Hold	4	∞	

The subsequent reactions (5 μL) were then run via an agarose/ethidium bromide gel (60 mins, 90 v) along with on well containing a mix of 4 μL Promega 1kb DNA ladder and 1 μL Promega Blue/Orange 6X loading dye. The gel was visualised under UV and the presence of bands conforming to 1500 bp on the ladder was confirmed.

Due to the presence of residual bands on the gel, a further PCR clean-up step was required. For this, the Promega Wizard® SV Gel and PCR Clean-Up System was used. The remainder of each PCR reaction (40 μL) was added to the wells of an agarose/ethidium bromide gel and run (60 mins, 90 v). Once separated, the 1500 bp bands representing the 16S gene were excised using a scalpel and the gel slices were prepared as per the kit protocol. The resultant pure DNA samples were again checked on the nanodrop to ensure the nucleic acid concentrations fell within the correct boundaries for Eurofins 16S sequencing. The gDNA samples were at such a concentration that they did not need to be diluted, and the 260/280 ratio fell between 1.8 and 2.0, indicating the samples are pure.

The samples were prepared as per the Eurofins guidelines (15 μL gDNA suspension + 2 μL F/R primer). Each sample was resuspended with forward and reverse primers separately to increase the chances of successful sequencing. The samples were sent to Eurofins for 16S sequencing, where they are sequenced using the cycle sequencing technology (dideoxy chain termination/cycle sequencing) on ABI 3730XL sequencing machines.

Sample No.	Primer	Barcode
1	F	TS02687806
2	F	TS02687807
4	F	TS02687808
5	F	TS02687809
6	F	TS02687810
7	F	TS02687811
1	R	TS02687812
2	R	TS02687813
4	R	TS02687814
5	R	TS02687815
6	R	TS02687816
7	R	TS02687817

A phylogenetic tree was constructed using the Tamura-Nei method (Tamura & Nei, 1993) and bootstrap resampling (1000 Bootstrap value) via the MEGAX package (S. Kumar, Stecher, Li, Knyaz, & Tamura, 2018). 'F' signifies that the sample was prepared using the forward primer, 'R' signifies that the reverse primer was used.

2.19 Bacterial culture and bioactivity screen of artificial soil microcosm systems

Natural soil collected in **2.14** was prepared as described in **2.14.** *S. spectabilis*, *S. noursei,* DS1 and DS5 were pre-cultured in 2xYT media until stationary phase transition. To prepare the bead-based artificial soil, a solution composed of 7.5 g/L^{-1} Phytagel™ and 2.4 g/L^{-1} sodium alginate was dropped into a solution of 2% $CaCl_2$ via sterile syringe, constantly stirred to prevent a film forming. Contact between the solutions resulted in the rapid formation of spherical beads, from which excess liquid was drained. The gel beads were then separated into 20 g aliquots, with each aliquot equilibrated for two hours with either 2xYT media or a trace element nutrient solution based on the results of previously conducted soil composition analysis ($AgNO_3$ – trace, $Pb(NO_3)_2$ – 235.2 mg/kg, $CoCl_2$ – 15.96 mg/kg, $CuCl_2$ – 60.89 mg/kg, $MoO3.H_2O$ – 1.53 mg/kg, $NiCl_2$ – 43.64 mg/kg, $NaHSeO_3$ – 2.55 mg/kg, $Sr(NO_3)_2$ – 86 mg/kg, $ZnSO_4$ – 370 mg/kg, $Ba(NO_3)_2$ – 278 mg/kg, 1 % wt/wt soluble starch and chitin). The excess liquid was again drained before inoculation.

For the sand-based media, inert silica sand was passed through a 2 mm pore sieve to remove larger particles and acid washed with 1% HCl to displace cations from cation exchange sites before rinsing with ddH_2O to remove residual chlorine. These steps are performed to allow complete control over the nutrient environment. The sand was then separated into 20 g aliquots and shaken with previously described trace element nutrient solution before being allowed to rest before inoculation.

For each microcosm system, 10^4 CFU/mL spores were added to 3 mL ddH$_2$O which was then added to the microcosm as the inoculum, in turn adjusting the moisture content of the microcosm to 15% wt/vol for all conditions except sand, where the spores were inoculated directly into the nutrient suspension. Where strains were grown in coculture, spores of both strains were added to ddH$_2$O for inoculation. Negative controls of soil, gel bead and sand blanks were established, with each culture performed in triplicate. The microcosms were then incubated (10 days, 30°C). Each microcosm was freeze-dried (Labconco FreeZone 2.5 benchtop freeze-drier), with the microcosm matrices transferred to Erlenmeyer flasks and extracted using ethyl acetate (EtOAc) (25mL to cover the matrices in the flask, shaken overnight at 180 rpm). After shaking, the samples were filtered by allowing the organic phase to pass through filter paper to be collected in pre-weighed glass scintillation vials, which were then dried under N$_2$ with the dry extract weight calculated.

To establish the bioactivity profiles of the microcosm systems, extracts were tested against members of the ESKAPE pathogen group. Each extract was dissolved in EtOAc to a concentration of 1 mg/mL. Overnight cultures of ESKAPE and *B. subtilis* were inoculated, with the OD$_{600}$ calculated and used to inoculate soft nutrient agar (SNA). This was then poured as a top layer to nutrient agar (NA) plates. 30 μL of each extract was added to a paper disc and each disc was transferred to the appropriate section on the top layer of SNA. The plates were incubated (overnight, 30°C) and checked for zones of inhibition.

Chapter 3 – Metabologenomics approaches for antibiotic discovery from *Streptomyces*

3.1 - Introduction

The phylum Actinomycetota comprises over 250 genera, many of which are a source of bioactive specialised metabolites (Ludwig *et al*., 2012), accounting for over 80% of all microbially produced antibics (Iliĉ *et al*., 2007). The genus *Streptomyces* are particularly biosynthetically-talented, with over six thousand specialised metabolite Gene Cluster Families (GCFs) unique to the genus (Gavriilidou *et al*., 2022). This makes *Streptomyces* an ideal candidate for novel drug discovery pipelines (Antoraz, Santamaría, Díaz, Sanz, & Rodríguez, 2015). *Streptomyces* sp. dedicate between 0.8 to 3.0 Mbp coding capacity to specialised metabolite function, with species harbouring between 20 and 50 Biosynthetic Gene Clusters (BGCs) per strain (Baltz, 2017) making manually mining their genomes a Herculean task. Mining these extensive datasets manually is fast becoming a bottleneck in discovery creating a need for *in silico* methods to be developed for the assessment of biosynthetic potential.

In the 2010's, there was an explosion of the development of bioinformatic genome mining tools, with analysis pipelines such as DeepBGC (which detects BGCs in bacterial and fungal genomes using neural networks for deep learning (Hannigan *et al*., 2019)) and PRISM (which identifies natural product BGCs using a structure prediction algorithm (Skinnider, Merwin, Johnston, & Magarvey, 2017)). Currently, the gold standard for annotation and analysis of BGCs is antiSMASH, which can detect BGCs across all known classes of specialised metabolite, as well as recognising gene similarities in the evolution of the query cluster and previously detected BGCs (Medema *et al*., 2011). Comprehensive analysis under the antiSMASH umbrella has been enhanced by the incorporation of genome mining tools. For instance, the most recent version 6.0 (K. Blin *et al*., 2021) contained updates which included access to the TIGRFAMs database to scan for protein family definitions (Haft *et al*., 2013).

antiSMASH can also predict the identity of a specific BGC by sequence comparison to experimentally verified BGCs submitted to the MIBiG database, an open-source library of known BGCs (Kautsar *et al*., 2020). Since initial publication (as of May 2023), antiSMASH has processed over 1.5 million jobs, cementing its place as an invaluable BGC mining platform for the discovery of specialised metabolites. Recently, antiSMASH-based genome mining was used to identify the structural gene and Open Reading Frames (ORFs) present in the novel humidimycin BGC from *Streptomyces humidus* CA-100629 (Sánchez-Hidalgo, Martín, & Genilloud, 2020). The discovery of the novel cyclodepsipeptide antibiotic atratumycin, which displayed bioactivity against *Mycobacterium tuberculosis* H37Ra and H37Rv, was enabled through antiSMASH analysis of *Streptomyces atratus* SCSIO ZH16 (Sun *et al*., 2019). However, genome mining alone does not paint the full picture of a strain's biosynthetic potential.

Genome mining shows the potential a microorganism has to biosynthesise metabolites, but that does not equate to what chemistry is actively being produced. This is achieved using metabolomics analysis. A recent study by Martinez *et al*. utilised LC-MS/MS to identify a family of luminacins, a class of antiangiogenic antibiotic, from *Streptomyces* 39 PL. A fragmentation pattern with a loss of 172 Da, representing a glycoside fragment, was conserved within luminacins. This was used as the basis for the search which resulted in the isolation of five luminacins, with one being a novel metabolite (Martinez, Mello, Zucchi, Melo, & Moraes, 2020). LC-MS/MS analysis was also used to identify the presence of five macrocyclic antibiotics with previously reported antimicrobials from the potentially novel strain *Streptomyces* sp. M54, isolated from the eusocial wasp *Polybia plebeja* (Matarrita-Carranza *et al*., 2021). A major bottleneck throughout mass spectrometry-guided natural product discovery has been rediscovery of known metabolites and the struggle to differentiate known from unknown, termed dereplication (Mohimani *et al*., 2018). The Global Natural Product Social (GNPS) molecular networking infrastructure is an open access knowledge base where raw or processed tandem mass spectrometry data can be shared (M. Wang *et al*., 2016).

Using GNPS molecular networking, data stemming from LC-MS/MS analysis can be rapidly dereplicated and unknown parent ions can be targeted for further analysis.

Whilst advances in the outlined genome mining techniques and metabolomic analyses have undoubtedly propelled natural product discovery forward, the last class of antibiotics to be clinically deployed were discovered in 1987, which is a clear sign of stagnation in the field (M. I. Hutchings, A. W. Truman, & B. Wilkinson, 2019). Many Actinomycetota are categorised as 'rare' due to their low isolation frequency, and as most classical methods of drug discovery were modelled on lab-based culture, this highlights the tremendous amount of chemical novelty still to be unlocked (Amin, Abdallah, Abolmaaty, Tolba, & Wellington, 2020). Metabologenomics is emerging as an exciting drug discovery prospect. The field is based on the large-scale linking of BGCs to produced metabolites, yielding the benefits of both datasets, and foregoing the limited information provided by each discipline in isolation. For example, metabologenomics approaches have recently been used to target metabolites from Actinomycetota such as *Streptomyces* GMR22 for the treatment of SARS-CoV-2 (Melinda *et al*., 2021). The software framework NP Linker was developed for linking complex genomic and metabolomic datasets. Current methods, such as mass spectrometry-guided peptidogenomics (Kersten *et al*., 2011), lack effectiveness in automatically validating identified links, and manually linking datasets is prohibitive to the point of being considered a bottleneck in the research of natural products. Using input data gathered from a combination of external data repositories and used generated data, the platform applies multiple scoring methods including a novel Input-Output Kernel Regression approach for more accurate and comprehensive links (Hjörleifsson Eldjárn *et al*., 2021). For the purposes of generating training data for machine learning platforms such as NP Linker, large-scale datasets are required for computational platforms to make predictions regarding unseen data (Bassel, Glaab, Marquez, Holdsworth, & Bacardit, 2011).

The One Strain/Many Compounds (OSMAC) approach (Schiewe & Zeeck, 1999), which uses alternative culture conditions to activate metabolic pathways, can be used to elicit additional chemistry that would not be produced under standard laboratory conditions. Some strategies include coculture, expanding the culture medium used and adding chemical elicitors or enzyme inhibitors to culture media (Wei, Lin, Li, Gu, & Zhu, 2010). To generate a large dataset for high-throughput analysis, scaled-down cultures utilising chemical elicitors to increase produced chemistry must be explored. Metabolites are then extracted from the culture media using solvent with the extract then analysed via LC-MS-MS.

The work encompassed in this chapter began with two specific hypotheses – the first being that the NP Linker platform could be further validated using previously derived data comparing the chemical output of Actinomycetota strains with knockout mutants. The NP Linker framework was applied to -omics data derived from *Streptomyces coelicolor* M145 and its knockout strains M1145, M1152 and M1154 to validate the presence or absence of links between BGCs and metabolites. Eight further strains were analysed for links. Secondly, it was hypothesised that an assay could be developed that miniaturised the process of gathering crude metabolite extract for the generation of high-quality datasets for LC-MS/MS analysis to be linked with the output of genome mining by NP Linker. Considering large-scale datasets required for accurate metabologenomics analyses, a high-throughput culture assay was developed for this purpose. A miniaturised assay utilising deep-well culture plates in place of traditional Erlenmeyer flasks was tested. A minimal media with chemical elicitors was developed to generate the minimum 1 mg/mL extract weight generally required for LC-MS/MS analysis.

3.2 Results

3.2.1 MLST-based phylogenetic analysis enabled the selection of phylogenetically diverse Actinomycetota

Six *Streptomyces* and three non-*Streptomyces* Actinomycetota were selected for analysis due to whole genome sequence availability and published antibiotic activity. A further 23 fully sequenced strains were selected for phylogenetic comparison and the maximum likelihood phylogenetic tree of 32 Actinomycetota strains was organised into two distinct clades (**Figure 3.1**). The non-*Streptomyces* strains included in the study formed their own, well supported clade (bootstrap value = 100%). *Micromonospora echinospora* and *Micromonospora chalcea* formed a subclade, indicating that they share a more recent common ancestor than either strain shares with *Nocardia farcinica*. The second major clade consisted of all but one *Streptomyces* and itself was organised into two large subclades. The first subclade was well supported (100%) and contained the query strains *Streptomyces collinus* Tu 365 and *Streptomyces scabiei* 87.22. Interestingly, *S. scabiei* 87.22 formed a tight subclade, supported by a bootstrap value of 100%, with *Streptomyces bottropensis* ATCC 25435, a strain which was recently found to also cause potato common scab. The second subclade was less well supported at 77% but still within the threshold for good confidence, considered to be 70% (Hillis & Bull, 1993). This clade contained the query strains *Streptomyces clavuligerus* ATCC 27064, *Streptomyces venezuelae* ATCC 10712, *Streptomyces coelicolor* M145, *Streptomyces lividans* TK23 and *Streptomyces lividans* 1326. The strains *S. coelicolor* and *S. lividans* are known to be close phylogenetically and indeed formed a tight subclade with a recent common ancestor.
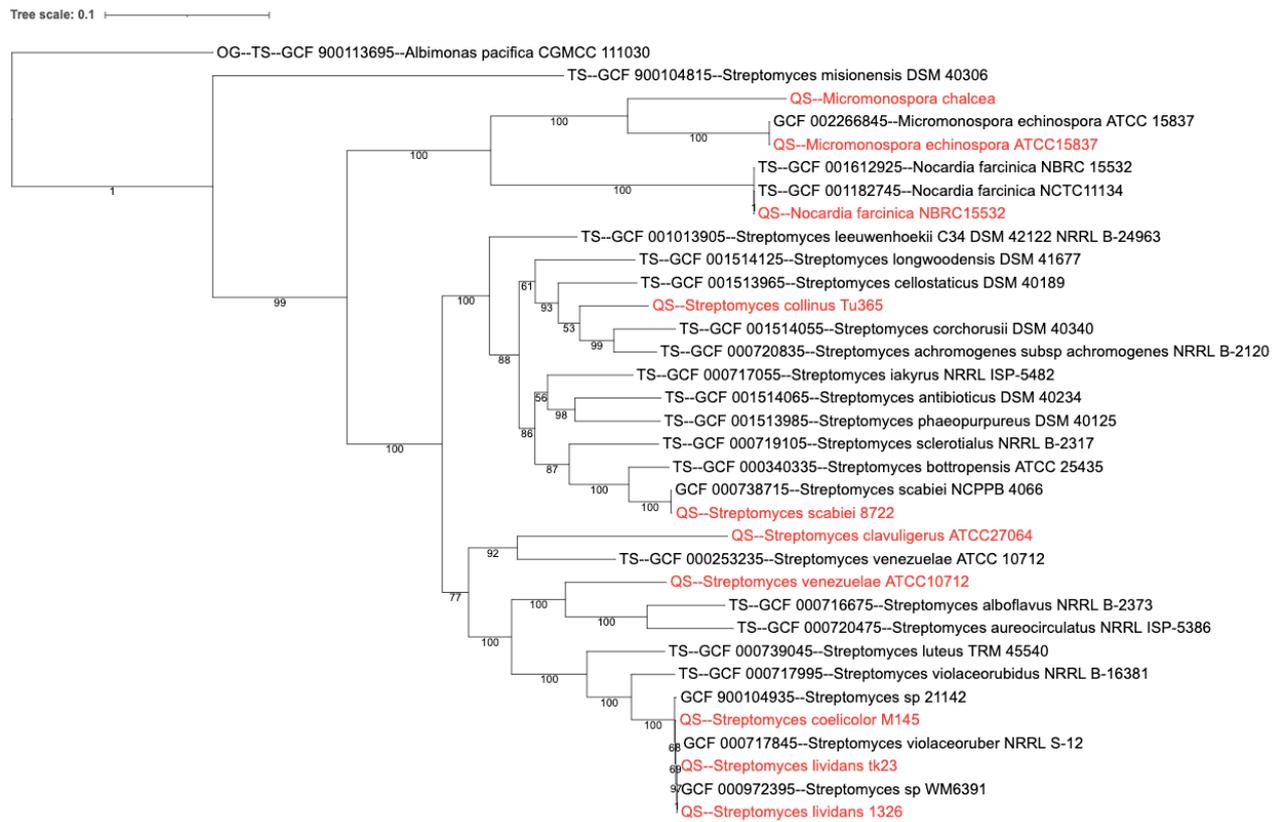
Figure 3.1 - Maximum likelihood tree based on whole genome sequences for nine Actinomycetota strains utilised for the study. Full sequences were obtained from the National Centre for Biotechnology Information (NCBI) with accession numbers as follows: *M. echinospora* - LT607413.1; *M. chalcea* - MAGP01000001.1; *N. farcinica* - CP031418.1; *S. collinus* – CP006259.1; *S. lividans* 1326 – CM001889.1; *S. coelicolor* – AL645882.2; *S. venezuelae* – CP029192.1; *S. clavuligerus* CP027858.1; *S. scabiei* 87.22 – FN554889.1. The tree was constructed with AutoMLST (1000 replicates, IQ-TREE Ultrafast Bootstrap analysis, denovo construction mode) and visualised using the interactive Tree Of Life (iTOL). Query strains selected for use in the study are labelled in red, and *Albimonas pacifica* CGMCC 111030 was used as an outgroup.

3.2.2 *Streptomyces* sp. have higher biosynthetic potential than other analysed genera

Genome mining was carried out on the nine Actinomycetota strains using antiSMASH v6.0.1 (K. Blin *et al*., 2021), which revealed a total of 240 BGCs at an average of 27 per strain. The strain with the highest number of BGCs was *S. scabiei* 87.22 with 34 in total, seven higher than the predicted average. This is perhaps to be expected as it has the largest genome within the set at 10.15 Mbp. Conversely, *N. farcinica* had both the smallest genome (6.43 Mbp) and the fewest BGCs (21) (**Supplementary Table 3.1**).

The most common class of BGC detected by antiSMASH v6.0.1 was Non-Ribosomal Peptide Synthases (NRPS), of which there were 42 (**Figure 3.2**). This falls in line with general trends as NRPs are a diverse group of specialised metabolites encompassing many clinically relevant antimicrobials. Perhaps the other most well characterised class of BGC are the Polyketide Synthases (PKS), which are subdivided into Type 1, Type 2, Type 3. A further PKS group, 'PKS Other', appears within the analysis. This group encompasses PKS-like clusters which cannot be categorised with Types 1, 2 and 3 or may merely have been misidentified.

A group of 35 BGCs were categorised together as 'Others'. This group contains the more uncommon classes of BGC found in Actinomycetota genomes, but also classes such as lassopeptides and alkaloids. BGCs that may be grouped into this class are often the source of medically important bioactive specialised metabolites so their grouping together should not lead to an assumption that they are not of interest. There were low numbers of Ribosomally synthesised and Post-translationally modified Peptides (RiPP) identified by antiSMASH. However, this class is notoriously difficult to identify given the post-translational nature of the final product with genes often found out with the primary cluster, so it is likely that more are present.

Upon GCF clustering by BiGSCAPE (Navarro-Munoz *et al*., 2020) it was found that the hopene cluster was conserved within all analysed *Streptomyces* strains as its hopanoid products are analogous to eukaryotic sterols in reducing lipid membrane permeability. One siderophore GCF contained BGCs across the *S. scabiei, S. collinus* and *S. clavuligerus* genomes had no similarity to any BGC within the MIBiG database (Kautsar *et al*., 2020), presenting itself as a priority target for drug discovery. Of the 240 BGCs predicted, 44 shared no sequence homology and 196 BGCs shared 1% - 100% sequence homology with BGCs submitted to MIBiG (**Supplementary Table 3.2**), meaning in total 18.3% BGCs lacked similarity to experimentally validated BGCs within the MIBiG data repository, underlining the high potential for chemical novelty within the dataset.
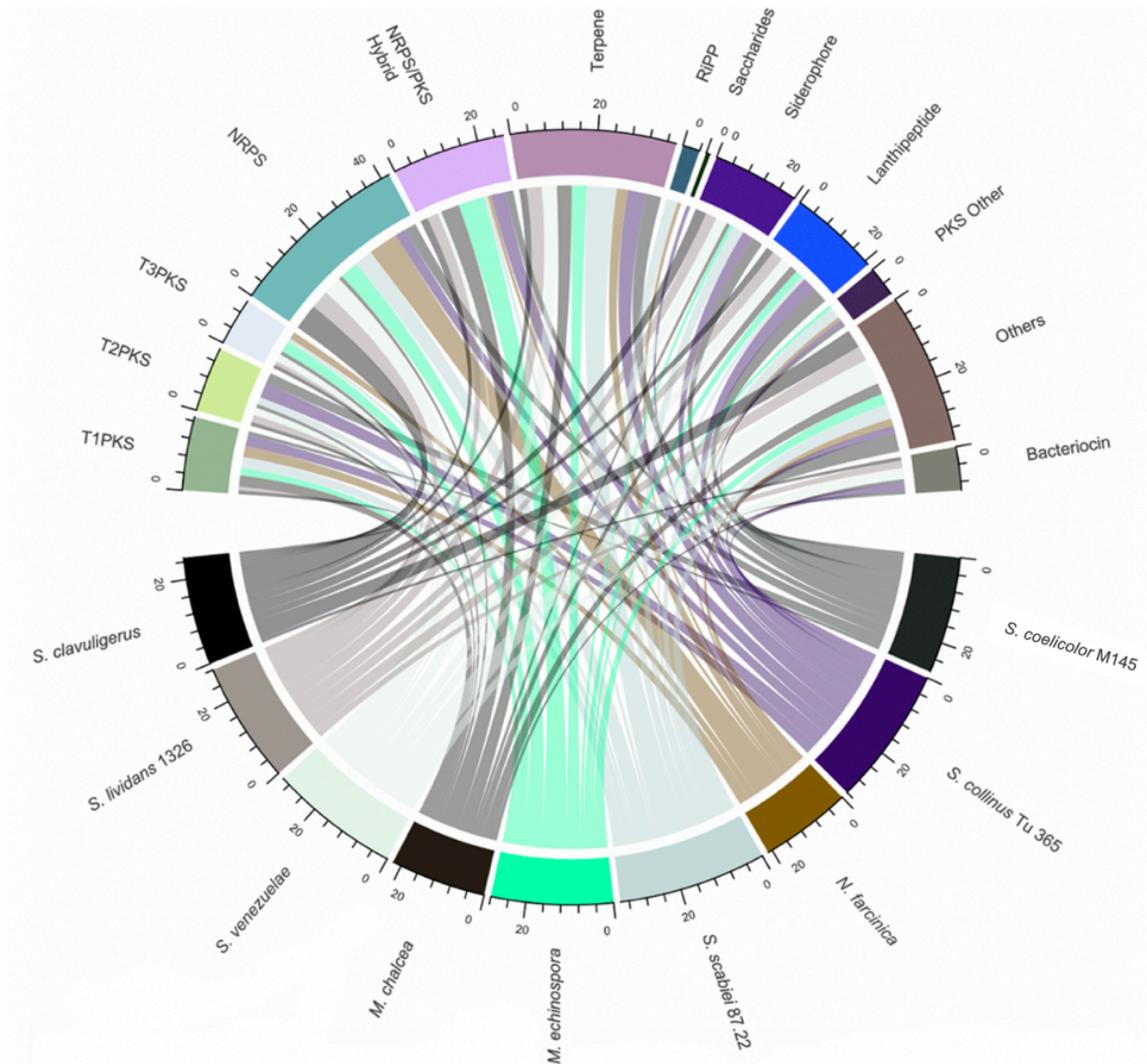
Figure 3.2 – Predicted BGC classes within Actinomycetota genomes. Ribbon plot showing predicted BGCs of the nine Actinomycetota strains (bottom of plot) selected for analysis, as predicted by antiSMASH v6.0.1 (K. Blin *et al*., 2021). The strains were selected based on availability of genome sequence and known antimicrobial activity. Each bar on the scale = 4. There was a total of 240 BGCs predicted across the pangenome, split into 12 specialised metabolite classes plus 'Other' (top of plot), which consists of marginal classes such as lassopeptides and alkaloids.

3.2.3 Comparative metabolomics revealed considerable chemical diversity of Actinomycetota

A GNPS molecular network of 145 microbial metabolite extracts (15 strains of Actinomycetota cultured in duplicate across five media conditions) was constructed, revealing 3622 nodes and 5944 edges (**Figure 3.3 a**). 418 nodes were detected exclusively within solvent and media blank controls. There were 278 Molecular Families (MFs) detected (cosine $\geq$ 0.7), ranging from having two to as many as 98 nodes in the largest family, indicating that these metabolites are structurally related. 1110 nodes (1337 in total, minus 227 solvent/media blank controls) were defined as singletons, in that they did not correlate with any other parent ions within the dataset, underlining the level of structural diversity within the detected metabolites and tying in with the fact that *Streptomyces* produce diverse chemistry. Chemical production was also dependant on culture media (data not shown), with culture on ISP2 medium resulting in the most nodes detected (945). Despite this high level of diversity, there remains a significant shared chemical space between Actinomycetota strains - the largest MF, made up of 98 nodes, consisted of spectra from all analysed strains and only one MF consisted of spectra relating exclusively to one strain. Within the network, spectra of a similar molecular mass to undecylprodigiosin (393.575 g/mol^{-1}) were detected within nodes related to the metabolite extracts of *S. coelicolor* M145 (393.279 Da *m/z*, 394.28601 Da precursor *m/z*, spectrum ID 1346) (**Figure 3.3 b**), and was not present within data relating to any of the three *S. coelicolor* deletion mutants. Similar results were observed for two other metabolites that had their BGC knocked out within the *S. coelicolor* deletion mutants. Actinorhodin (632.496 Da *m/z*, 633.504 Da precursor *m/z*, spectrum ID 2063) (**Figure 3.3 c**) and coelimycin P1 (348.173 Da *m/z*, 349.181 Da precursor *m/z*, spectrum ID 1198) (**Figure 3.3 d**) were detected within *S. coelicolor* metabolite extracts but were again absent from data from to the deletion mutants. These data indicated that all three metabolites were produced by *S. coelicolor* M145 in the examined culture extracts, allowing targeted comparison and linking via computational pattern matching.

a)

b)

c)

d)

KEY:

- S. coelicolor M145
- S. coelicolor M1146
- S. coelicolor M1152
- S. coelicolor M1154
- S. coelicolor CH999
- S. collinus
- S. scabiei 87.22
- S. venezuelae
- S. clavuligerus

- S. lividans 1326
- S. lividans TK24
- S. goldiniensis
- M. echinospora
- M. chalcea
- N. farcinica
- Blank
- Shared

- 3622 total nodes

84

Figure 3.3 – GNPS molecular network of produced chemistry from 15 Actinomycetota strains across five media conditions. a) GNPS molecular network showing 3622 parent ion nodes produced by 15 Actinomycetota strains. Grey nodes represent media components and pale blue nodes represent parent ions produced by two or more strains. **b), c)** and **d)** inset metabolites highlighted in red. Coloured nodes are produced exclusively by one strain, as per key. **b), c)** and **d)** show the nodes manually linked to the metabolites undecylprodigiosin, actinorhodin and coelimycin P1 with the metabolite chemical structures.

3.2.4 – Computational pattern matching revealed links between metabolites and BGCs

The four BGC pathways that had been knocked out in the *S. coelicolor* deletion mutants (Δred, Δact, Δcda, Δcpk) were used as a starting point for computational pattern matching of known chemistry, utilising *S. coelicolor* M145 wild type as a comparison. The scoring method is herein referred to as Metcalf. Using NP Linker, the link between *S. coelicolor* M145 region 31 (100% sequence homology to MIBiG ID BGC0001063, undecylprodigiosin) and spectrum ID 1346 (ISP2 media, 393.279 Da *m/z*, 394.28601 Da precursor *m/z*, spectrum ID 1346) had the highest Metcalf score within the ranked links (2.8284, where the average score was 1) (**Table 3.1**). The same method of identifying mass spectrometry features similar to the expected accurate masss of the deleted metabolites was employed to link the predicted BGCs of actinorhodin (MIBiG ID BGC0000194) (**Table 3.1**) and coelimycin P1 (MIBiG ID BGC0000038) (**Table 3.1**) to parent ions produced by *S. coelicolor* M145 in ISP4 (632.497 Da *m/z*, spectrum ID 2063 and 348.173 Da *m/z*, spectrum ID 1198 respectively). These links go some way to solidifying NP Linkers position as an exciting prospect for metabologenomics-based drug discovery, as it shows the ability of the platform to identify links between known BGCs and spectra. It should be noted that despite the high Metcalf scores, the indicted parent ions for the three liked spectra mentioned above do not match the exact mass within the required 5 ppm accuracy for high resolution data (the expected parent ion *m/z* for undecylprodigiosin should be 394.253 and not the observed 393.278, for actinorhodin the expected parent ion *m/z* is 635.1395 compared to the observed 632.496 and finally the expected parent ion *m/z* for coelimycin P1 is 349.1217 and not the observed 349.181). These data represent potential false positives within the ranked links yielded from NP Linker analysis.

Each BGC within the dataset was then mined for links. A link was investigated if the spectra ranked with the highest Metcalf score had an *m/z* close to the molecular mass of the BGC product after selecting each BGC within the NP Linker output. The fragmentation patterns were then observed. Ten further specialised metabolites were linked to their BGCs using this

method (**Table 3.1**): albaflavenone, melanin, nocobactin NA-b, coelichelin, naringenin, kirromycin, holomycin, desferroxamine B, geosmin and germicidin A/B. These links lack the experimental validation of the deletion mutants, but the links represent a concrete point from which to begin experimentally verifying the links between these BGCs and spectra.

Table 3.1 – Specialised metabolites linked to BGCs by NP Linker. Table of specialised metabolites from LC-MS/MS analysis of nine Actinomycetota strains linked to BGCs predicted by antiSMASH. Linking was undertaken via the NP Linker platform using computational pattern matching and a standardised scoring system known as 'Metcalf'. The table includes the producer strains, media and detected parent ions. The numerical value of the Metcalf score is arbitrary and varies between datasets. The highest scores within the datasets are the strongest inferred links. In this dataset, the highest score was 2.8284, and the average was 1.

Identified Metabolite	Strain(s)	Media	Parent ion *m/z* (Da)	Metcalf score
Albaflavenone	*S. lividans* 1326 *S. lividans* TK24 *S. goldiniensis* *M. chalcea*	ISP1 ISP2 ISP3	261.129 +(M+ACN+H) adduct	2.2361
Melanin	*S. clavuligerus*	ISP1	319.239	2.8284
Undecylprodigiosin	*S. coelicolor* M145	ISP2	393.278	2.8284
Nocobactin NA-b	*N. farcinica*	ISP2 ISP4	772.448	2.8284
Actinorhodin	*S. coelicolor* M145	ISP4	632.496	1.8708
Coelichelin	*S. lividans* 1326	ISP3	606.440 +(M+ACN+H) adduct	1.8708
Naringenin	*S. clavuligerus*	ISP1	273.078	2.8284
Coelimycin P1	*S. coelicolor* M145	ISP4	349.181	2.8284
Kirromycin	*S. collinus* *S. lividans* TK24	ISP2 ISP3 ISP4 ISP5	797.422	2.8284
Holomycin	*S. clavuligerus*	ISP1	256.229 + (M+ACN+H) adduct	2.8284
Desferroxamine B	*S. coelicolor* M145	ISP2	559.510	2.8284

	S. lividans 1326	ISP5		
Geosmin	*M. echinospora*	ISP1	183.092	2.1381
	S. coelicolor M145	ISP2		
	S. lividans 1326			
Germicidin A/B	*S. coelicolor* M145	ISP2 ISP4	183.103	2.8284
		ISP3 ISP5		

3.2.5 Strain selection for the development of a miniaturised culture method

In this next hypothesis, a miniaturised culture and extraction method was developed using strains unrelated to the earlier work in this chapter. Ten strains of *Streptomyces* were selected for inclusion within the study based on factors such as known chemistry, antimicrobial activity and genome size and quality (**Table 3.2**). The genomes of all ten strains were analysed for BGCs as a proxy of biosynthetic potential.

Table 3.2 - Strains selected for inclusion within the study with NCBI accession numbers, genome coverage and genome size. The strains were again selected based on the availability of genome sequence and known antimicrobial activity. Another factor considered was that there should be no overlap with previously analysed strains.

Strain	NCBI Accession No.	Genome Coverage	Genome Size (Mbp)
Streptomyces noursei ATCC 11455	CP070326	226.3x	9.8
Streptomyces rimosus ATCC 10970	CP025551	1500x	9.37
Streptomyces clavuligerus ATCC 27064	CP027858	60x	7.59
Streptomyces fradiae ATCC 10745	CP023696	120x	8.08
Streptomyces venezuelae ATCC 10712	CP029197	100x	9.05
Streptomyces kanamyceticus ATCC 12853	CP023699	50x	10.13
Streptomyces nodosus ATCC 14899	CP023747	50x	7.77
Streptomyces spectabilis ATCC 27465	CP023690	50x	9.81
Streptomyces vinaceus ATCC 27476	CP023692	50x	7.67
Streptomyces platensis ATCC 23948	CP023691	50x	8.50

3.2.6 Phylogenetic analysis of ten whole genome sequenced *Streptomyces* sp. yielded eight distinct clades

AutoMLST (Alanjary *et al*., 2019) analysis of the selected sequenced *Streptomyces* strains shows phylogenetic diversity as evidenced by clades and subclades (**Figure 3.4**). The tree is divided into eight clades, each supported by a bootstrap value of 100%. 50 strains comprising of 40 species are included by autoMLST with *Microtetraspora glauca* NRRL B-3735 selected as an outgroup. *S. rimosus* ATCC 10970, *S. noursei* ATCC 11455 and *S. platensis* ATCC 23948 are the only query strains involved in the study that appear in the same clade, indicating a more recent common ancestor. Given one of the primary study aims is to maximise produced chemistry by using coculture to elicit the production of potentially novel specialised metabolites, Phylogenetically diverse strains were selected, allowing for analysis of a broad range of BGCs.
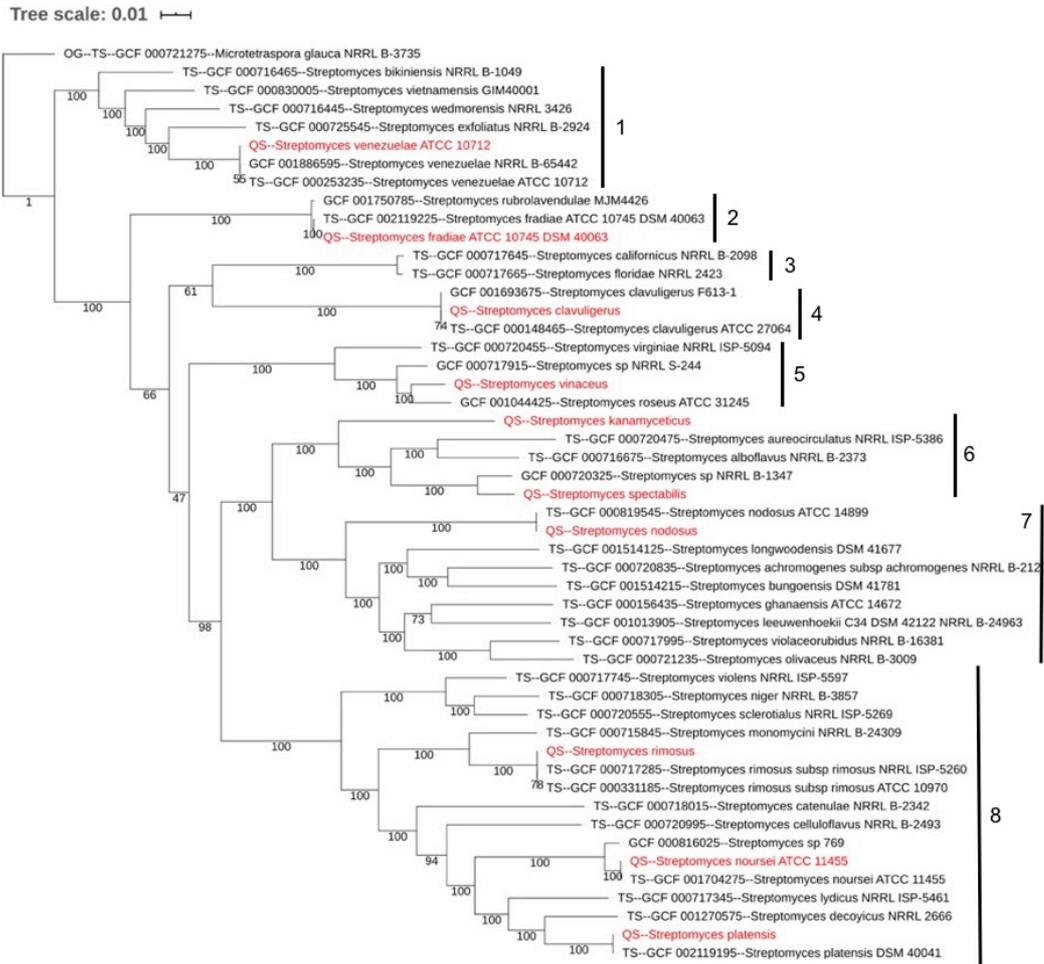
Figure 3.4 – Maximum likelihood phylogenetic tree of selected *Streptomyces* strains. The tree was constructed with AutoMLST (1000 replicates, IQ-TREE Ultrafast Bootstrap analysis, *denovo* construction mode) and visualised using the interactive Tree Of Life (iTOL). Bootstrap values displayed at nodes. Query strains are shown in red. Clades are indicated by black bar and numbered in descending order. Genomes retrieved using the following NCBI accession numbers: *S. noursei* ATCC 11455 – CP011533.1, *S. rimosus* ATCC 10970 – CP025551.1, *S. clavuligerus* ATCC 27064 - CP027858.1, *S. fradiae* ATCC 10745 – CP023696.1, *S. venezuelae* ATCC 10712 – CP029197.1, *S. kanamyceticus* ATCC 12853 – CP023699.1, *S. nodosus* ATCC 14899 – CP023747.1, *S. spectabilis* ATCC 27465 – CP023690.1, *S. vinaceus* ATCC 27476 – CP023692.1, *S. platensis* ATCC 23948 – CP023691.1.

3.2.7 Genome mining of ten *Streptomyces* strains revealed 340 predicted BGCs

BGC profiles were predicted for all ten strains by antiSMASH v6.0.1 (K. Blin *et al.*, 2021) **(Supplementary Table Set 3.3)**. A total of 340 BGCs were predicted across the ten strain pangenome, with antiSMASH v6.0.1 able to predict the identity of 289 based on comparison to the most similar known clusters within the MIBiG database (Kautsar *et al.*, 2020). A prediction of specialised metabolite class was made for the remaining 51 BGCs. In total, 15% of all BGCs within the pangenome shared no homology to any known BGC within MIBiG, highlighting potential novelty within the dataset.

The number of BGCs per strain ranged from 22 (*S. nodosus* ATCC 14899) to 46 (*S. spectabilis* ATCC 27465) (**Figure 3.5**). The most abundant specialised metabolite class throughout the ten strains was NRP with 67 NRPS BGCs occurring throughout the selected strain genomes, followed by terpenes with 47. There were also 58 total PKS BGCs, subdivided into Type 1, Type 2 and Type 3, as well as 28 NRPS/PKS hybrid clusters. These numbers line up with commonly accepted opinion that most antimicrobials can be categorized as NRPs or PKSs. There were 58 BGCs designated as 'Others'. This group encompassed less abundant classes of BGC within the dataset including alkaloids, arylpolyenes and hglK-ES (**Figure 3.5**). The average NRPS content throughout the 10 strains was 21.6%, *S. rimosus* ATCC 10970 had higher than average NRPS content with 31%. Average PKS genome content was 20.2%. *S. spectabilis* ATCC 27465 and *S. platensis* ATCC 23948 displayed higher than average PKS content with 32% and 28.6% respectively. Saccharide genome content was generally low, however *S. fradiae* ATCC 10745 displayed a higher percentage of saccharide at 10%, 4.8% higher than the average (5.2%).

Streptomyces are one of few bacterial genera which possess linear chromosomes. These linear chromosomes are generally known to be compartmentalised, consisting of a central region, where common core genes are located, and terminal arms susceptible to DNA

recombination and rearrangements (Lioy *et al*., 2021). Analysis of spatial distribution of BGCs within genomes using normalised positioning revealed this to be the case for each strain within the dataset (**Figure 3.6**). Most strains had an even distribution of BGCs between the left and right terminal arms, with notable exceptions being *S. fradiae* ATCC 10745 and *S. vinaceus* ATCC 27476 which skewed heavily towards the right and left arms respectively. This indicates that these genomic regions have undergone a higher level of recombination. In general, the arms of the *Streptomyces* linear genomes are more prone to recombination events, with more conserved genes found towards the OriC region, and the results observed tie into that theory with predicted BGCs gathered towards the arms.

The data were then analysed using BiGSCAPE-CORASON (Navarro-Munoz *et al*., 2020) to cluster BGCs into GCFs before visualization via Cytoscape (Shannon *et al*., 2003) (**Figure 3.7**). This resulted in ten GCFs including hopene, ficellomycin and ectoine. Three GCFs corresponded to BGCs with no match in the MIBiG database and therefore presented themselves as a good starting point for computational pattern matching of the genomic and metabolomic data from this analysis.
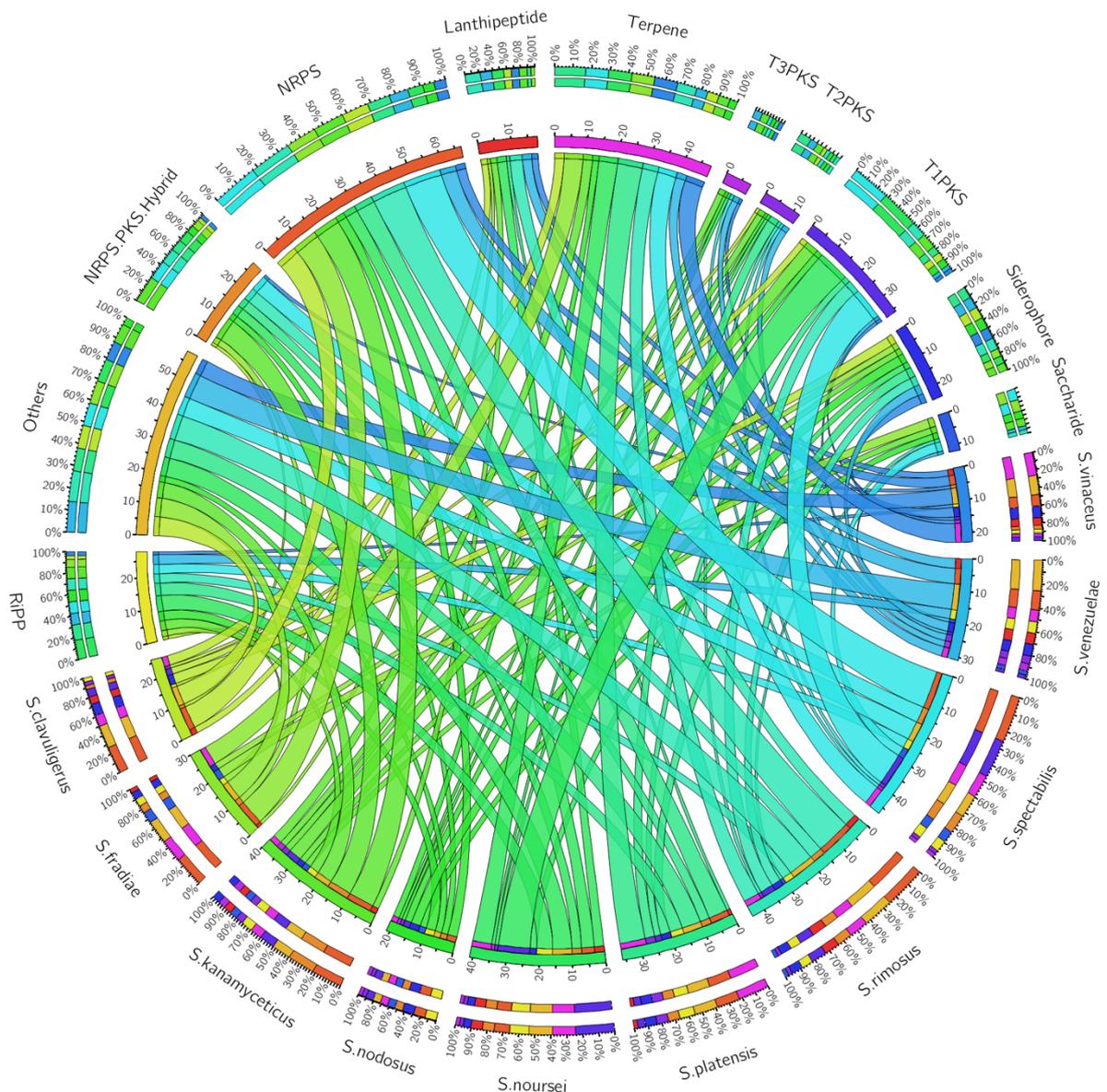
Figure 3.5 – Ribbon plot comparing BGC content and class within the analysed pangenome. Ribbon plot created using Circos Online (http://mkweb.bcgsc.ca/tableviewer/visualize/) showing percentages of each class of BGC (top of plot) within each analysed *Streptomyces* strain (bottom of plot) as predicted by antiSMASH. In total, there were 340 BGCs detected within the analysed pangenome, averaging 34 per strain. The ten strains were selected based on genome availability, known antimicrobial activity and lack of overlap with previously analysed strains. For this plot, each bar = 5. The 'Others' heading consists of marginal BGC classes such as bacteriocins and alkaloids.
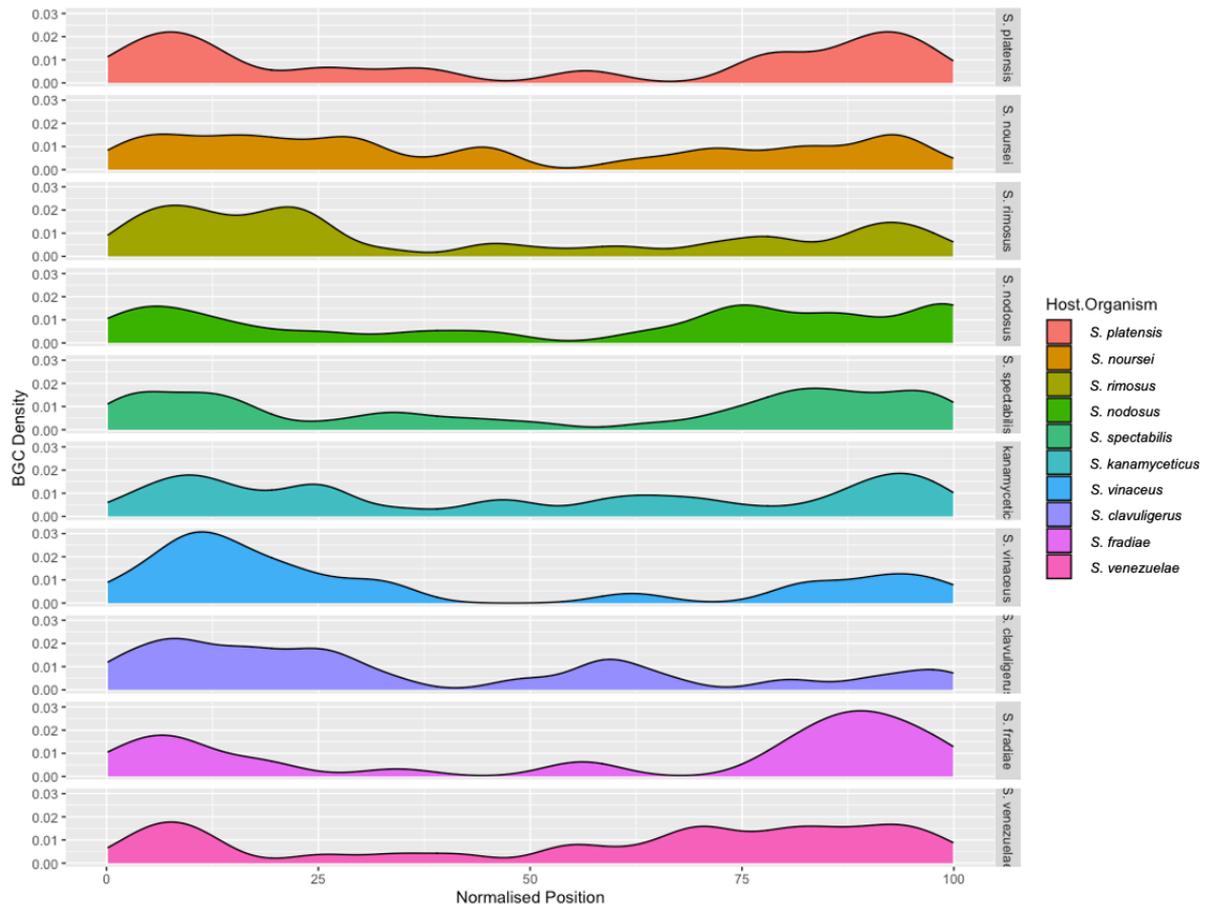
Figure 3.6 – Normalised distribution of BGC position along the genomes of the ten analysed strains. In the normalised position, the OriC is located in the middle of each strain plot. BGCs are evenly distributed between terminal arms on the linear chromosome, except for *S. fradiae* which heavily skews to the right, and *S. vinaceus* which heavily skews to the left. In every strain, BGCs were mostly located towards the chromosomal arms. These regions are more prone to recombination events whereas the region close to the OriC is conserved.
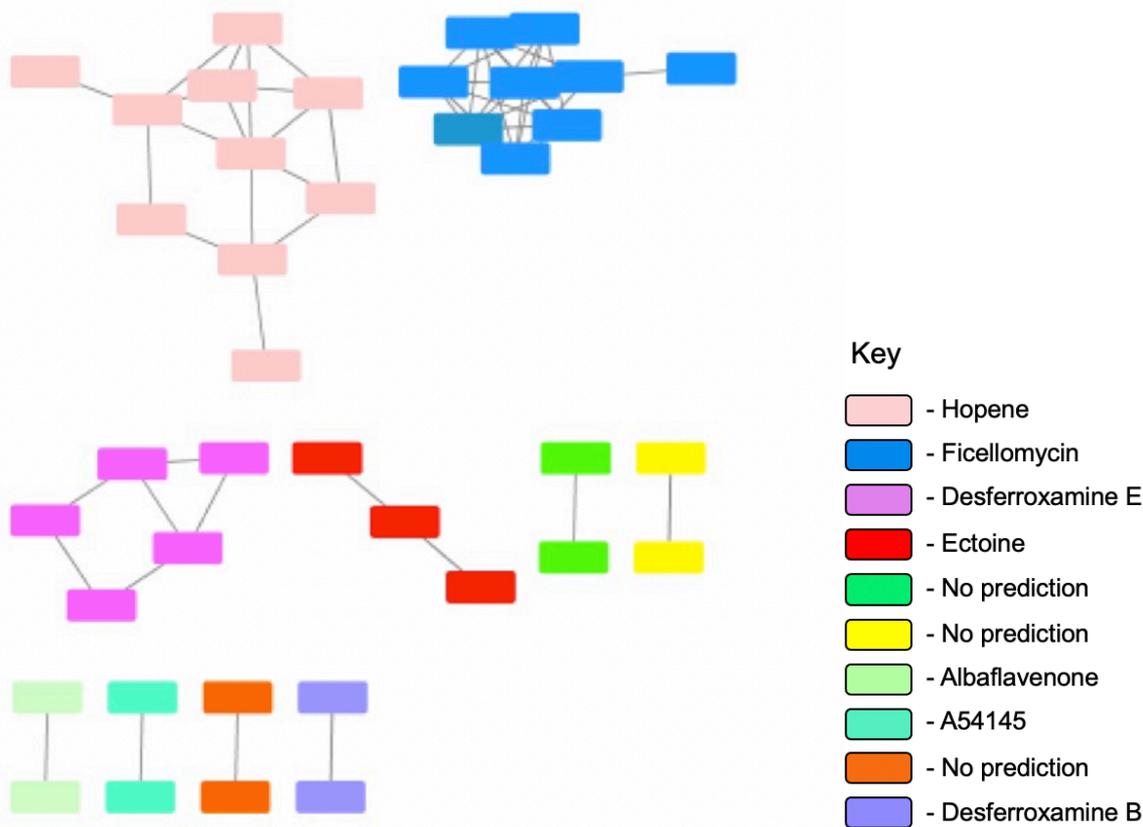
Figure 3.7 – Network showing GCFs clustered by BiGSCAPE-CORASON within the dataset. This similarity network shows ten GCFs clustered by BiGSCAPE-CORASON. Seven GCFs have predictions of identity in antiSMASH. Three of the GCFs were families of terpenes (green), PKS's (yellow) and RiPPs (orange) with no identified most similar known BGC as predicted by antiSMASH. As PKS's account for many candidate antimicrobials and RiPP products are generally bioactive, these GCFs with no predicted identity represent a good starting point for targeted linking in the search for novel antimicrobial compounds.

3.2.8 Miniaturised culture and metabolite extraction yielded over 1 mg/mL dry mass

The diversity in growth and sporulation of strains is dependent on the type of culture media. Strains were cultured on several media to gain an insight into behaviours and determine which media would be best suited for the assays. *S. fradiae* ATCC 10745 showed minimal growth and no sporulation on GYM media. Culture of the same strain on ISP2 also led to a similar lack of growth with sporulation occurring on around a third of the culture, whilst culturing on SFM led to increased growth and universal sporulation (**Figure 37 a**). This is in direct contrast to what was observed when assessing cultures of *S. spectabilis* ATCC 27465. When grown on GYM, a red pigment was expressed combined with sporulation of the whole culture whilst reduced growth was observed when grown on SFM and a lack of red pigment was observed on ISP2 (**Figure 3.8 b**).

ISP2 was selected for initial testing because it yielded the most parent ions in previous work in this study. Both solid and liquid media, as well as culture scale were investigated to optimize a higher throughput of metabolite extracts generated. The metabolite crude extracts generated confirmed that an extract weight higher than 1 mg/mL could be obtained on both solid and in liquid media on 24-well plates (Axygen™). 1 mg/mL is generally the concentration of extract required for LC-MS/MS protocols, so the reduction in scale whilst generating this amount of extract allows for a higher throughput method (**Supplementary Table 3.4**). A drawback with this version of the reduced scale method is the fact that just over 1 mg/mL is generated, effectively destroying the sample after the first test. A change in method was implemented to increase nutrient stress for elevated production of specialised metabolites and in turn, increased dry extract weight. Liquid GYM (DSMZ recommended media) and NMMP media replaced ISP2 due to improved growth of *Streptomyces.* For the carbon source in the media, n-acetylglucosamine (GlcNAc) was introduced due to its properties as a chemical elicitor. Contrarily, it has been shown that an excess of GlcNAc can repress specialised metabolite production, so increments of 0, 10 and 50 mM GlcNAc were used. The extraction method was

also altered, and a liquid/liquid technique was implemented to remove the freeze-drying step and reduce the overall time of the assay. This method heavily increased the extract dry weight produced from all microbial metabolite extracts compared to media blanks (**Supplementary Table 3.5**).
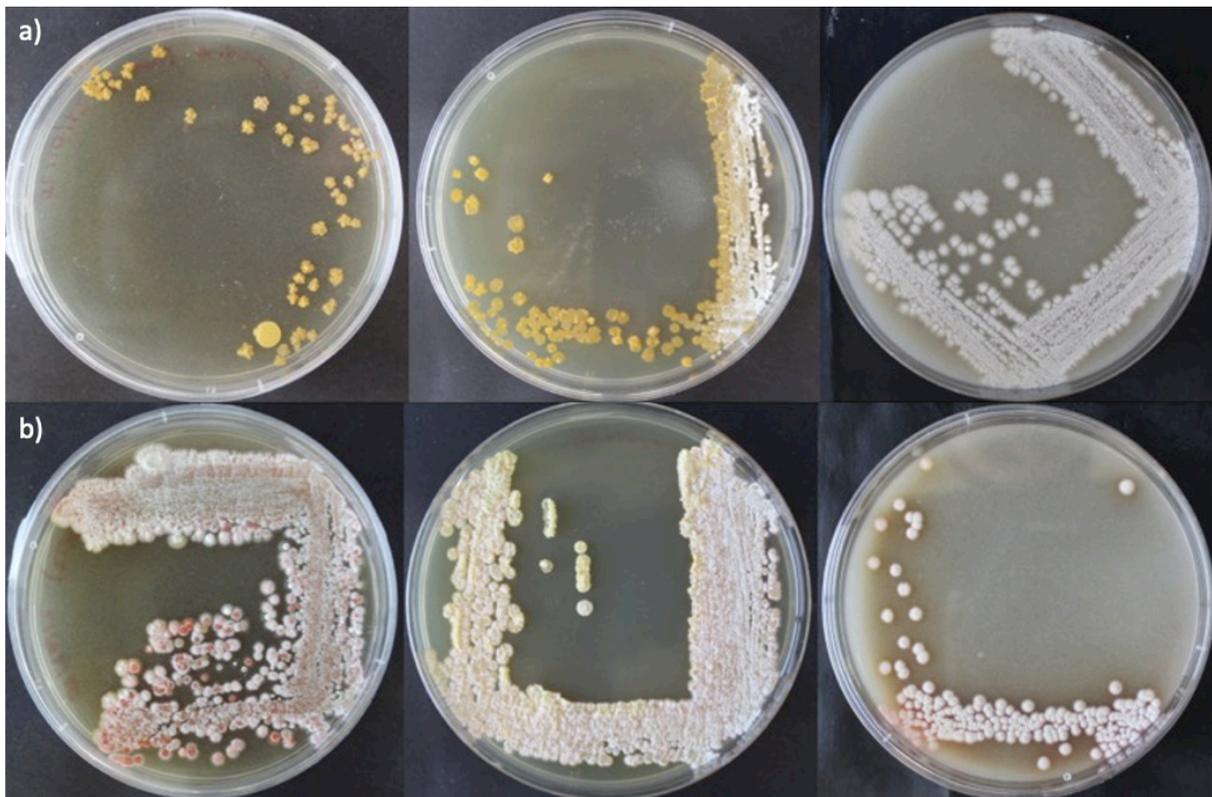
Figure 3.8 – Comparative growth of *Streptomyces* strains on diverse media. The strains were inoculated from spore stocks and incubated (5 days, 30°C) before being removed and analysed. a) Growth of *S. fradiae* ATCC 10745 on (l to r) GYM, ISP2 and SFM showing reduced growth and a lack of sporulation on GYM, partial sporulation and complete sporulation on SFM. b) shows growth of *S. spectabilis* ATCC 27465 on (l to r) GYM, ISP2 and SFM. In contrast to a), the highest level of growth and pigmentation was observed on GYM media, with pigment secretion into the media observed when cultured on MS.

3.2.9 Gravimetric dry mass analysis determined that the optimum extraction time point from reduced-scale culture was strain-dependent

Seven strains of *Streptomyces* were subjected to gravimetric dry mass analysis to determine when each strain transitioned to stationary phase (**Figure 3.9 a-g**) for the miniaturised assay in development. The data was also used to determine optimum extraction time, as stationary phase is the point in the *Streptomyces* life cycle where the majority of specialised metabolites are produced. Of the seven strains analysed, three – *S. spectabilis, S. kanamyceticus* and *S. nodosus* – were calculated to have reached stationary phase after 96 hours (**Figure 3.9 a-c**). The remaining four strains – *S. noursei, S. fradiae, S. vinaceus* and *S. platensis* – transitioned into stationary phase after 120 hrs (**Figure 3.9 d-g**). Going forward with the analysed strains, for the version of the assay conducted on Axygen™ 24-well plates, the optimum time for extraction would be 120 hours to allow all seven strains of *Streptomyces* to transition to stationary phase.
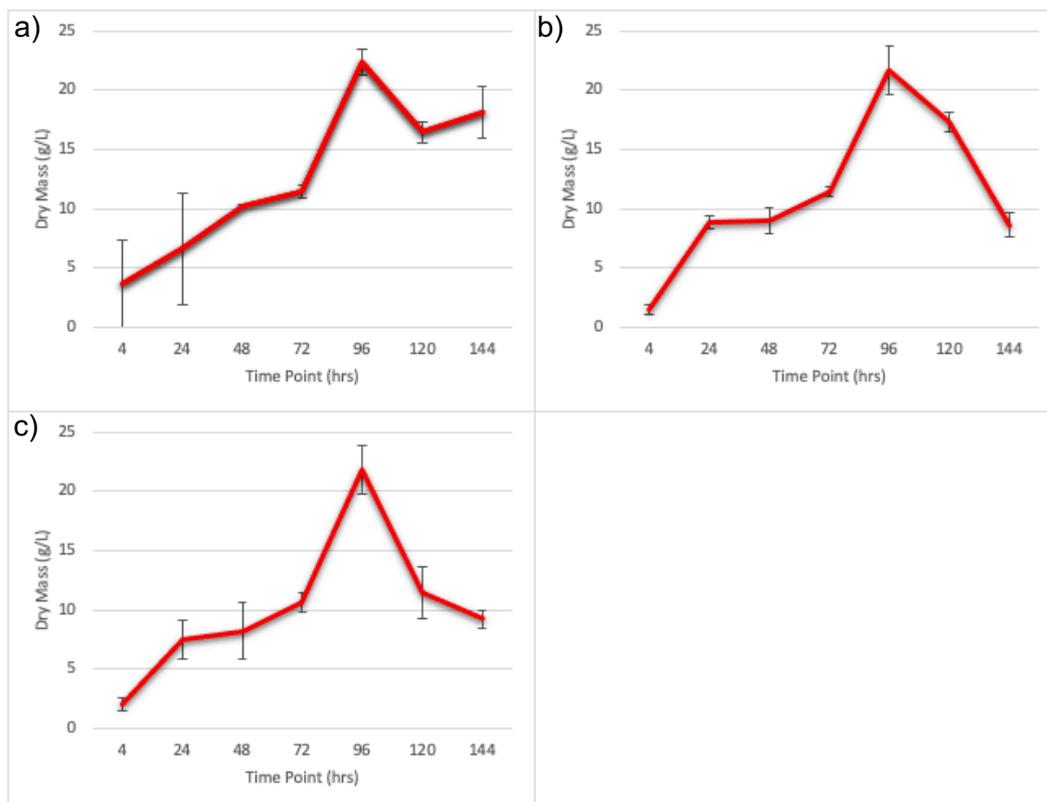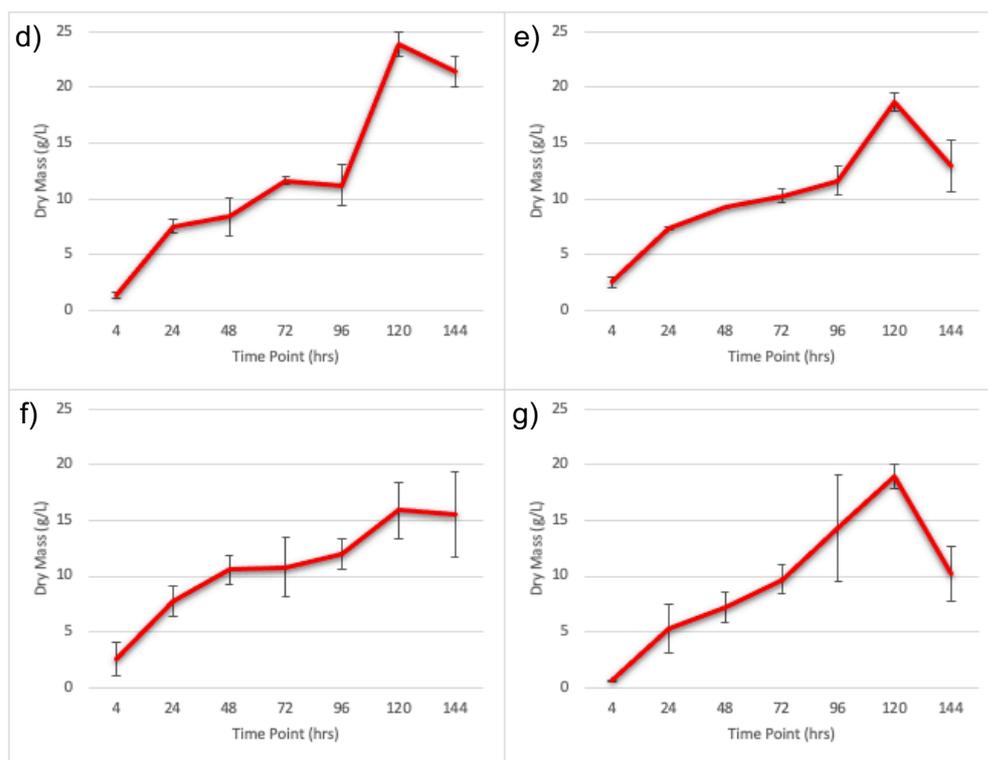
Figure 3.9 – Growth curves showing seven *Streptomyces* strains reaching stationary phase.
Growth curves of a) *S. spectabilis*, b) *S. kanamyceticus,* c) *S. nodosus,* d) *S. noursei,* e) *S. fradiae,* f) *S. vinaceus* and g) *S. platensis*. Each strain was cultured in triplicate in liquid GYM media. 1 mL aliquots were obtained from each culture every 24 hours for six days. The spores were then separated from the liquid media using gravimetric dry mass analysis to pull the media through pre-weighted filter paper, which was then dried and re-weighted. Three strains (a-c) reached stationary phase by 96 hours. The remaining four strains reached stationary phase by 120 hours. Error bars indicating standard deviation within triplicate cultures are shown.

3.3 Discussion

In this study, genome mining of Actinomycetota was carried out using antiSMASH to identify BGCs for linking and analysis and revealed extensive biosynthetic diversity. 240 BGCs, averaging 27 per genome, were revealed in the first set of nine Actinomycetota analysed. The second set of 10 *Streptomyces* had 340 BGCs predicted at 34 per genome. A recent comprehensive study on phylogeny and biosynthetic diversity of 1,110 *Streptomyces* genomes downloaded from NCBI revealed an average of 39.64 BGCs per genome (Belknap, Park, Barth, & Andam, 2020). Surprisingly, both sets of Actinomycetota analysed in this work had lower average BGCs per genome than the comprehensive study. However, the first set of nine Actinomycetota analysed contained strains from the genera *Nocardia* and *Micromonospora* which tend to have less BGCs per genome on average than *Streptomyces,* reducing the average from this set. It should also be noted that the referenced study utilised a much larger, and as a result presumably more diverse, dataset. If this work was expanded to include more genomes, the average number would potentially increase. Furthermore, the first set of nine Actinomycetota shared 44 BGCs with no sequence homology to any BGC in the MIBiG database, at an average of 5.5 per genome. A study comparing abundance of BGCs from 11 phylogenetically linked genomes showed an average of 50 BGCs per genome including 89 singletons, around 8 BGCs per genome, not associated with any known natural product (Chung *et al.*, 2021). Interestingly, despite the lower number of BGCs per genome in the Actinomycetota analysed in this work, the percentage of unknown BGCs is similar (16% and 18%). These unknown BGCs represent excellent future targets for metabologenomics analysis.

Of the initial four deleted BGCs which were targeted for linking, CDA was the only BGC where the corresponding metabolite was not identified within the molecular network. This could be due to strains being cultured across media conditions which do not elicit the production of CDA, as only ISP1-5 were used. In one study, CDA was detected in the extracts coculture of

Streptomyces violaceoruber with *Streptomyces* MG7-G1 on SFM media supplemented with CaCl₂ (Schindl, Sharma, & Spiteller, 2020). In work specifically studying CDA production from *S. coelicolor* used a combination of the *S. coelicolor* 2377 mutant (commonly used for studies on CDA production) and SV2 media (Hojati *et al*., 2002). In both described studies, culture media contained calcium, which is absent from ISP1-5. Often to ensure the production of a specific metabolite, culture conditions play a large role and to promote CDA production, calcium should be present as an ingredient or supplement. Culture conditions supporting specific metabolite production is also seen with undecylprodigiosin (produced on ISP2 media), actinorhodin (ISP4) and coelimycin P1 (ISP4), the metabolites that were linked to their BGCs. Previous studies have shown that undecylprodigiosin is produced by *S. coelicolor* on ISP2 media (Chaudhary *et al*., 2014) (Schaberle, Orland, & Konig, 2014), and both actinorhodin and coelimycin P1 are produced when *S. coelicolor* is grown on ISP4 media (Uguru *et al*., 2013) (Nguyen, Riebschleger, Brown, Gorgijevska, & Nybo, 2021), highlighting the dependency on specific culture conditions for production of certain specialised metabolites. As such, the use of diverse media conditions will result the elicitation of more specialised metabolites and in turn, more comprehensive links.

In terms of discerning what constitutes a highly ranked link inferred by NP Linker, the score is relative to the dataset. Links were prioritised by mining the list of ranked links and matching examples where the score was amongst the highest in the dataset and the mass-to-charge ratio (*m/z*) of the spectra was close to the accurate mass of the metabolite in question. This method was effective with the gene knockouts because the accurate mass of each metabolite was known. As a result, this could generally be matched with *m/z* because the use of Electrospray Ionization (ESI) results in little fragmentation, giving an *m/z* close to the accurate mass of the metabolite (Banerjee & Mazumdar, 2012). The process of linking was confirmed by looking at the fragmentation patterns of the spectra, as metabolites fragment in ways that can be predicted (Schymanski, Meringer, & Brack, 2009). Therefore, the original files for the relevant linked spectra can be inspected using MZmine to confirm the presence of metabolite

fragments relating to the predictions based on structural identity (Watson, 2013). It is worth noting that the quality of NP Linker output depends on the quality of input data. The strains analysed in this study had fully sequenced genomes, however an incomplete genome can lead to incorrect prediction by antiSMASH due to clusters being fragmented across contigs (Baltz, 2017). A study by Skinnider *et al.* used manual BGC annotations to assess the frequency of false positives predicted by antiSMASH within a set of 200 randomly sampled clusters and suggested that around 55% of clusters predicted by antiSMASH represented false positives (Skinnider *et al.*, 2020). Therefore, utilisation of predicted BGCs from antiSMASH may cause false links inferred by NP Linker.

Regarding the accuracy of the NP Linker output and whether the links inferred are true links between a known metabolite and BGC, the evidence is much more robust when knockout mutants are used. In the case of the *S. coelicolor* deletion mutants (Δred, Δact, Δcda, Δcpk), the presence and absence of each metabolite within the dataset can be compared alongside the fragmentation patterns of the metabolites in the raw MS/MS data. The degree of certainty in the link is reduced where a deletion mutant is not available but can be reinforced by findings in the literature. Naringenin, typically considered a plant secondary metabolite, was linked to its BGC in *S. clavuligerus.* Interestingly, naringenin has been previously been reported as being produced by *S. clavuligerus* (Álvarez-Álvarez *et al.*, 2015). The author claims that this is the first reported instance of naringenin having been produced by a prokaryote, despite it being isolated from *Streptomyces graminofaciens* BA14348 in 1990 {Kondo, 1990 #508}. The dithiopyrrolone antibiotic holomycin was also linked to an *S. clavuligerus* BGC. Holomycin was first identified as an *S. clavuligerus* specialised metabolite in 1979 (Kenig & Reading, 1979) with the first characterisation of its BGC following much later in 2010 (B. Li & Walsh, 2010). To confirm any putative links between unknown BGCs and metabolites, wet laboratory techniques such as gene deletion or heterologous expression are required. In this work, NP Linker is shown to be a useful tool for prioritising links between metabolites of interest and their BGCs for further experimental analysis. This validation study shows that the NP Linker

standardised strain correlation scoring method can successfully link known metabolites to BGCs.

In conclusion, the metabologenomic tool NP Linker validly inferred links between 13 Actinomycetota specialised metabolites and their corresponding BGCs, including undecylprodigiosin, actinorhodin and coelimycin P1. In future, more diverse culture media should be used to maximise produced chemistry, including media for the targeted production of specific metabolites. Alongside this, a larger group of Actinomycetota should be analysed to produce more comprehensive links. Furthermore, a miniaturised culture method was developed in which enough extract dry mass – 1 mg/mL minimum - was produced to facilitate LC-MS/MS analysis at a reduced culture scale of 5 mL liquid media.

Chapter 4 - Eliciting bioactive specialised metabolites from *Streptomyces* in a soil microcosm system

4.1 Introduction

Coculture is a prominent method for the elicitation of bioactive specialised metabolites from streptomycetes, which involves the culture of two or more bacterial strains together so that their interactions may result in an observable change in behaviour, potentially unlocking previously silent biosynthetic pathways (Nicault *et al*., 2021). There are many different methods of coculture. For example, on agar plates *Streptomyces violaceoruber* was grown in coculture with *Streptomyces* sp. MG7-G1, which induced droplet production from the aerial mycelium of *S. violaceoruber*. The metabolite profile of the droplets produced in coculture, and agar plates of cocultures and monocultures were analysed via LC-MS/MS. Several metabolites were only produced by *S. violaceoruber* in coculture, including deacylated calcium-dependent antibiotics (daCDAs) (Schindl *et al*., 2020). Coculture can also be carried out in liquid media. The mycolic acid-containing bacterium *Tsukamurella pulmonsis* TP-B0596

and *Streptomyces* sp. TAKO-2 were co-inoculated in baffled Erlenmeyer flasks. This coculture lead to the production of julichrome Q_6 and julichrome $Q_{8,8}$ which were absent when the strains were grown in monoculture (Hoshino, Okada, Onaka, & Abe, 2016). Furthermore, microscale coculture of marine Actinomycetota was carried out in 500 μL of culture media in 96-deepwell plates and resulted in the elicitation of unique metabolites that were not present in strain monoculture extracts (Adnani *et al*., 2015). Cross-Kingdom coculture has also proved fruitful for successful specialised metabolite elicitation. For example, coculture of *Streptomyces* sp. 13F051 and the fungus *Leohumicola minima* 15S071 on solid media resulted in the production of the novel anticancer polyketide ulleungdolin, which was found to have been elicited from *Streptomyces* sp. 13F051 in response to fungal metabolites rather than physical interaction (Hwang *et al*., 2022). *Streptomyces* sp. have also been utilised in coculture to elicit the production of specialised metabolites from fungal species. Coculture between *Streptomyces rapamycinius* and *Aspergillus nidulans* resulted in the production of orsellinic acid and a yellow polyketide pigment, two specialised metabolites that are not produced when *A. nidulans* is grown in monoculture. Initially, it was thought that physical interaction between the mycelia of both species was responsible for the aforementioned elicitation (Schroeckh *et al*., 2009). It was later shown that the macrolide polaramycin B, produced by *S. rapamycinius,* was responsible for eliciting the production of both the orsellinic acid and the yellow pigment. The isolated polaramycin B was found to trigger the same response in *A. nidulans* even in the absence of bacterial cells (Berger *et al*., 2022). These interactions are important to consider because in nature, complex microbial communities are incredibly diverse and would not simply consist of bacteria-bacteria interactions.

In soil, *Streptomyces* interact in a diverse and dynamic microbial ecosystem (A. R. Muok, D. Claessen, & A. Briegel, 2021). It is important to consider that *Streptomyces* sp. may utilise the physical structure of soil to enable survival. For example, we know that in soil the *Streptomyces* produce substrate mycelium consisting of hyphae (Manteca, Fernandez, & Sanchez, 2005) which grow by tip extension and branch off within the soil, presumably on the

hunt for nutrients (Chater & Chandra, 2006) (Seipke *et al.*, 2012). The prevalence of these saprophytes in soil would appear to support the hypothesis that *Streptomyces* would grow in a laboratory-based soil microcosm system. However, antibiotic production in *Streptomyces* is notoriously variable even in tightly controlled conditions and attempts to reproduce the behaviour of cells *in situ* have not been straightforward. As an example of this variability, novobiocin production by *S. coelicolor* M512 showed a 39% standard deviation when cultured in baffled Erlenmeyer flasks (Stefanie Siebenberg, Prashant M. Bapat, Anna Eliasson Lantz, Bertolt Gust, & Lutz Heide, 2010). This is a standard laboratory liquid culture method, but the shaking movement causes culture media to splash onto the walls of the flask, resulting in growth away from the primary culture which contributes to variability (Büchs, 2001). If one were to speculate, the combination of natural environmental variability (for example, day-to-day changes in weather altering environmental conditions) and this variability in the behaviours of *Streptomyces* sp. would cause a lack of reproducibility in cultures which may make the idea of undertaking such experiments unappealing. Environmental variability does not extend to the laboratory however, and a soil sample could be transported to the laboratory for inoculation and analysis. This type of lab-based soil microcosm system has been utilised before, for successfully plasmid recovery testing (Wellington *et al.*, 1990) and determination of the metabolic activity of *Streptomyces* isolates in soil that they were not isolated from (Katsifas, Koraki, & Karagouni, 2000). A recent study used natural soil microcosm systems which had been amended with different sugars to analyse the inhibitory profiles of *Streptomyces* isolated from the soil (Dundore-Arias, Castle, *et al.*, 2019). These studies indicate that there is much still to discover about the behaviour of *Streptomyces* in soil that may prove to be advantageous in the fight against AMR.

In this chapter, *Streptomyces* strains were cultured in amended natural soil microcosm systems in both monoculture and coculture. It was hypothesised that the growth of *Streptomyces* would be supported in these conditions and the use of natural soil microcosm systems would represent a viable culture method for maximising chemistry. Furthermore, it

was hypothesised that in these conditions, coculture of *Streptomyces* would elicit chemistry and antimicrobial activity compared to monoculture.

4.2 Results

4.2.1 Stationary phase transition is strain dependant in *Streptomyces*

Gravimetric dry mass analysis was utilised to understand when each of the seven *Streptomyces* strains entered stationary phase (**Figure 4.1 a-g**). This method has been used previously to measure *Streptomyces* growth as mycelial aggregation renders OD_{600} readings unreliable (Shepherd, Kharel, Bosserman, & Rohr, 2010). The results showed that stationary phase, unsurprisingly, varied within the genus. *S. fradiae* and *S. platensis* reached stationary phase earliest at 48 hours. These were followed by *S. noursei, S. vinaceus* and *S. kanamyceticus* at 72 hours and finally *S. nodosus* and *S. spectabilis* at 96 hours. Based on these findings, five strains with the closest growth profiles were carried forward to interaction assays. These were *S. noursei, S. vinaceus, S. spectabilis, S. kanamyceticus* and *S. nodosus*, which reached stationary phase between 72 and 96 hours. Evolutionarily, the five selected strains represented four separate clades (**Figure 3.8**) indicating, based on this small dataset, that growth and phylogeny are not linked.
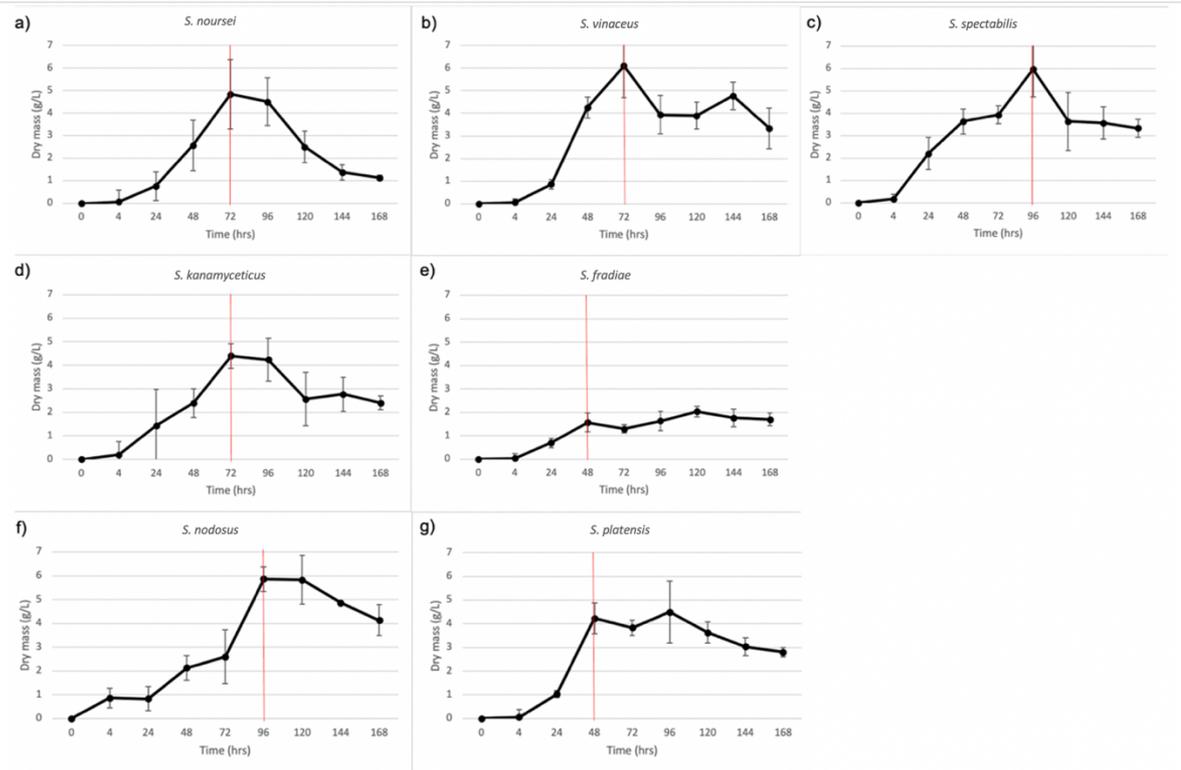
Figure 4.1 – Growth curves showing when each strain reached stationary phase Dry mass (g/L) growth curves to determine for **a)** *S. noursei*, **b)** *S. vinaceus*, **c)** *S. spectabilis*, **d)** *S. kanamyceticus*, **e)** *S. fradiae*, **f)** *S. nodosus*, and **g)** *S. platensis* reached stationary phase. 1 mL aliquots were obtained from each culture every 24 hours for six days. The spores were then separated from the liquid media using gravimetric dry mass analysis to pull the media through pre-weighted filter paper, which was then dried and re-weighted. *S. platensis* and *S. fradiae* reached stationary phase after 48 hours, the earliest of the seven strains, and were excluded from further analysis. Timepoint of stationary phase transition is marked by a red line on graph, with error bars indicating the standard deviation across replicates (n = 3).

4.2.2 *S. spectabilis* and *S. noursei* inhibit the growth of other *Streptomyces*

Challenge assays were used to elicit chemistry on petri dishes, which is why selecting strains with similar growth profiles was important. Each strain was challenged by four others, giving 20 strain/strain interactions across four media, resulting in 80 interactions. Herein the strains being challenged are named the 'protagonist' and were inoculated in a cross through the middle of the plate. The challenger strains are the 'antagonists'. Based on an inhibition scoring system, five strain-strain interactions were selected which allowed for streamlining of subsequent assays. These interactions were: *S. spectabilis* against *S. noursei, S. nodosus* and *S. vinaceus*; *S. noursei* against *S. nodosus* and *S. vinaceus* (**Figure 4.2**). When considering the strains in isolation, *S. spectabilis* and *S. noursei* were the most inhibitory towards other *Streptomyces* and significantly more inhibitory than *S. nodosus* (both $P = \leq 0.0001$) (**Figure 4.2**). Interestingly this is reflected in the BGC content, as both *S. spectabilis* and *S. noursei* genomes contain many BGCs (43 and 38 respectively) and *S. nodosus* contains 21. This would indicate that both *S. spectabilis* and *S. noursei* have the potential for greater specialised metabolite production which may explain the higher levels of inhibition, although further experiments such as gene knockouts would be required to confirm this. Further assays were set up using Gause's no 1 and GYM medium supplemented with soil amended with chitin, to test whether *S. nodosus* and *S. vinaceus* would be more inhibitory towards other streptomyces in the presence of N-acetylglucosamine (GlcNAc), the chitin monomer. Surprisingly, when cultured on soil-supplemented GYM media, very little inhibition was observed. *S. nodosus* was significantly more inhibitory towards *S. spectabilis* and *S. kanamyceticus* and *S. vinaceus* was significantly more inhibitory towards *S. spectabilis* when cultured on Gause's no 1 media compared to GYM (**Figure 4.3**).
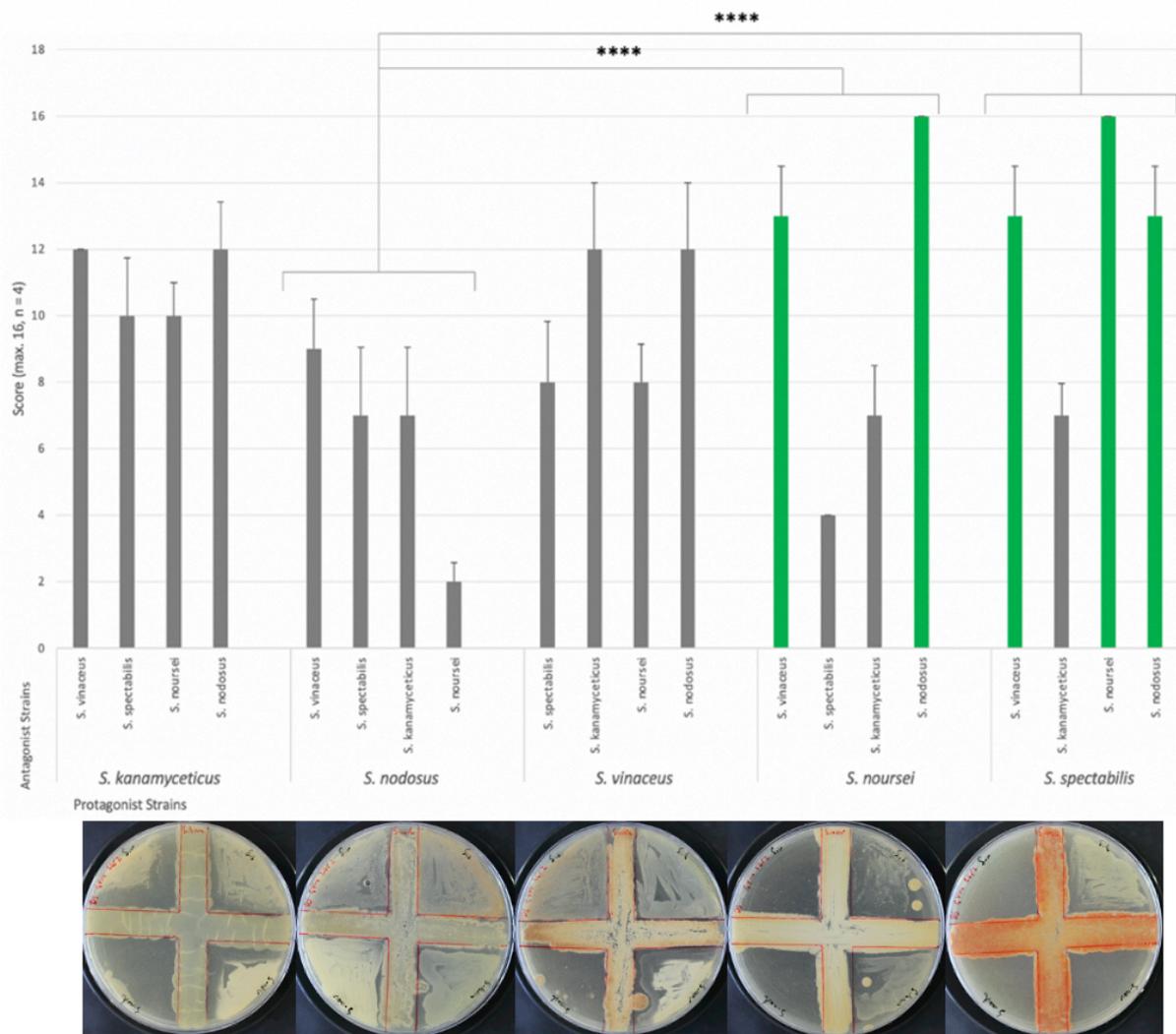
Figure 4.2 – Inhibition interactions from challenge assays between five *Streptomyces* strains. The scoring system was: if inhibition of < 1 cm occurred – 1 pt; > 1 cm – 2 pts; Complete inhibition – 4 pts. Protagonist strains were inoculated with spores to the centre of the plate and incubated (5 days, 30°C). Upon removal, the four antagonist strains were inoculated to the quadrants of the plate around the protagonist strain. The cultures were then incubated once more (5 days, 30°C). Interactions which scored highest and therefore prioritised for further analysis are marked in green. Error bars indicate the standard deviation of replicates where n = 4 (four media conditions, triplicate).
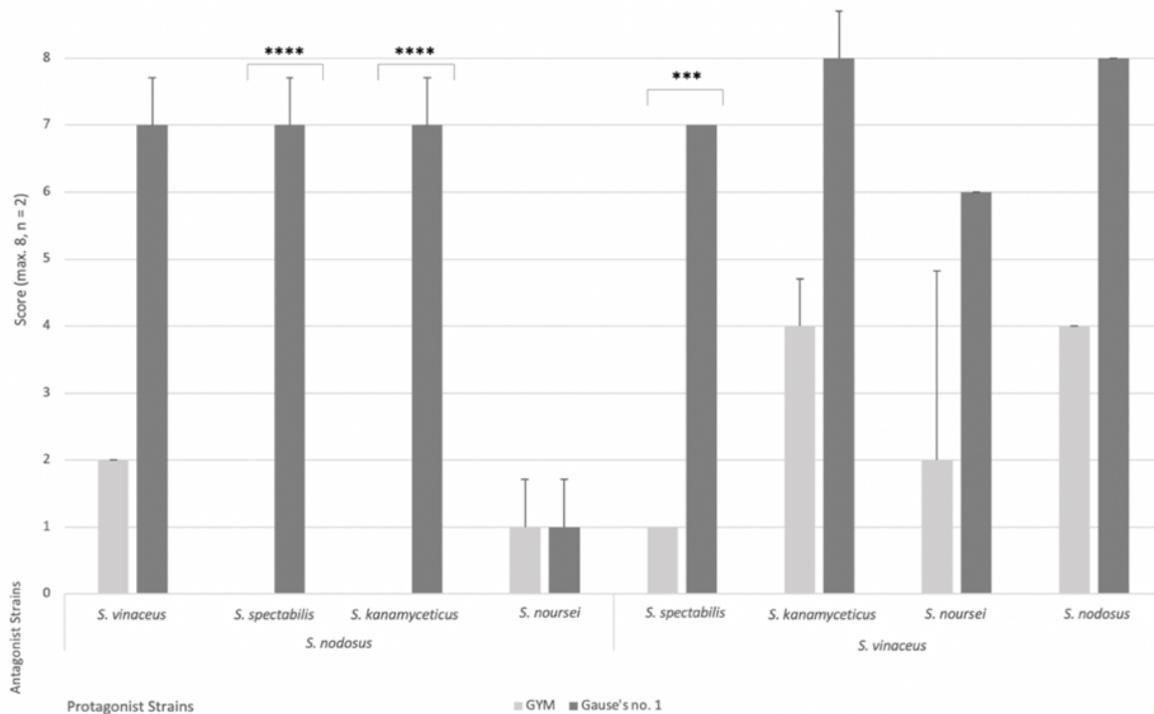
Figure 4.3 – *Streptomyces* challenge assays on soil-supplemented media. Scoring system was the same as **Figure 4.2**. Maximum score was 8, n = 2 as duplicate cultures were counted. Error bars represent standard deviation between duplicate cultures.

4.2.3 Soil analysis revealed a lower-than-average C:N ratio

To better replicate the behaviour of *Streptomyces* sp. in natural soil microcosm systems, soil categorised as brown earth was collected from Drumpellier Country Park in Coatbridge using sterile sampling bags, and composition analysis was conducted. The moisture percentage shows that the soil could be dried completely (**Table 4.1**). This is essential in allowing the moisture concentration of the soil microcosm systems to be fully controlled. Loss on ignition (LOI) determines the percentage of organic matter within the soil. The soil was burned at $450^{\circ}C$, resulting in a loss of 19.98% of dry weight. The pH ($CaCl_2$) of the soil was found to be 5.72, slightly under neutral but still within the ideal range for the growth of plants. This is beneficial for soil microbes as within this pH window plants produce high levels of root exudates as a carbon source allowing for microbial survival. The carbon:nitrogen (C:N) ratio

of the soil was 18:1, lower than the ideal 24:1 (USDA, 2011). This being the case, the soil within the microcosm systems was supplemented with soluble starch to increase the C:N ratio.

Table 4.1 – Characteristics of the soil sample. Carbon (C) and nitrogen (N) data expressed after drying at 50°C basis.

Sample ID	Moisture (%)	LOI 450°C (%)	pH (H_2O)	pH ($CaCl_2$)	C (%)	N (%)
Drumpellier Soil	0.00	19.98	5.94	5.72	10.50	0.57

Table 4.2 – Elemental trace metal composition (mg/Kg) of the analysed soil. On an oven dry (105°C) basis.

Element	Concentration	Element	Concentration
Ag	<0.01	**Mo**	1.53
As	27.47	**Ni**	43.64
Ba	278.0	**Pb**	235.2
Cd	0.99	**Pt**	<0.01
Co	15.96	**Se**	2.55
Cr	58.05	**Sr**	86.0
Cu	60.89	**Zn**	370.3
Hg	0.23		

4.2.4 Bioassays of soil microcosm cocultures revealed coculture fractions caused more pathogen inhibition than monocultures

Monoculture and coculture extracts were fractionated using Solid Phase Extraction (SPE), utilising a polarity gradient resulting in metabolite mixtures split across methanol/ddH$_2$O 80/20, methanol, acetonitrile and ethyl acetate fractions. Each culture was set up in triplicate, resulting in 12 fractions per culture. In total, there were 108 fractions: cocultures of *S. spectabilis*/*S. noursei, S. spectabilis/S. vinaceus, S. spectabilis/S. nodosus, S. noursei/S. vinaceus* and *S. noursei/S. nodosus,* as well as monocultures of each strain. There were 42 instances of bioactivity observed against both pathogens across the 108 extract fractions tested. Overall, 28 and 14 fractions showed bioactivity against *E. coli* and *S. aureus* respectively (**Figure 4.4**) (**Supplementary Figure 4.1**)*.* The *S. spectabilis*/*S. noursei* coculture showed the highest bioactivity against *E. coli*, with activity observed in 7 of its 12 fractions. As with the inhibition assays outlined in the previous section **4.2.2**, these strains possess the highest number of BGCs, so it could be hypothesised that they would produce more specialised metabolites, increasing the chances of observed bioactivity. Against *S. aureus, S. vinaceus* showed the highest level of bioactivity with 4 of its 12 fractions. MeOH/H$_2$O fractions were the most bioactive, indicating that the metabolites responsible for the observed bioactivity from these fractions are likely to be polar. In general, observed bioactivity was limited to small (1-2 mm) Zones Of Inhibition (ZOI), potentially due to the small number of bioactive metabolites that were captured in the extraction process or the large number of fractions. No bioactivity was observed within soil blank fractions, confirming that the observed bioactivity was a result of bacterial metabolites.

Figure 4.4 – Radar plot of bioactivity of mono- and coculture metabolite fractions from _Streptomyces_. Strains were inoculated to the soil microcosms in monoculture and/or coculture, and incubated (5 days, 30°C). Upon removal, the soil culture was freeze dried, transferred to a 250 mL Erlenmeyer flask, covered with 25 mL EtOAc, and shaken (180 rpm, overnight). The solvent and culture were separated by filtration into a pre-weighted scintillation vial, dried under N_2 and re-weighted. Extracts were then resuspended in EtOAc and to 1 mg/mL and 30 μL was inoculated to a paper disc. The discs were then placed on _S. aureus_ and _E. coli_ pathogen overlay plates and incubated (overnight, 37°C) with ZOIs measured upon removal. In each group n = 12, with all four fractions in triplicate.

4.2.5 Comparative metabolomics of soil microcosm systems revealed the elicitation of bioactive specialised metabolites

A total of 42 microbial metabolite extracts (consisting of four strains in soil microcosm monoculture and four coculture conditions – *S. spectabilis* and *S. nodosus* coculture was not included because due to a lack of bioactivity) were used to build a molecular network. The network consisted of 983 parent ions and 1159 edges (**Figure 4.5 a**). There was a total of 41 molecular families of structurally related metabolites identified, with the largest consisting of 42 parent ions. 763 parent ions were singletons and did not structurally correlate to any other parent ion detected. This indicates a high level of chemical diversity within the dataset. GNPS annotation identified the specialised metabolites oleanolic acid in files corresponding to *S. noursei,* and desferroxamine E in files corresponding to *S. nodosus.* The oleanolic acid BGC is not predicted by antiSMASH as being present within the BGC suite of *S. noursei*, but desferroxamine E is present within the suite of *S. nodosus.* In total, 91 parent ions appeared within coculture data exclusively, indicating that they had been elicited by the presence of a challenger strain within the microcosm systems. 248 parent ions were detected within data corresponding to monocultures, indicating that these metabolites may be actively suppressed within cocultures. Many parent ions appeared within the network which were detected exclusively within certain culture conditions. The strain with the highest number of exclusively related parent ions was *S. noursei*, with 34. 34 parent ions were also found to correspond to data related only to coculture of *S. spectabilis* and *S. noursei*, which was the highest amongst cocultures and potentially expected due to the two strains involved within the coculture having the highest number of BGCs predicted within their genomes (**Figure 4.5 b**).

There were some areas of particular interest. For example, one molecular family consisted of 8 nodes, with mass-to-charge ratios (*m/z)* ranging from 408.362 to 438.410, and all nodes appearing exclusively within coculture files. All nodes were detected in data relating to *S. spectabilis* and *S. noursei* coculture. These related to the peptidyl nucleoside antibiotic

blasticidin S predicted within the genome of *S. spectabilis* upon mining using antiSMASH. Blasticidin S is an antibiotic with an accurate mass of 422.4. One node within the molecular family had an *m/z* of 422.378 Da, and upon inspection of the fragmentation pattern within the raw data peaks (**Figure 4.6**), seven fragments matching the predicted blasticidin S fragmentation pattern were found indicating that the detected metabolite is blasticidin S and has been elicited from *S. spectabilis* in soil microcosm coculture with *S. noursei*, and as such is potentially responsible for bioactivity observed in **Figure 4.4**. Further experimental work in the form of gene knockouts or heterologous expression would be required to fully determine the identity of the metabolite (P. N. Tran, M.-R. Yen, C.-Y. Chiang, H.-C. Lin, & P.-Y. Chen, 2019).

a)

b)
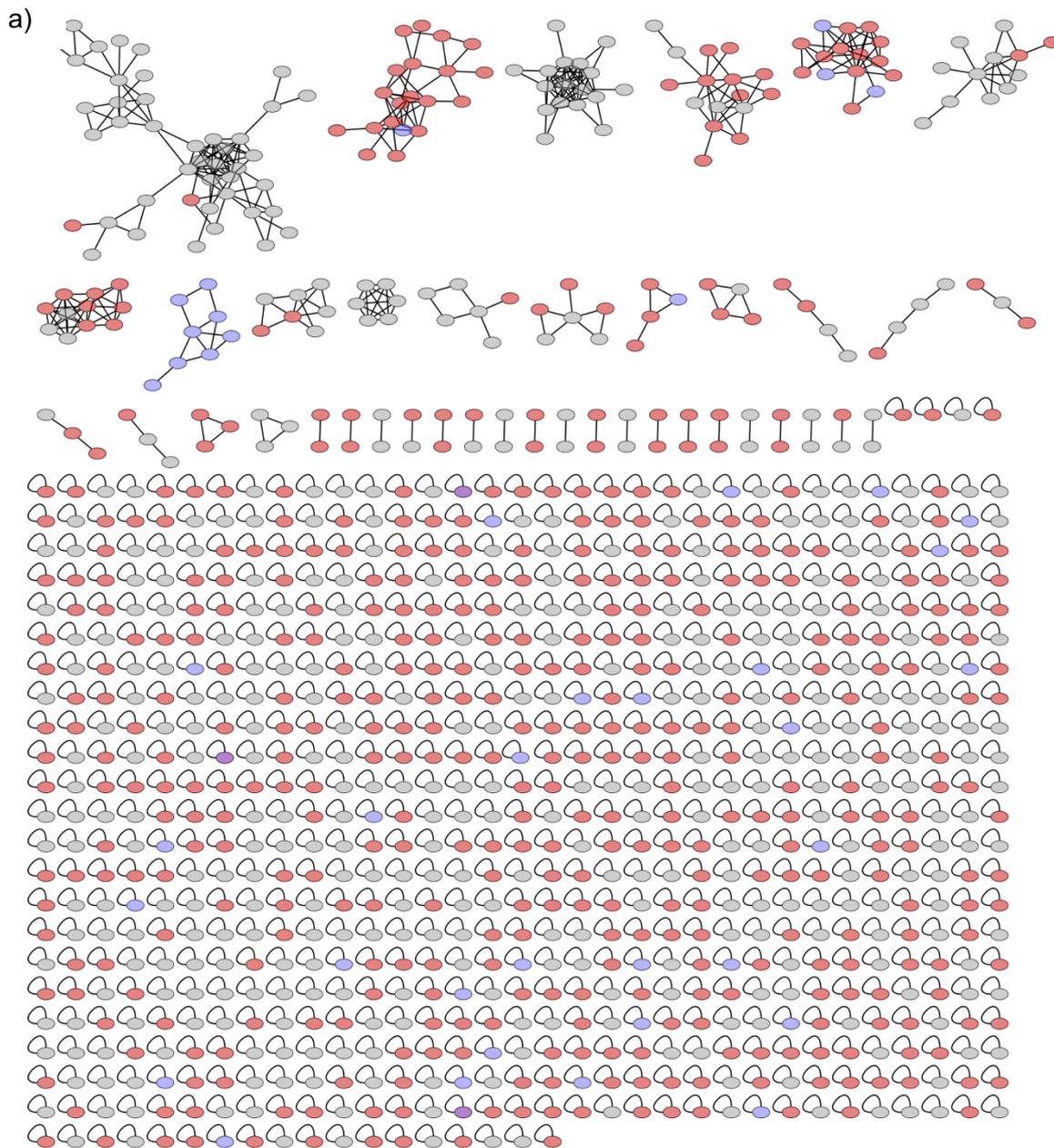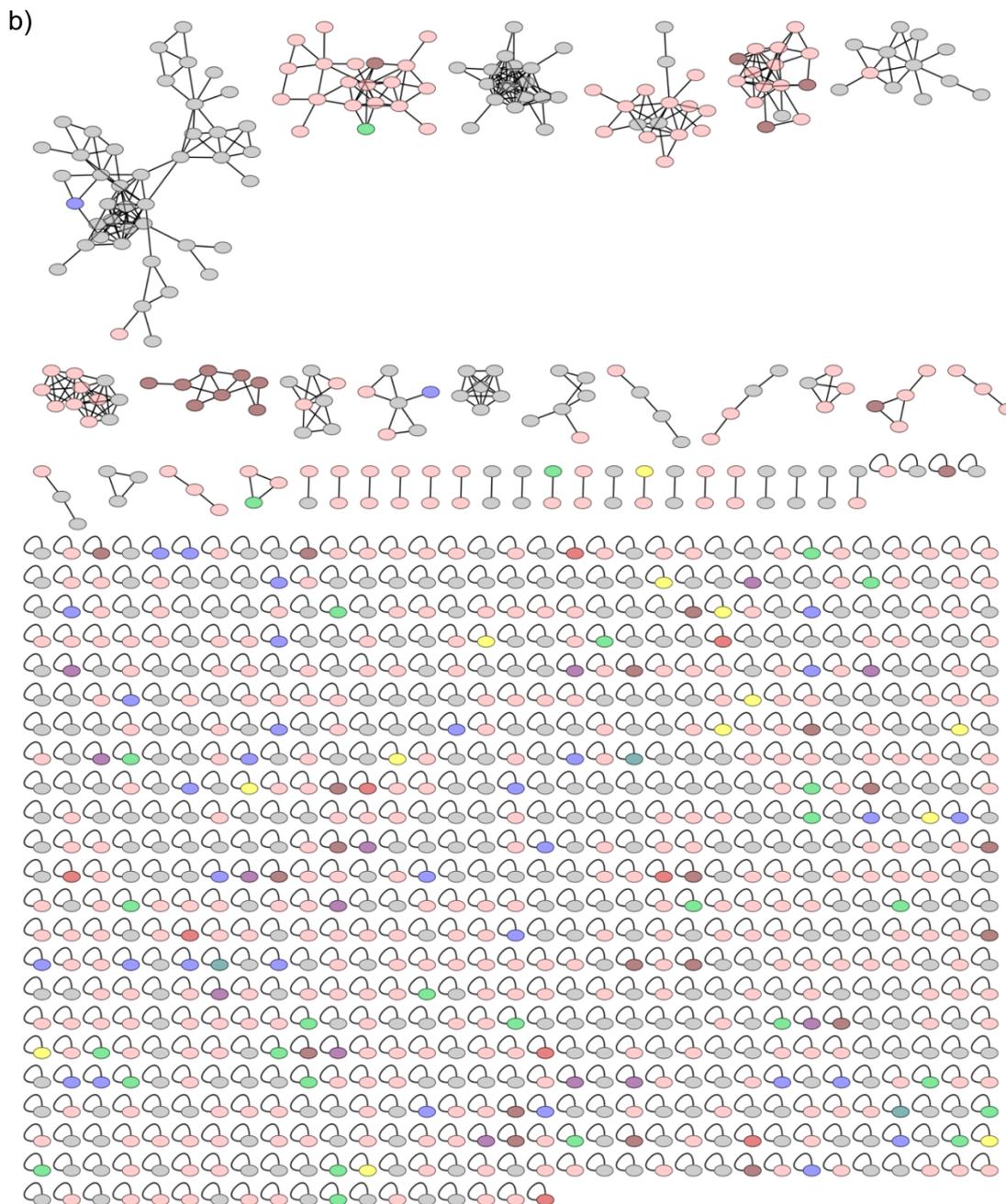
Figure 4.5 - GNPS molecular networks of 983 parent ions produced by four strain monocultures and four coculture conditions. **a)** grey nodes correspond to soil blanks (456), pink nodes correspond to nodes detected in both monoculture and coculture (188), red nodes were detected exclusively in monoculture (248) and blue nodes exclusively in coculture (91). **b)** node colours refer to organism(s) in culture. Red – *S. spectabilis* (9 nodes), blue – *S. noursei* (34), green – *S. vinaceus* (28), yellow – *S. nodosus* (13), brown – *S. spectabilis/S. noursei* (34), purple – *S. noursei/S. vinaceus* (14), jade – *S. noursei/S. nodosus* (3), grey – blank (456).
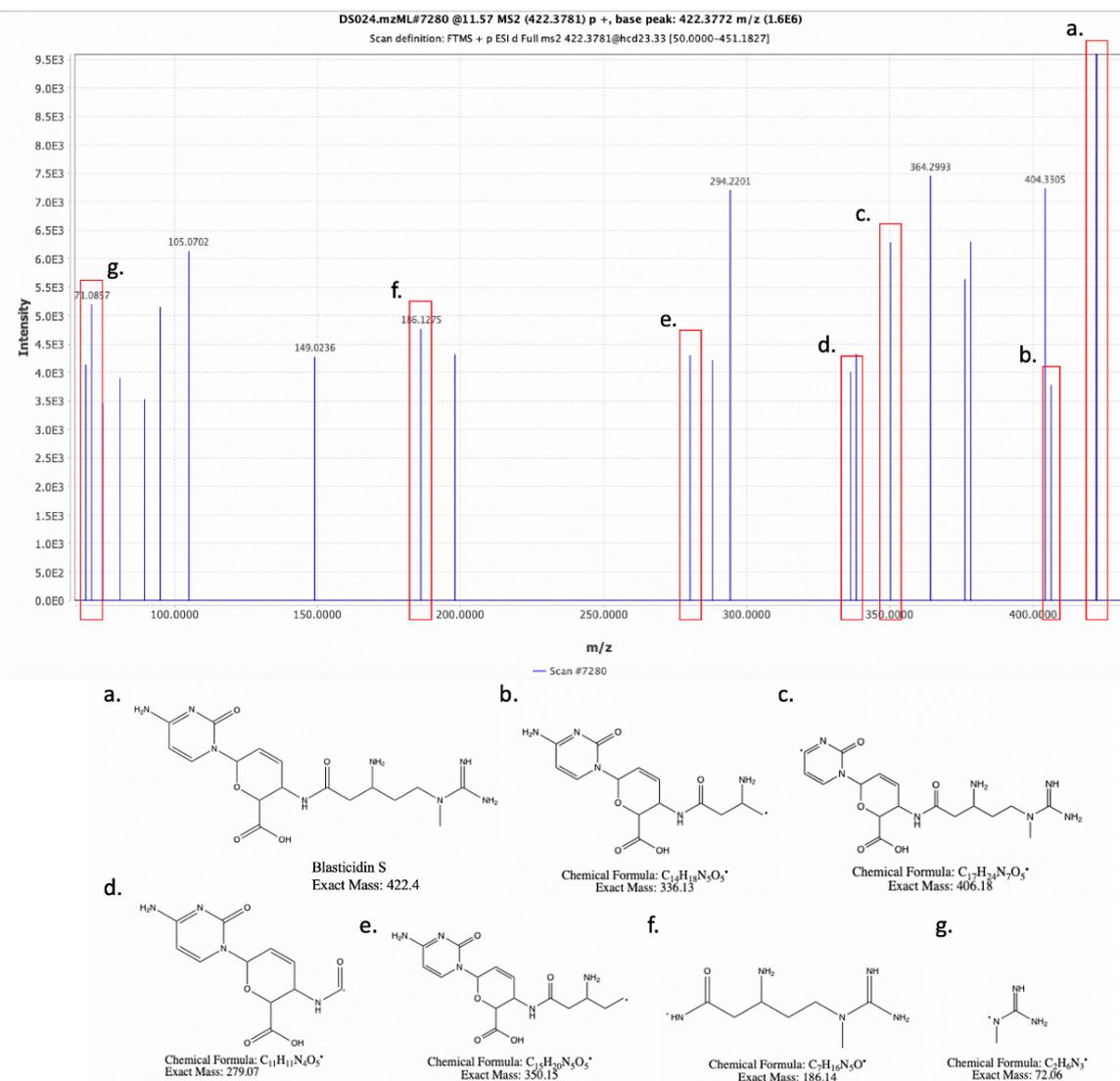
Figure 4.6 – Raw peak data of potential blasticidin S related metabolite produced by *S. spectabilis* in coculture with *S. noursei*. Raw peak data for scan no. 7280 (*m/z* 422.378 Da, RT Mean 694.334). The parent ion was detected within *S. spectabilis* and *S. noursei* soil microcosm coculture extracts after the culture was freeze dried and extracted using EtOAc. Peak **a** corresponds to the accurate mass of the intact Blasticidin S metabolite and **b - g** are metabolite fragments which correspond to how Blasticidin S is predicted to fragment during MS2 analysis.

4.2.6 Metabologenomic analysis of soil microcosm systems revealed a link between a produced metabolite and a putative siderophore BGC

Several links between metabolites produced within the soil microcosm systems and BGCs were made using NP Linker. For example, spectrum ID #13883 (which had a mass-to-charge ratio (*m/z*) of 1411.44 Da and a mean retention time (RT Mean) of 656.041) was linked to a singleton Gene Cluster Family, ID #100 (**Supplementary Figure 4.2**). The BGC was identified by antiSMASH as the glycopeptide antibiotic zorbamycin, which has an accurate mass of 1412.5, within the *S. noursei* genome. Although the *m/z* and the accurate mass of zorbamycin differ slightly, this could potentially be attributed to ionisation, as with Electron Spray Ionisation (ESI) a proton is lost. The ion was also produced by *S. spectabilis* and the zorbamycin BGC was predicted within the *S. spectabilis* genome. The BGC within the *S. noursei* and *S. spectabilis* genomes only shared 4% and 8% genes with zorbamycin respectively and few zorbamycin fragments were observed within the raw spectral data. This indicates that the metabolite detected may not be zorbamycin but shares some genes and has a similar accurate mass. Another link included the GCF with the ID #80 containing predicted BGCs from all four *Streptomyces* strains analysed was linked to a spectrum (ID #7908) also detected within the data relating to all analysed *Streptomyces* strains (**Figure 4.7 a**). For three out of four predicted BGCs, antiSMASH predicted similarity to the siderophore ficellomycin. However, with *S. noursei*, the linked BGC was predicted to be a siderophore with no hit to a known metabolite BGC. The BGC predicted within the *S. spectabilis* genome contained two extra genes not found within the related clusters (**Figure 4.7 b**), identified by antiSMASH as an aminotransferase class III-fold pyridoxal phosphate-dependent enzyme and an iron transporter. BLAST analysis of the aminotransferase class III-fold pyridoxal phosphate-dependent enzyme showed it shared 100% identity with the diaminobutyrate--2-oxoglutarate transaminase family protein within *S. spectabilis* (NCBI accession numbers WP_221515291.1 and WP_229878709.1). This protein reversively catalyses the conversion of L-aspartate beta-semialdehyde (ASA) to L-2,4-diaminobutyrate (DABA) by transamination with L-glutamate.

The structure of the protein was predicted using AlphaFold (Jumper *et al.*, 2021) (**Figure 4.7 c**). The *m/z* of the linked spectrum was 403.232, which is vastly different from the accurate mass of ficellomycin (312.37). The ionization method used for the analysis was ESI, a soft ionization method, so it would be expected that the *m/z* of any linked spectra would be similar to the accurate mass of the BGC product. With this in mind, combined with the fact that percentage similarity within the ficellomycin predictions was again low (3% in every BGC suite), it could be said that the produced metabolite is a siderophore although it is likely not ficellomycin due to the low number of shared genes.

Rosetta scoring, a feature-based correlation scoring method facilitated by NP Linker (Hjörleifsson Eldjárn *et al.*, 2021), linked spectrum ID #236 to a BGC within the *S. noursei* genome. This was identified as the linaridin Ribosomally synthesised and Post-translationally modified Peptide (RiPP) legonaridin, which upon isolation was found to have an *m/z* of 901.223 (M. E. Rateb *et al.*, 2015). This differs significantly from the *m/z* of the Rosetta-linked spectrum (261.131). Links were also inferred by the Rosetta method between spectrum ID #7841 (*m/z* 388.211, RT Mean 642.826) and BGC ID's #50 and #23 (region 28 and region 5 of the *S. spectabilis* and *S. nodosus* BGC suites respectively). antiSMASH identified the BGC as kanamycin, the product of which has an accurate mass of 484.499 Da and again differs significantly from the *m/z* of the Rosetta-linked metabolite. The analysis produced few links, likely due to the fact that the dataset was small. An expanded dataset would allow greater validation through more links.
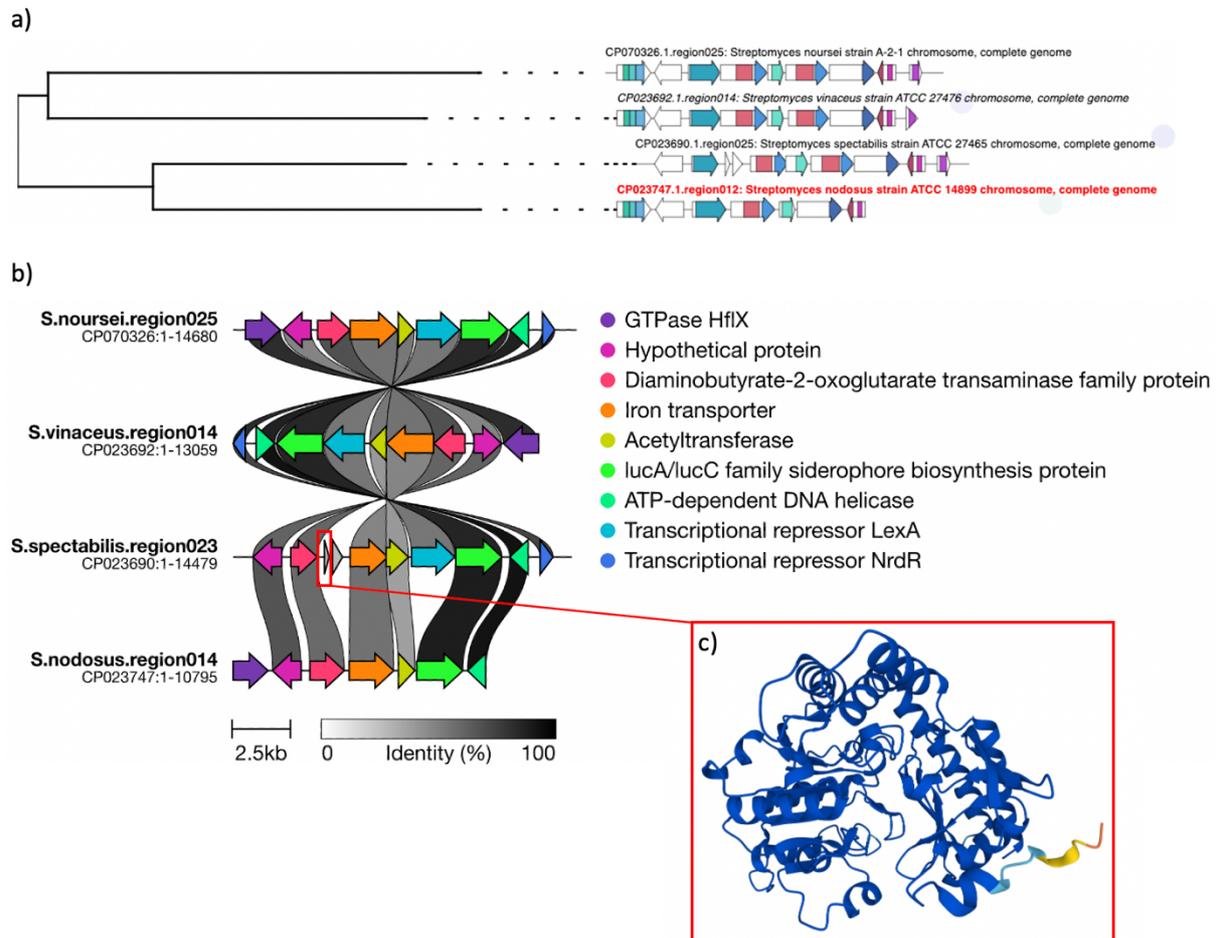
Figure 4.7 – Analysis of GCF predicted to be linked to a putative siderophore by NP Linker. a) Phylogeny of the four BGCs that make up the GCF (#80) which was linked to spectrum ID #7908. **b)** Identity of genes found within the related BGCs. The BGC within the *S. vinaceus* genome is flipped, and the version in *S. spectabilis* contains an extra two genes – identified by antiSMASH v6.0.1 as an aminotransferase class III-fold pyridoxal phosphate-dependent enzyme and an iron transporter. **c)** Comparison via NCBI BLAST revealed 100% sequence homology to the diaminobutyrate--2-oxoglutarate transaminase family protein within *S. spectabilis*, the structure of which has been predicted using AlphaFold.

4.3 Discussion

Each of the five *Streptomyces* strains used in the analysis showed antibacterial activity against cocultured environmental strains. This is also highlighted in a study where the model strain *S. coelicolor* A3(2) was shown to inhibit 40 soil streptomycetes, chosen as targets due to their importance in soil environments (Z. Zhang *et al*., 2020). This activity was to be expected as *Streptomyces* is known to be a prolific producer of antimicrobials capable of inhibiting other members of the genus. One of the strains responsible for the highest level of inhibition was *S. spectabilis.* Previous work has shown that a strain with 99.9% sequence homology to *S. spectabilis* produced the characteristic red pigment metacycloprodigiosin. The isolated pigment exhibited strong antibacterial activity against clinically important pathogens such as MRSA (Meng-Xi, Hui-Bin, Jie-Yun, Jing-Xiao, & Zhen-Wang, 2021). It should be noted that similar red prodigiosin pigments (undecylprodigiosin) have been produced by bioactive *S. spectabilis* cultures in this work. *S. noursei* was the next most inhibitory towards *Streptomyces.* Molecular network analysis of LC-MS/MS data of *S. noursei* revealed Noursamycins A and B, which when isolated were found to inhibit Gram-positive bacteria (Mudalungu *et al*., 2019). The level of inhibition from both *S. spectabilis* and *S. noursei* was perhaps predictable, as these strains have the highest number of BGCs from the set of analysed strains and therefore should be capable of producing the most antimicrobial specialised metabolites. Interestingly, in the challenge assays where *S. nodosus* and *S. vinaceus* were cultured on GYM supplemented with chitin-amended soil, there was no increase in bioactivity as anticipated. This was surprising because the genomes of both strains contain GntR-family regulatory genes as predicted by antiSMASH. The monomer of chitin, N-acetylglucosamine (GlcNAc), acts as an important signalling molecule in the switch between primary and specialised metabolism by binding GntR-family genes (Swiatek, Tenconi, Rigali, & van Wezel, 2012). In one study GlcNAc was found to significantly increase yields of the anti-cancer glycopeptide antibiotic bleomycin (H. Chen, Cui, Wang, Wang, & Wen, 2020) and in famine conditions,

GlcNAc exposure promotes antibiotic production in *S. coelicolor* A3(2) (Swiatek, Urem, Tenconi, Rigali, & van Wezel, 2012). Conversely, GlcNAc has been shown to have a negative effect on tacrolimus production by *Streptomyces tsukubaensis*. In this case, GlcNAc represses fkbN and ppt1 transcription, both gene clusters involved in tacrolimus biosynthesis (Ordóñez-Robles, Rodríguez-García, & Martín, 2018). It has also been shown that high concentrations of extracellular GlcNAc can prevent *S. coelicolor* A3(2) developing beyond a vegetative state, with the effect absent in mutant strains possessing defective GlcNAc transport systems (Rigali *et al*., 2006). Therefore, it is not necessarily always the case that chitin amendments to the soil will elicit chemistry and result in more bioactivity. Instead, the effect is strain dependant. These results did not affect the strain/strain pairings carried forward into the soil microcosm assays.

The peptidyl nucleoside antibiotic blasticidin S was detected in *S. spectabilis* monoculture metabolite extracts which inhibited *E. coli*. This metabolite was first isolated from *Streptomyces griseochromogenes* (Takeuchi, Hirayama, Ueda, Sakai, & Yonehara, 1958) and inhibits cytosolic protein synthesis in both eukaryotes and prokaryotes by blocking termination of ribosomal translation (de Carpentier *et al*., 2020). It exhibits strong fungicidal activity and as such, was the first antibiotic used to control rice blast fungus in China (Niu, Zheng, & Tan, 2017) (K. T. Huang, Misato, & Asuyama, 1964). Because of this eukaryotic toxicity, blasticidin S is not clinically used as an antibiotic. However, novel blasticidin S analogues with increased drug-like properties and potent antimicrobial activity against *E. coli* and *S. aureus* have been reported (Davison *et al*., 2017). In order to confirm the identities of metabolites of interest identified in this work, further experiments are required. For instance, the BGC within *S. spectabilis* genome putatively identified as blasticidin S could be deleted using CRISPR/Cas9-based editing tools (Tao, Yang, Deng, & Sun, 2018). CRISPR/Cas9 tools have been successfully used for BGC deletion in *Streptomyces* models. For example, a CRISPR/Cas9 genome editing plasmid was used to delete single genes such as *actII-ORF4* and *glnR* in *S. coelicolor* M145 as well as whole BGCs such as actinorhodin, undecylprodigiosin and CDA

with efficiencies of between 60% and 100% (H. Huang, Zheng, Jiang, Hu, & Lu, 2015). In future, such tools could be used to delete the putatively identified blasticidin S BGC and the mass spectrometry data checked for the presence of blasticidin S peaks. Thus, providing further evidence that the now absent metabolite was produced by the predicted BGC.

The methanol/water fraction was generally the most bioactive, including the fraction from the *S. spectabilis* monoculture when pitted against *E. coli* and *S. aureus*. Mixed methanol/water fractions are often the source of bioactivity in work of this type. In one such study, a methanol/water mix was used to extract metabolites from *Streptomyces* SA32. These solvent extracts were tested and inhibited all ESKAPE pathogens, although the metabolites responsible for the bioactivity were not identified (Radjasa, Oedjijono, & Ryandini, 2021). The polarity of a specialised metabolite also determines which fraction they are captured by. For instance, *S. spectabilis* is a known producer of the polar aminoglycoside antibiotic spectinomycin (K. R. Kim, Kim, & Suh, 2008) which has previously been shown to have activity against both *E. coli* and *S. aureus* (Wendlandt *et al*., 2013). It could be speculated that spectinomycin is potentially the metabolite responsible for the observed bioactivity against these pathogens. The methanol/water fraction from the *S. vinaceus* monoculture was bioactive against *S. aureus*. *S. vinaceus* is responsible for the production of the polar specialised metabolite, viomycin. This metabolite has also shown activity against *S. aureus* (Thomas, Chan, & Ozanick, 2003). If these are indeed the specialised metabolites responsible for the activity against *S. aureus,* the polarity of both metabolites would result in capture by the highly polar methanol/water fraction. Bioactivity is then lost within the cocultures, so it could be the case that coculture is suppressing production of the metabolites responsible for the observed activity. It should also be noted that the microbial metabolite extraction from soil is not often undertaken and may require further optimisation.

One major factor to consider when attempting to utilise natural soil as a culture medium is the lack of reproducibility. For example, one soil sample from the same location may vary in terms

of nutrients from one day to the next. The soil used in this study exhibited a lower-than-average C:N ratio according to USDA guidelines, highlighting potential nutrient variability between samples. The layer of soil sampled in this study, known as the Humus form and comprised mostly of mull humus, is known to be the primary sight for biological and chemical processes underpinning the function of the terrestrial ecosystem (Z. S. J. Ponge, 2010). A study of 148 variables concluded that geology and climate were the major determinants in humus forms in temperate forests where such soil samples could be gathered from (J.-F. Ponge, Jabiol, & Gégout, 2011). Topsoil pH is also known to affect soil microbiological parameters (Hellwig *et al*., 2018), with rainfall being a contributing factor in the increase of soil acidity (L. Li *et al*., 2022). Thus, it can be predicted that alterations in these parameters would cause downstream changes in nutrient conditions of samples collected at different time points. In future an artificial soil microcosm system based on the nutrient composition of the soil will be developed and the chemistry of the strains inoculated will be compared to that of natural soil microcosm systems.

To summarise, coculture of *Streptomyces* in natural soil microcosm systems elicits the production of bioactive specialised metabolites and solvent extracts from cocultures exhibited more bioactivity against pathogens than monocultures, including blasticidin S. Interestingly, molecular networking revealed that 157 more parent ions were produced by strains in monoculture than in coculture. This highlights that the culture condition itself plays an important role in eliciting specialised metabolites. Metabologenomic analysis also linked a putative siderophore to its corresponding BGC, however more links would have been established with a wider dataset.

Chapter 5 – Development of a standardised artificial soil microcosm system

5.1 Introduction

The Actinomycetota genus *Streptomyces* is ubiquitously distributed in soil (Seipke *et al*., 2012) but despite this, studies into the behaviour of *Streptomyces* and bacteria in general *in situ* are rarely undertaken (Sheth, Cabral, Chen, & Wang, 2016). Work in Chapter 4 showed natural soil as a viable alternative to standard laboratory media for microbial culture yet obtaining reproducible conditions for long-term experiments is challenging. One cannot guarantee the continuity of soil nutrient composition if multiple samples are taken from the same coordinates at different time points and attempts to replicate the natural environment in a laboratory setting often fall short (Behie *et al*., 2017). Environmental variation in abiotic parameters such as rainfall can cause fluctuations in soil pH due to leached alkali elements (Ulén, 2020) (L. Li *et al*., 2022) which has been shown to influence bacterial diversity (Feng *et al*., 2023). This could have implications for streptomycetes, which grow optimally at neutral to alkali pH (Kontro, Lignell, Hirvonen, & Nevalainen, 2005) (Ripa, Nikkon, Zaman, & Khondkar, 2009). Variations in temperature have also been shown to affect the rate of microbial respiration in soils. Interestingly, a study of 25 diverse soil samples (varying elevations, geographic locations, and climate conditions) revealed the average optimum temperature for microbial respiration was 42.4°C, far higher than the assumed model value (Y. Liu *et al*., 2018). Furthermore, climate and soil nutrients such as sulphur and nitrogen are closely correlated and an increase in temperature due to climate change will alter the levels of these environmental nutrients (Shao *et al*., 2021) (Ni *et al*., 2022).

To overcome such issues in variability, several studies have focussed on artificial soil systems in a laboratory setting. For example, transparent substrates have been used for tackling problems with the physical modelling of interactions in soil, imaging soil thermal processes and imaging plant root interactions with bacterial species (Iskander, Bathurst, & Omidvar,

2015). This includes 3D imaging of rhizosphere interactions between the human pathogen *E. coli* O157:H7 and lettuce roots, that showed micro-colony formation which contributed to the survival of bacteria in extreme environments (Downie *et al*., 2012). The transparent polymer Nafion has been used as a substrate with tuneable mineral content. Nafion was compared to the naturally occurring crystal cryolite as a transparent soil substrate with both able to support the growth of *B. subtilis* NCIB3610, showing that could be useful in future study of microbial growth dynamics (Sharma, Palatinszky, Nikolov, Berry, & Shank, 2020). However, Nafion is often prohibitively expensive at around $1,000/kg and does not physically absorb nutrients, which has led to the development of hydrogel bead-based artificial soil systems for studies where high-standard imaging is of lower priority. For example, transparent soil was developed for plant root growth and monitoring by dropping a gellan gum/alginate solution which forms gel beads. The media supported soybean root growth to a significantly similar level than that of natural soil (Ma *et al*., 2019). A similar technique using a Phytagel™ (Sigma Aldrich) and sodium alginate solution supported the growth of *B. subtilis* B5-P1, *Pedobacter* sp. D749, *Rhodococcus globerulus* D757, *Stenotrophomonas indicatrix* D763, and *Chryseobacterium* sp. D764. Moreover, lipopeptide production was also monitored using HPLC-MS showing that metabolites could be extracted from the media using organic solvent (Lozano Andrade, Nogueira, Wibowo, & Kovács, 2022). While the application of these methods to the growth of bacteria has proved advantageous, certain caveats should be considered before *in situ* use with Actinomycetota specifically. Actinomycetota are essential to rhizospheric interactions because filamentous growth and sporulation promote the formation of a strong physical bond to soil particles (Olanrewaju & Babalola, 2019). With this in mind, the importance of the physical soil particle may be essential for bacterial replication in soil. To our knowledge, no studies have investigated artificial soils to support the growth of Actinomycetota in a laboratory setting with a view to enhancing specialised metabolite production.

In this study, two environmental *Streptomyces* isolates as well as two sequenced strains were inoculated in monoculture and coculture to a hydrogel-based and sand-based artificial soil.

The strains were also inoculated in natural soil and laboratory standard culture media microcosm systems for comparison of chemistry. It was hypothesised that the artificial soil microcosm systems, with nutrient conditions tuned to that of natural soil, would provide a viable and sustainable alternative to previously described natural soil microcosm systems and that specialised metabolites may be elicited as the culture would more closely resemble natural ecosystems.

5.2 Results

5.2.1 *Streptomyces* from soil showed bioactivity against *B. subtilis* ATCC 23857

Seven isolates were morphologically identified as *Streptomyces*. They were isolated on Gause's no. 1 in undiluted (five isolates) and 1/100 diluted (two isolates) Ringer's solution. The isolates were named DS1 – DS7. A number of methods were employed to narrow the seven isolates down to two for future coculture experiments. Firstly, plugs from plates of all seven isolates cultured in triplicate on MS media were tested for antibiotic activity against *B. subtilis* ATCC 23857 (**Figure 5.1**) (**Supplementary Figure 5.1**). As expected, Zones Of Inhibition (ZOIs) were visible around all isolate plugs, confirming they all produced inhibitory metabolites. No inhibition was observed around the negative control, which was a blank MS plug, and spectinomycin was used as a positive control. DS2 plugs caused the most inhibition of *B. subtilis* at 16.3 mm, narrowly more than DS3 (16.1 mm) and DS5 (15.8 mm). Across triplicate cultures, all isolate plugs caused >10 mm pathogen inhibition, ranging from 10.3 mm to 16.3 mm. On MS medium, DS5 formed large, smooth colonies with dark grey spores and a black pigment diffusing into the media (**Figure 5.3 a**). DS1 formed colonies of a similar size with light grey spores. Droplets appeared after four days of incubation and large patches of very dark red pigment diffused into the media (**Figure 5.3 b**). The dark red pigment could potentially be undecylprodigiosin, commonly produced by *Streptomyces*.
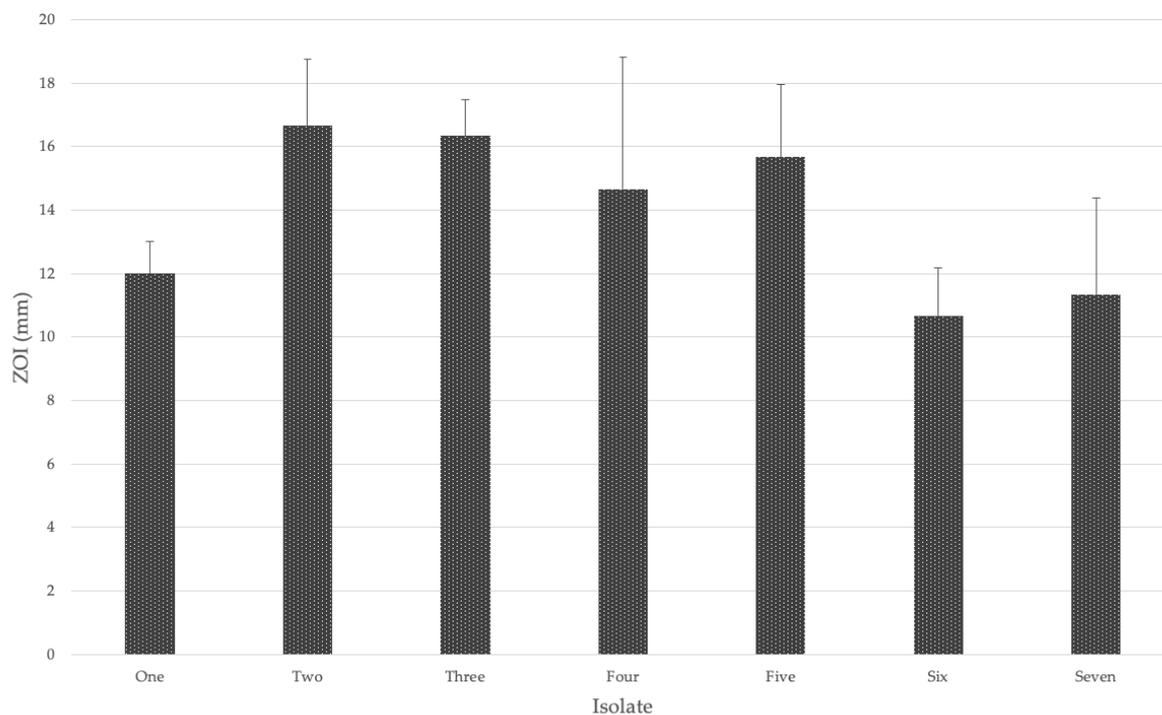
Figure 5.1 – Bar chart showing zones of inhibition (mm) of isolates against *B. subtilis* **ATCC 23857.** Seven strains were isolated from Drumpellier park soil and morphologically identified as Actinomycetota (DS1 – DS7) and subcultured until pure. Each isolate was cultured in triplicate on MS medium, and then then tested for bioactivity against the common Actinomycetota indicator strain *B. subtilis.* DS2 showed the highest level of inhibition, with DS6 displaying the lowest. ZOIs of isolates range between 10.3 mm and 16.3 mm. n=3, with error bars indicating standard deviation between replicates.

5.2.2 16S rRNA gene sequencing confirmed isolates were *Streptomyces*

Phylogenetic analysis and BLAST comparison suggested that DS2, DS4, DS6 and DS7 were the same strain (**Table 5.1**) (**Supplementary Figure 5.3**). DS1 showed high similarity to *Streptomyces thermocarboxydus* and DS5 showed high similarity to *Streptomyces* sp. BV9 (**Table 5.1**). As DS1 and DS5 were morphologically distinct and inhibited *B. subtilis,* the two isolates were taken forward for further analysis. Evolutionarily, DS1 claded with *S. fradiae*

ATCC 10745 and DS5 with *S. kanamyceticus* ATCC 12853, although bootstrap values were low (61 and 57 respectively) (**Figure 5.2**). In general, the 16S rRNA gene for phylogenetic analysis of *Streptomyces* is not very reliable. As such, it was the BLAST analysis and bioactivity that were used for selection purposes.

Table 5.1 – BLAST comparison of 16S rRNA gene sequences of soil isolates. 16S rRNA gene sequencing was performed to confirm the identity of the isolates to genus level. Sequences were obtained with both forward and reverse primers. Sequencing of DS3 failed, and it was excluded from further analysis. Of the six isolates sequenced, only DS1 and DS5 matched at species level with both forward and reverse primer sequences, and as such these strains were taken forward for further analysis. Accession numbers for matched 16S rRNA gene sequences were retrieved from the NCBI database.

Sample ID	BLAST Result	Query Cover (%)	Similarity to Isolate 16S rRNA gene (%)	Accession (NCBI)
1F	*Streptomyces thermocarboxydus* HD07	99	99.34	KT163790.1
1R	*Streptomyces thermocarboxydus* EGI124	97	99.53	MN704433.1
2F	*Streptomyces* sp. HBUM206355	99	99.43	MT540269.1
2R	*Streptomyces fulvissimus* NA06532	98	99.07	CP054926.1
4F	*Streptomyces* sp. HBUM206355	98	99.34	MT540269.1
4R	*Streptomyces fulvissimus* NA06532	98	99.33	CP054926.1
5F	*Streptomyces* sp. BV9	99	99.35	MF511780.1
5R	*Streptomyces* sp. BV9	98	99.53	MF511780.1
6F	*Streptomyces* sp. STR43	99	98.6	KF803388.1
6R	*Streptomyces fulvissimus* NA06532	98	99.63	CP054926.1
7F	*Streptomyces* sp. HBUM206355	96	99.10	MT540269.1
7R	*Streptomyces pratensis* WZS030	97	99.25	MH482911.1

Tree scale: 0.01

MT135558.1 *Kocuria* sp.

Streptomyces nodosus ATCC 14899

69

Streptomyces spectabilis ATCC 27465

42

Streptomyces fradiae ATCC 10745

61

Isolate 1

Streptomyces platensis ATCC 23948

68

Streptomyces rimosus ATCC 10970

35

Streptomyces noursei ATCC 11455

24

Streptomyces clavuligerus ATCC 27064

84

Streptomyces venezuelae ATCC 10712

73

Streptomyces vinaceus ATCC 27476

42

Streptomyces kanamyceticus ATCC 27476

57

Isolate 5

Figure 5.2 – Phylogeny of soil isolates DS1 and DS5 compared to *Streptomyces* type strains. A phylogenetic tree was constructed to 1000 bootstraps, assessing evolutionary relatedness of DS1 and DS5 compared to 10 *Streptomyces* strains previously analysed in this study. 16S rRNA gene sequences for type strains were retrieved from NCBI, with *Kocuria* sp. used as an outgroup.
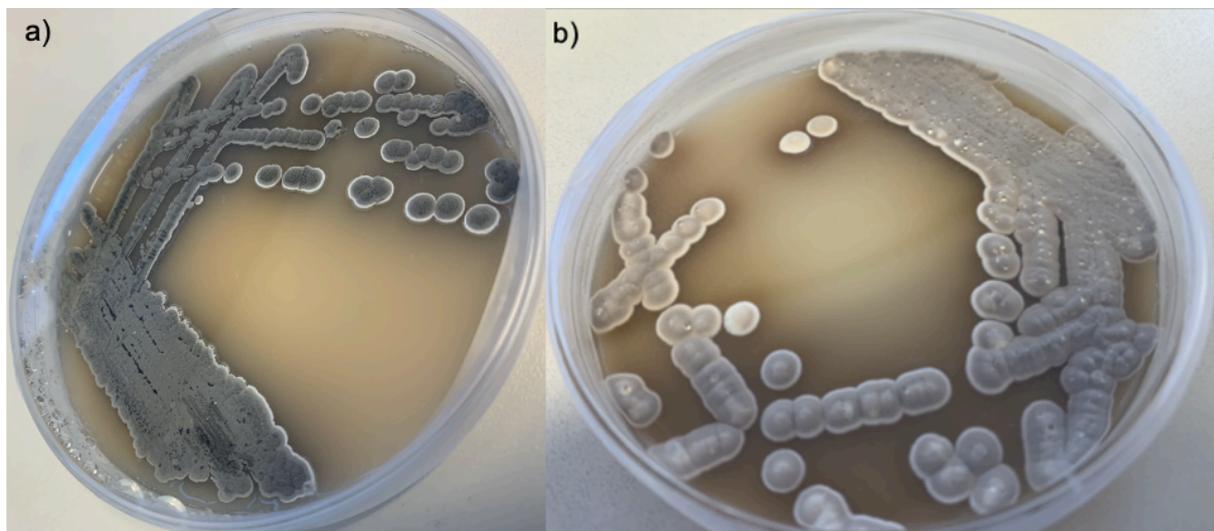
Figure 5.3 – *Streptomyces* isolated from soil. On MS media **a)** DS5 produced dark grey spores and a black pigment which diffused into the media and **b)** DS1 formed light grey spores, with droplets appearing after four days of incubation and large patches of very dark red pigment diffusing into the media.

5.2.3 Light microscopy of Drumpellier park soil revealed features characteristic of Actinomycetota

Light microscopy of DS1 and DS5 revealed characteristics of *Streptomyces* morphology. DS1 had fragmented hyphae and a small number of singular spores could be seen (**Figure 5.4 a**). Most hyphal fragments within the sample appear to have differentiated into polysporous chains (**i**). DS5 (**Figure 5.4 b**) had a greater number of visible components with examples of a full life cycle visible. Many singular spores were observed with several in a state of budding (**ii**), whilst both polysporous chains and undeveloped hyphae (that have not developed beyond a vegetative state) could be seen.
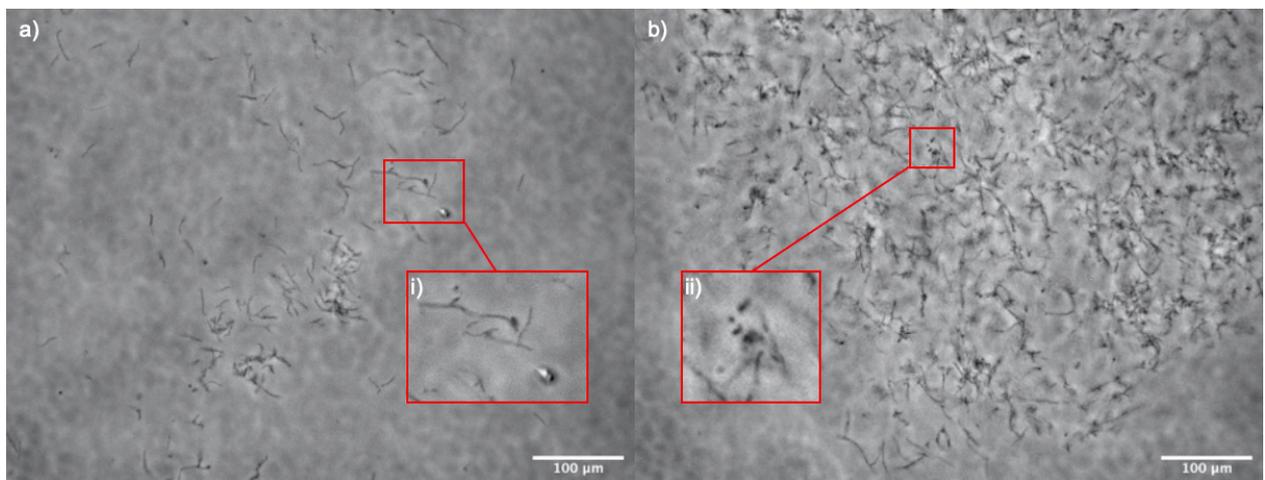
Figure 5.4 – Light microscopy of Drumpellier Park soil DS1 and DS5. To further confirm the isolates belonged to the phylum Actinomycetota, light microscopy was performed and several characteristics of Actinomycetota morphology and life cycle were identified. **A)** shows DS1 with fragmented polysporous hyphae (**i**) and singular spores visible. **B)** DS5 showed spores in various stages of the Actinomycetota life cycle including budding (**ii**).

5.2.4 Soil isolates inhibited *Streptomyces*

S. spectabilis and *S. noursei* were the strains previously shown to be the most bioactive in Chapter 4. Therefore, cross inhibition assays were established with each of the four tested *Streptomyces* strains (*S. spectabilis, S. noursei,* DS1 and DS5 as the centre strain. *S. noursei,* whilst having previously displayed potent inhibitory activity against other streptomycetes*,* did not inhibit DS1 or DS5 at all whilst continuing to show inhibitory behaviour against *S. spectabilis.* Interestingly, a violet/purple pigment which did not diffuse into the media was elicited from both soil isolates in coculture with *S. noursei* (**Figure 5.5 a**), with this pigmentation absent in any other cocultures. Violet/purple pigments such as violacein (as opposed to a mixture of blue and red pigments) have been isolated from rare *Streptomyces* sp. previously but were not detected in this work. Speculatively, this pigment could be the simultaneous production of both actinorhodin and a prodigiosin as can be commonly seen within streptomycetes.

S. spectabilis inhibited both DS1 and DS5, but previously observed inhibition of *S. noursei* was absent (**Figure 5.5 b**), potentially due to a focusing of inhibitory metabolites towards the location of the two soil isolates. Both soil isolates displayed inhibitory behaviour towards other streptomycetes. DS1 inhibited both *S. spectabilis* and DS5 (**Figure 5.5 c**) and DS5 inhibited *S. noursei* as well as DS1 (**Figure 5.5 d**). This behaviour indicated that both DS1 and DS5 were capable of producing antimicrobial specialised metabolites in coculture to a level which interrupted the growth of opposition strains. They therefore would be suitable candidates for comparative metabolomics analysis.
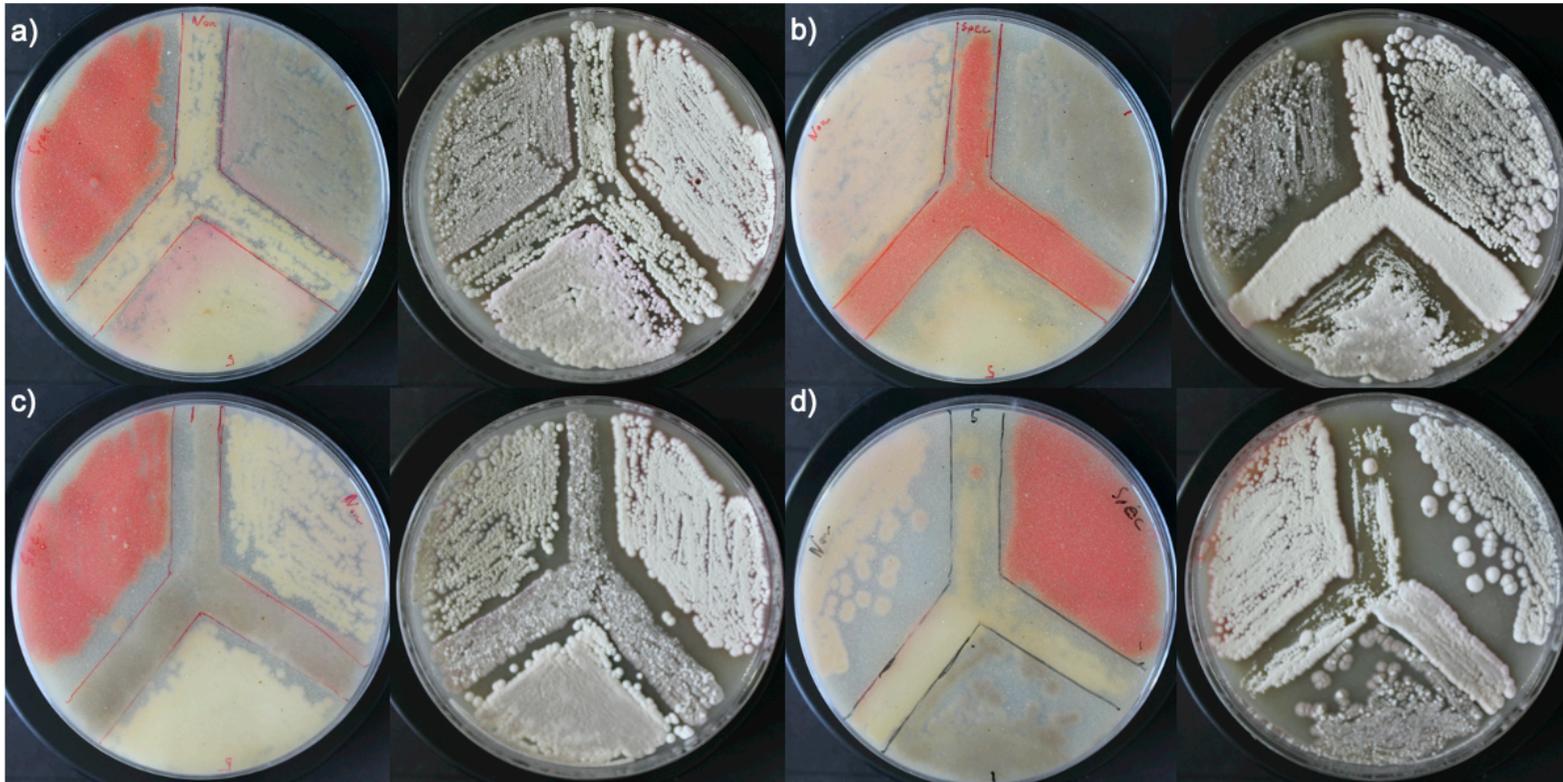
Figure 5.5 – Petri dishes showing the back and front of inhibition coculture assays between _S. spectabilis, S. noursei_ and soil DS1 and DS5. In order to test whether the isolates DS1 and DS5 could inhibit or would be inhibited by other Streptomycetes, interaction assays were set up. DS1 and DS5 were tested against _S. spectabilis_ and _S. noursei,_ the strains which were the most inhibitory towards other streptomycetes earlier in previous chapters of this work. One strain was inoculated to the middle of a plate and incubated (5 days, 30°C), and the three other strains were inoculated in the appropriate sectors around the centre strain before a second incubation (5 days, 30°C). **a)** shows _S. noursei_ as the strain through the centre of the plate, **b)** _S. spectabilis_ can be observed as the centre strain, **c)** shows soil DS1 as the centre strain whilst **d)** shows Isolate 5 as the centre strain.

5.2.5 Comparative metabolomics analysis of natural and artificial soil microcosm systems revealed unique chemistry

When comparing four strains in monoculture against four cocultures across four media conditions in triplicate, 96 bacterial metabolite extracts were generated. These extracts were analysed by LC-MS/MS and a molecular network revealed 1836 parent ion nodes and 2410 edges (**Figure 5.6**). 701 of these nodes related to either solvent or media blanks and therefore were discounted from further analysis. In total, there were 156 molecular families, with each family consisting of structurally related metabolites. The largest of these families contained 84 parent ion associated nodes with all four microcosm conditions represented. The GNPS platform automatically identified and annotated linoleic acid, mucic acid and 13-docosenamide (an amide of docosenoic acid) within this family. 887 nodes were classed as singletons, at 47.6% this was almost half of the entire dataset. These metabolites had no correlation to any other parent ion in the dataset revealing considerable chemical diversity.

In total, 327 parent ions were detected as being shared between groups of microcosm systems, accounting for 28.81% of bacterial metabolites (**Figure 5.7**). 353 parent ion nodes were observed from the microcosm system consisting of phytagel beads equilibrated with liquid 2xYT media, which was the highest in all microcosms (**Supplementary Figure 5.4**). This is perhaps to be expected, as the composition of the media is specifically designed to facilitate the growth of microbes. The natural soil microcosm systems shared 289 parent ions, showing them as viable constructs for future bacterial metabolomics studies. Interestingly, in the natural soil microcosm systems two molecular families, one consisting of 15 nodes with an *m/z* range of between 1148 and 1327 and another consisting of 16 nodes with an *m/z* range of 1335 and 1427, were detected. These molecular families were unannotated and could not be manually linked to any BGC predicted by antiSMASH. They therefore represent targets for computational pattern matching via NP Linker. The artificial soil microcosm systems had 142 specific parent ions, proving them to be a source of unique chemistry. The artificial soil

microcosm system had the highest number of annotated metabolites via GNPS with 19 in total and 11 specifically detected within this particular system (**Table 5.2**). These included linoleic acid and genistein produced by *S. spectabilis*, and 2'-Hydroxy-a-naphthoflavone produced by *S. spectabilis* and *S. noursei* both individually and in coculture with each other. Some unique chemistry was detected within the sand/nutrient solution microcosm systems, albeit at a much lower level than the other three conditions, with 22 specific parent ions, mostly occurring as singletons within the network. The success of both the natural and artificial soil microcosm systems in facilitating the production of unique and diverse chemistry highlights their value as future approaches for the field of natural product chemistry **(Figure 5.6)**. 36 *Streptomyces* metabolites within the microcosm systems matched with the GNPS spectral library (**Table 5.2**). These included undecylprodigiosin, produced by *S. spectabilis* in monoculture as well as in coculture with both *S. noursei* and DS1. The detection of undecylprodigiosin was expected due to the pigmentation previously observed within soil microcosm systems. However, the pigment was only detected within the 2xYT and sand-based microcosm systems. The siderophore desferrioxamine E (also known as nocardamine) and a structurally similar cyclic peptide known as desmethylenylnocardamine were both detected in *S. noursei* in 2xYT media microcosm systems. Cycloheximide (observed *m/z:* 282.172 Da, accurate mass: 281.35 g/mol), a fungicidal polyketide, was detected in *S. noursei* monoculture as well as coculture with DS5 in 2xYT and sand microcosm systems. It was initially expected that antifungal metabolites would be elicited in response to fungal spores present within soil microcosm. It was therefore surprising that an antifungal metabolite was detected in the biologically inert 2xYT and sand microcosm systems but was absent in the soil systems. Known metabolites were detected as being produced only when cultured in non-laboratory standard media conditions. Oleanolic acid (observed *m/z:* 457.358 Da, accurate mass: 456.7 g/mol), which has been reported as having both antitumour and antiviral effects, was produced by *S. spectabilis, S. noursei* and within coculture of both strains in soil microcosm systems. This being the case, it would be reasonable for one to assume that novel antimicrobial metabolites are being missed when strains are cultured exclusively in laboratory standard conditions.
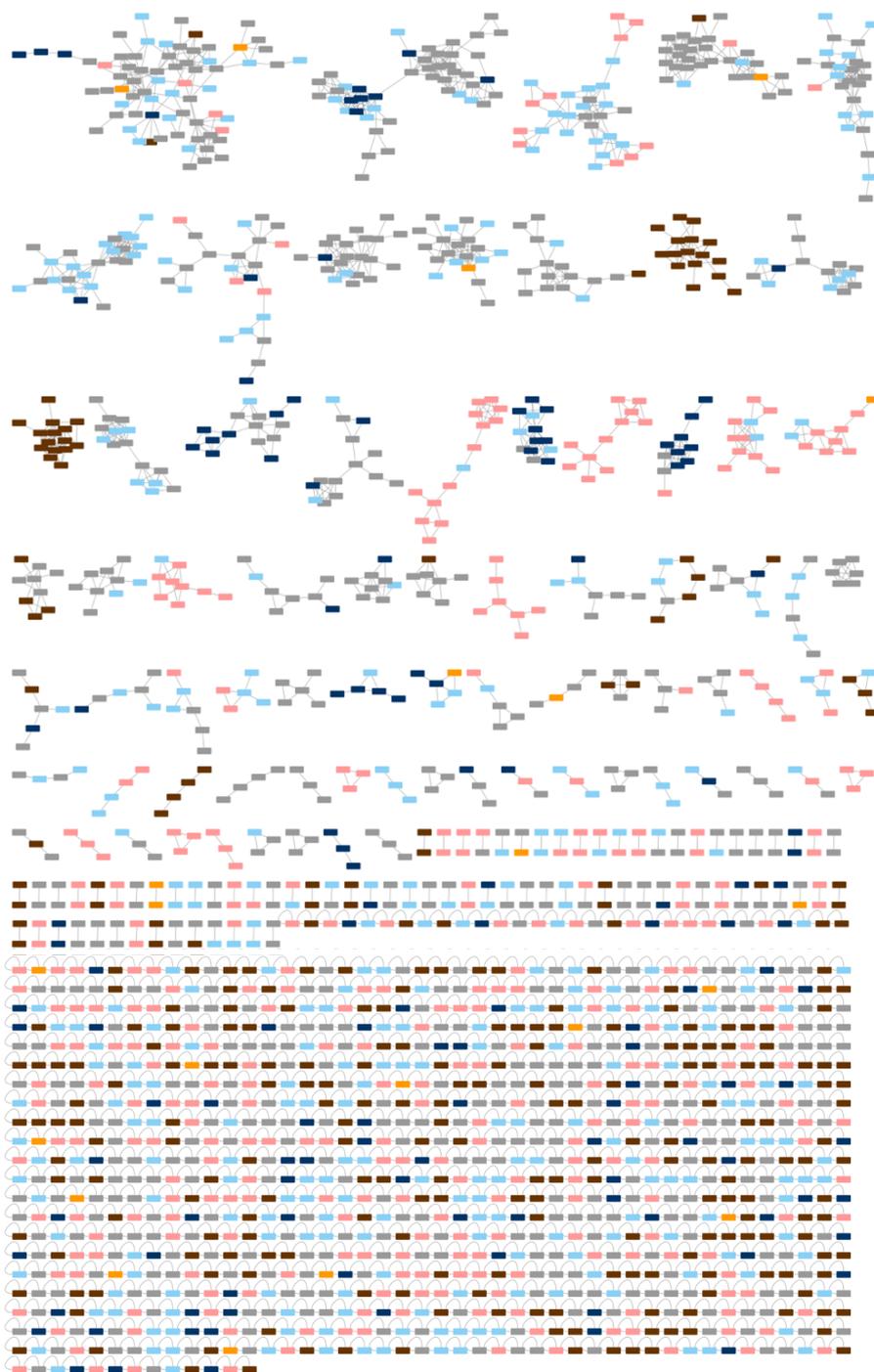
Figure 5.6 – GNPS molecular network of spectra detected within soil, artificial soil, sand and 2xYT microcosm systems. Each microcosm system was freeze dried, and the metabolites were extracted via shaking the dried culture with EtOAc. Each extract was dissolved to 1 mg/mL in acetonitrile and analysed via LC-MS/MS. The resultant data was used to construct a molecular network using the GNPS platform. The network has 1836 nodes and 2410 edges and features 156 molecular families and 887 singletons. It is colour coded by microcosm system type: Brown – soil, orange – sand, navy – artificial soil, pink – 2xYT, blue – shared by multiple microcosm system types, grey – blank.
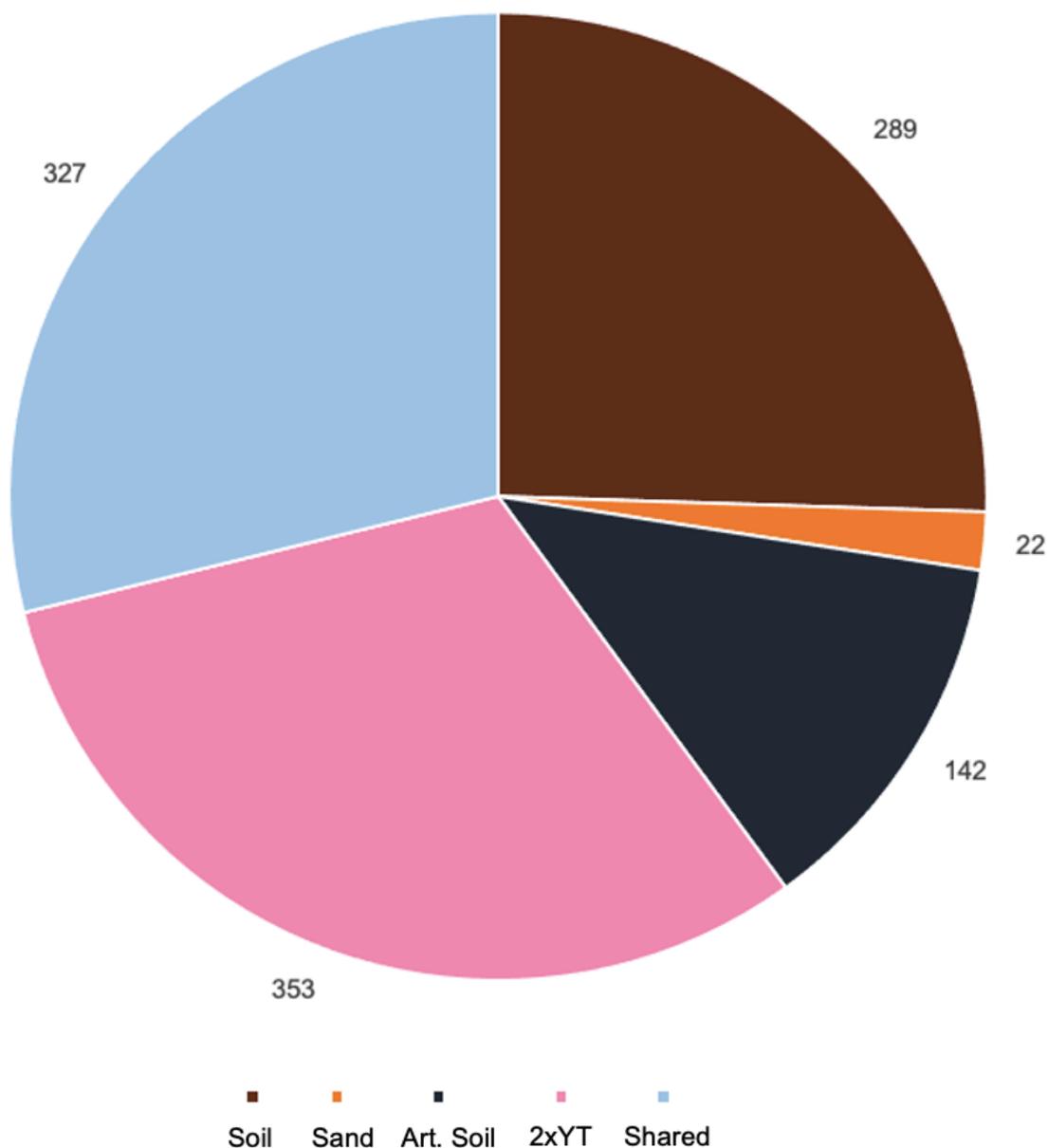
Figure 5.7 – Pie chart displaying the breakdown of parent ions detected within each microcosm system. From the GNPS molecular network in **Figure 5.6**, the number of nodes associated with each type of microcosm system were calculated. Each slice represents a specific microcosm system (soil, sand, artificial soil, 2xYT). The 'Shared' slice represents any parent ion detected in more than one type of microcosm system and solvent/media blanks were excluded from the figure. Of the four types of microcosm systems, culture in 2xYT resulted in the most metabolites specific to the culture condition. This is perhaps to be expected given that this culture medium is specifically designed to facilitate the growth of bacteria. Culture in artificial soil microcosm systems resulted in the detection of 142 metabolites exclusive to this condition, proving to be a valuable source of unique chemistry.

Table 5.2: Metabolites identified via GNPS, plus producer strains and media conditions

'Cultured strain(s)' refers to the strains in the specific culture extract which the metabolite was detected.

A forward slash ('/') between two strain names denotes a coculture. PPM error is relative to the theoretical mass of the precursor.

Metabolite	Cultured strain(s)	Microcosm condition	PPM Error
Linoleic acid	*S. spectabilis*	Artificial soil	0.0
Spectral Match to Val-Leu	*S. noursei, DS1/S. spectabilis,* DS1, DS5/*S. noursei, S. spectabilis/S. noursei,* DS1/DS5	All	4.72
1-methyl-4-methylidene-7-(propan-2-yl)-1,2,3,3a,4,5,6,8a-octahydroazulen-1-ol	*S. noursei, S. spectabilis/S. noursei,* DS1/DS5	2xYT	0.0
Spectral Match to N-Lauroylsarcosine	*S. spectabilis/S. noursei*	Artificial soil	3.59
5'-Deoxy-5'-(methylsulfinyl)adenosine	DS1/*S. spectabilis*	Artificial soil	3.21
Cyclo(L-Phe-D-Pro)	*S. noursei, DS1/S. spectabilis,* DS1, DS5/*S. noursei, S. spectabilis/S. noursei,* DS1/DS5	All	0.0
Spectral Match to N-Acetyl-D-glucosamine	*S. spectabilis/S. noursei*	2xYT	4.93
5'-Deoxyadenosine	*S. spectabilis/S. noursei,* DS1/*S. spectabilis*	2xYT, Artificial soil	0.0
Cycloheximide	*S. noursei,* DS5/*S. noursei*	2xYT, Sand	0.0
Spectral Match to His-Pro	*All*	2xYT	4.28
Spectral Match to Diisodecyl phthalate	*DS1/DS5*	Soil	0.0
Undecylprodigiosin	*S. spectabilis, S. spectabilis/S. noursei,* DS1/*S. spectabilis*	2xYT, Sand	2.55
Oleanolic acid	*S. spectabilis, S. noursei, S. spectabilis/S. noursei*	Soil	2.15
Adenosine (2R,3R,4S,5R)-2-(6-aminopurin-9-yl)-5-(hydroxymethyl)oxolane-3,4-diol	*S. spectabilis, S. noursei, S. spectabilis/S. noursei,* DS1/*S. spectabilis*	Artificial soil, 2xYT, Sand	0.0
Octadeca-9,12,15-trienoic acid	*S. spectabilis, S. noursei*	Artificial soil	0.0

Cyclo(D-Trp-L-Pro)	*S. spectabilis/S. noursei,* DS1/*S. spectabilis*	2xYT, Sand	0.0
Desmethylenylnocardamine	*S. noursei*	2xYT	1.77
Desferrioxamine E	*S. noursei*	2xYT	1.62
Cyclo-(Val-Phe)	*S. noursei, S. spectabilis/S. noursei,* DS1/DS5, DS1/*S. spectabilis,* DS5/*S. noursei*	2xYT, Artificial soil, Sand	4.01
Cycloheximide	*S. noursei,* DS5/*S. noursei*	2xYT, Sand	0.0
4',7-dihydroxyisoflavone|Daidzein|Ddze|7-hydroxy-3-(4-hydroxyphenyl) chromen-4-one	*S. spectabilis*	Artificial soil	0.0
Germicidin A	*S. noursei, S. spectabilis/S. noursei,* DS1/*S. spectabilis,* DS5/*S. noursei*	2xYT	0.0
Spectral Match to 1-Hexadecanoyl-sn-glycerol	*S. spectabilis, S. noursei, S. spectabilis/S. noursei*	Artificial soil, 2xYT	0.0
3-hydroxy-6-[[(E)-3-hydroxy-2,4-dimethylhept-4-enoyl] amino]-2,4-dimethyl-5-oxohexanoic acid	*S. spectabilis/S. noursei,* DS1/*S. spectabilis*	2xYT	2.75

5.2.6 Computational pattern matching confirmed the presence of undecylprodigiosin in both natural and artificial soil microcosm systems

A parent ion with the *m/z* 394.285 Da was detected within both 2xYT and sand systems across *S. spectabilis* monoculture, *S. spectabilis/S. noursei* coculture and DS1/*S. spectabilis* coculture. The ion was annotated by GNPS as undecylprodiginine via spectral similarity to the GNPS spectrum CCMSLIB00000072310. NP Linker, which uses computational pattern matching to infer links between BGCs and produced metabolites, inferred a link between this spectra and "CP023690.1.region031" (**Figure 5.8**), a tag which corresponds to the 31st BGC of the *S. spectabilis* genome. The standardised score for this inferred link was amongst the highest in the dataset (11, with 11 being the highest standardised score, although this score was common throughout the dataset). The mass spectra fragmentation pattern for this

metabolite in GNPS also closely matched predicted fragments of the undecylprodigiosin molecule. It can therefore be concluded that region 31 of the *S. spectabilis* BGC suite, predicted by antiSMASH as undecylprodigiosin, is likely responsible for the production of the linked parent ion (*m/z* 394.285 Da).

Figure 5.8 – NP Linker output showing predicted link between *S. spectabilis* BGC and spectra. Links were inferred between the genomic and metabolomic datasets using the NP Linker platform. Frame a) shows the linked BGC to be region 31 of the *S. spectabilis* genome, identified by antiSMASH as undecylprodigiosin. Frame b) shows details of the linked spectrum (spectrum ID: 59074, MF ID: 70, *m/z*: 394.285 Da) and frame c) shows the GNPS annotation of the linked spectra.

5.2.7 Rosetta-based computational pattern matching of natural and artificial soil microcosm systems yielded no confirmed links

An additional filtering layer was applied to the dataset, combining the standardised strain correlation scoring system (known as 'Metcalf') with the novel Rosetta scoring system. This scoring approach yielded links for four BGCs (**Table 5.3**): Regions 17 and 28 from the *S. spectabilis* BGC genome and regions 17 and 21 of the *S. noursei* genome. Region 17 in both genomes is predicted by antiSMASH to be the ectoine BGC, a specialised metabolite commonly utilised by microorganisms to combat osmotic stress in highly saline environments. The accurate mass of ectoine is 142.158 g/mol^{-1} and due to the soft ionisation method used (Electron Spray Ionisation – ESI) one would expect any linked metabolites to have a mass to charge ratio (*m/z*) similar to this accurate mass. However, none of the three detected metabolites linked to the BGC in the *S. spectabilis* genome and neither of the two metabolites linked to the region in the *S. noursei* genome had an *m/z* corresponding to the accurate mass of ectoine.

This result was the same across the board for Rosetta links. Region 28 of the *S. spectabilis* BGC suite was predicted by antiSMASH to be a match to kanamycin, with none of the four linked spectra matching the accurate mass of 484.499 g/mol^{-1}. Region 21 of the *S. noursei* BGC suite was predicted to match the RiPP legonaridin, which during high resolution election spray ionisation mass spectrometry (HR-ESIMS) was confirmed to have an *m/z* of 901.223 Da, however none of the linked spectra had an *m/z* of more than 314.019 Da.

Table 5.3 – Links predicted between BGCs and detected metabolites by the Rosetta scoring method. Scores are dataset dependent, with the highest BGC score within the dataset being 1.254 and the highest spectral score 0.76305468.

NPLinker BGC ID	BGC ID	MIBiG BGC ID	BGC score	spectrum ID	GNPS ID	spectral score	*m/z* (Da)
35	*S. noursei* region 17	BGC0000860	1.252	2942	CCMSLIB00000565731	0.67704755	191.136
35	*S. noursei* region 17	BGC0000860	1.252	3105	CCMSLIB00000565731	0.51613545	193.083
35	*S. noursei* region 17	BGC0000860	1.252	13161	CCMSLIB00000218171	0.54778882	242.128
40	*S. noursei* region 28	BGC0000703	0.01410811	5419	CCMSLIB00000219050	0.73595385	236.112
40	*S. noursei* region 28	BGC0000703	0.01410811	12999	CCMSLIB00000219052	0.58610102	241.068
40	*S. noursei* region 28	BGC0000703	0.01410811	13180	CCMSLIB00000219046	0.62201772	244.190
40	*S. noursei* region 28	BGC0000703	0.01410811	56993	CCMSLIB00000567313	0.77620131	388.211
43	*S. spectabilis* region 17	BGC0000860	1.254	2942	CCMSLIB00000565731	0.67704755	191.136
43	*S. spectabilis* region 17	BGC0000860	1.254	3105	CCMSLIB00000565731	0.51613545	193.083
72	*S. spectabilis* region 21	BGC0001735	0.2378125	67	CCMSLIB00000221757	0.51839668	170.081
72	*S. spectabilis* region 21	BGC0001735	0.2378125	85	CCMSLIB00000216133	0.56640536	179.045
72	*S. spectabilis* region 21	BGC0001735	0.2378125	87	CCMSLIB00005435936	0.50916573	179.079
72	*S. spectabilis* region 21	BGC0001735	0.2378125	3762	CCMSLIB00000425434	0.51724216	209.059
72	*S. spectabilis* region 21	BGC0001735	0.2378125	3960	CCMSLIB00000221757	0.50249432	211.107
72	*S. spectabilis* region 21	BGC0001735	0.2378125	4429	CCMSLIB00000567253	0.55563824	220.039
72	*S. spectabilis* region 21	BGC0001735	0.2378125	15114	CCMSLIB00000567253	0.7254812	252.109
72	*S. spectabilis* region 21	BGC0001735	0.2378125	15967	CCMSLIB00000216133	0.5292004	261.123
72	*S. spectabilis* region 21	BGC0001735	0.2378125	17071	CCMSLIB00000567253	0.76305468	268.104
72	*S. spectabilis* region 21	BGC0001735	0.2378125	40471	CCMSLIB00000567253	0.52250636	314.019

5.3 Discussion

Actinomycetota were isolated from the soil collected in Chapter 4 and cocultured with sequenced *Streptomyces.* When cocultured on plates with *S. noursei*, a violet/purple pigment was elicited from both DS1 and DS5. The pigment was absent from any negative control and did not diffuse into the media. There are interesting examples of similarly coloured pigments being produced by *Streptomyces* sp. For example, in the Circular Mausoleum tomb at the Roman Necropolis of Carmona in Spain, the walls are covered in a dense colonisation of microbes with violet stains occurring throughout the mortar of the walls. LC-MS analysis of *Streptomyces parvus* isolated from the stains revealed the violet/purple pigmented granaticins (Dominguez-Moñino *et al*., 2017). Interestingly, cocultures of DS5 and *S. noursei* within soil microcosm systems displayed bioactivity against the Gram-positive *B. subtilis* and granaticins have been reported to have antibiotic activity against Gram-positive bacteria (Corbaz *et al*., 1957). Given the scope of Actinomycetota pigments and metabolites in general, it seems more than a coincidence that a lesser seen violet/purple pigment is elicited from two discrete isolates from the same geographic location by *S. noursei*. Whole genome sequencing and genome mining of these two isolates to determine whether either isolate has the granaticin BGC in their genomes is an interesting point of work to be undertaken in future. NP Linker could then be used to link the BGC to the metabolite and confirm its identity.

Using the GNPS platform, a molecular network of 1836 nodes was generated, with 701 of these nodes related to a non-microbial source such as media or solvent blanks. 136 of these nodes were detected exclusively within coculture extracts, indicating that these nodes correspond to metabolites elicited through the coculture of strains. It should be noted that of the 1836 nodes, only 24 have GNPS annotations with only seven annotated metabolites elicited via coculture, and as such there are many more elicited metabolites in the molecular network that lack such annotation which are prime targets for future study. Metabolites such

as genistein and methyl dihydrojasmonate were annotated by GNPS, but with a ppm error of greater than the threshold for MS accuracy, so were excluded from further analysis. These results show that bringing together Actinomycetota isolates in diverse culture conditions is a good source of eliciting unique chemistry. Interestingly, as in previous chapters of this work, elicitation was not limited to coculture.

Metabolites were also elicited in monoculture of strains in artificial soil microcosm systems. One such annotated metabolite, produced by *S. spectabilis*, was linoleic acid. A study previously demonstrated that linoleic acid inhibited the growth of all Gram-positive pathogens it was tested against, including *S. aureus* and *B. subtilis* (Dilika, Bremner, & Meyer, 2000). Genistein was also produced by *S. spectabilis* in the artificial soil microcosm and has been shown to have antibacterial activity against *S. aureus* and *B. anthracis* (Hong, Landauer, Foriska, & Ledney, 2006), as well as being a chemotherapeutic anticancer agent (Spagnuolo *et al*., 2015). These results show that both monoculture and coculture of Actinomycetota in artificial soil microcosm systems can be considered key in unlocking unique and potentially novel antimicrobial metabolites. Of the 142 parent ions detected exclusively in the gel-based artificial soil microcosm system, only 11 were annotated by GNPS and none were linked to a BGC using either computational pattern matching method. Many of these unidentified parent ions were detected from culture extracts which were bioactive against *B. subtilis*. In future, metabolites from artificial soil microcosm systems within the molecular network should be manually dereplicated and investigated further.

It is clear that coculture within artificial soil microcosm systems results in the production of potentially important specialised metabolites. However, only 28.81% of the metabolites detected were shared between the different microcosm systems. This indicates that there is little crossover between the chemistry of natural and artificial soil microcosms systems and more work should be done to develop a reproducible natural soil environment for laboratory analysis. This is perhaps to be expected, as culture conditions are one of the key drivers of

chemical diversity and varying culture media is one of the fundamental parameters of the OSMAC approach (Bode, Bethe, Hofs, & Zeeck, 2002). This is highlighted in a study of *Streptomyces* C34 cultured in eight different media conditions. It was found that altering the composition of the media affected both the quantity and diversity of metabolites produced and resulted in the isolation of novel specialised metabolites (Mostafa E. Rateb *et al*., 2011). Furthermore, a recent study cultured *Streptomyces* CPCC 200267 across eleven culture media found that HPLC profiles differed greatly between culture media (Fang *et al*., 2023). As such, the failure to replicate the chemistry of natural soil microcosm systems was not a disappointment. Artificial soil microcosm systems were the source of unique and potentially novel chemistry and should be considered an option when attempting to broaden produced chemistry from *Streptomyces* in drug discovery projects.

In conclusion, an artificial soil microcosm system was established which elicited specialised metabolites from *Streptomyces,* in both monoculture and coculture. Actinomycetota were isolated from soil, which when cocultured in the artificial soil microcosm systems, produced unique specialised metabolites. In future, their genomes should be sequenced so these metabolites can be identified and studied further. Initially it was hypothesised that creating an artificial soil based on the nutrient composition of natural soil would replicate chemistry, however this was not the case. Instead, there was not a high percentage of crossover in terms of chemistry and a completely different set of metabolites were detected.

Chapter 6 – General summary, future work, and conclusions

6.1 General summary and discussion

In 1940, a bacterial enzyme was identified which was capable of destroying penicillin (Abraham & Chain, 1940). This discovery actually predated both the awarding of the Nobel Peace Prize in Medicine and Physiology for its discovery and the year it became available over the counter for the first time in the United States by five years (Gaynes, 2017). In short, Antimicrobial Resistance (AMR) is a phenomenon that has long plagued the field of natural product drug discovery. To attempt to overcome come this, it is imperative that the natural product discovery field is shunted forward by the continued advancement of microbial culturing methods and analytical tools (Atanasov *et al*., 2021). This work contributes towards the rejuvenation of natural product drug discovery by describing new methods for eliciting potentially novel antimicrobial specialised metabolites, as well as outlining metabologenomic methods for analysing the resultant datasets.

Metabologenomics is a term coined to describe the field whereby complex genomic and metabolomic datasets are co-analysed to link BGCs to the metabolites they encode. One of the initial hypotheses for this work was that the emerging field of metabologenomics would help accelerate the discovery of novel antimicrobial specialised metabolites from Actinomycetota. Some early studies utilising metabologenomic methods resulted in the discovery of two novel antimicrobials from *S. rimosus* NRRL B-2659, tambromycin and tyrobetaine (Goering *et al*., 2016) (Parkinson *et al*., 2018). A similar pipeline was used to link stravidin to its BGC (Montaser & Kelleher, 2020) with another recent study successfully applying a metabologenomic pipeline to link fungal specialised metabolites to their corresponding BGCs (Caesar *et al*., 2023). The platform NP Linker was utilised in this study for metabologenomic analyses, which built on a previously established pattern matching

method (Doroghazi *et al*., 2014) by implementing standardisation and multi-level scoring for more accurate prediction of links (Hjörleifsson Eldjárn *et al*., 2021). The NP Linker platform was validated by linking actinorhodin, undecylprodigiosin and coelimycin P1 to their BGCs, which was reinforced by mining the raw data to confirm the linked metabolites were identified correctly. 10 more known metabolites were linked to their BGCs using this method however, the results were limited by a short timeframe and small dataset.

The reasons for the slowing of the natural product discovery pipeline are well documented (Zorzet, 2014). Most BGCs are silent or cryptic when bacterial strains are cultured in standard laboratory conditions, where bacteria are traditionally grown in isolation on solid or liquid media (Gupta *et al*., 2017). OSMAC approaches have yielded exciting results, but it is clear that work still has to be done in expanding culture conditions to maximise produced chemistry (Pan, Bai, Chen, Zhang, & Wang, 2019). Taking this into consideration, a decision was made to cease the use of commonly used culture media and pivot the study into a direction more aligned with *Streptomyces* ecology. Previous studies have shown that when cultured in a soil microcosm system, *Streptomyces* survive and are metabolically active (Wellington *et al*., 1990) (Schlatter *et al*., 2009). Therefore, a model natural soil microcosm system replaced standard culture media in this study, and it was hypothesised that this would more closely resemble a natural environment and elicit novel specialised metabolites. It was also noted that in nature, *Streptomyces* exist in a complex microbial community. Coculture has previously been utilised to elicit the production of specialised metabolites from *Streptomyces* (Zarins-Tutt *et al*., 2016) (Baral *et al*., 2018) and this technique was implemented to prompt more elicitation and further imitate a natural soil environment. It was found that in coculture did indeed elicit the production of bioactive specialised metabolites, including blasticidin S, a peptidyl nucleoside which inhibits *E. coli* (Davison *et al*., 2017). Metabologenomic analysis also linked a putative siderophore to its BGC. This work reiterated the need to move away from standard laboratory conditions and confirmed that both model soil microcosm systems and metabologenomic tools

such as NP Linker are potentially vital for the future discovery of bioactive specialised metabolites.

One of the main issues with implementing natural soil microcosm systems is the lack of reproducibility between soil samples. Conditions in nature are volatile and fickle, with daily fluctuations in rainfall and pH affecting the nutrient environment (Ulén, 2020) (Feng *et al*., 2023). The results gathered in Chapter 4 of this study showed that culture in natural soil was a viable way to diversify chemistry produced by *Streptomyces,* but the issue of reproducibility must be overcome if this culture method were to be utilised as a long-term option. As such, a nutrient solution was created based on the composition of the soil. Furthermore, this solution was combined with an artificial soil construct to replicate conditions within a natural soil microcosm system. It was therefore hypothesised that this artificial soil microcosm system could be implemented in long-term drug discovery studies to elicit the production of potentially novel specialised metabolites. Interestingly, only around a quarter of the parent ions detected were shared between the natural and artificial soil microcosms, indicating that the artificial soil does not replicate the conditions in a natural soil environment. This factor is up for debate however, as variability of antibiotic production in *Streptomyces* culture is often high (S. Siebenberg, P. M. Bapat, A. E. Lantz, B. Gust, & L. Heide, 2010). Whilst reproducibility was not achieved in the study, the artificial soil microcosm systems were the source of unique and diverse chemistry and. Combined with microbial coculture, these systems can be utilised in future as an alternative to standard laboratory culture conditions for the elicitation of specialised metabolites.

6.2 Future work

One of the initial focuses of the study was to utilise and in turn, validate the NP Linker platform (Hjörleifsson Eldjárn *et al*., 2021). Whilst the analyses were successful, more accurate and comprehensive links would be obtained with larger genomic and metabolomic datasets. In the

initial validation experiments for NP Linker in this work, 15 Actinomycetota strain genomes and metabolomes were analysed. However, in larger studies which involve machine learning tools there are often hundreds and, in some cases, even hundreds of thousands of strains involved. For instance, the study in which the term 'metabologenomics' was coined analysed the genomes and metabolomes of 178 Actinomycetota species (Goering *et al*., 2016). At the larger end of the scale, the recent compendium of specialised metabolite biosynthetic diversity involved the computational analysis of around 170,000 bacterial genomes and around 47,000 metagenome assembled genomes (Gavriilidou *et al*., 2022). Indeed, whilst these numbers are beyond the scope of this work given the restrains on time and space, future studies involving should look to expand the data analysed to increase the possibility of finding valid links.

Clearly, one of the next steps in this work should be to apply the pipelines developed to the discovery and isolation of novel bioactive specialised metabolites. If NP Linker strongly inferred a link between a metabolite with a match within the GNPS spectral library and a BGC with no predicted identity in antiSMASH, this link would be an ideal candidate for further investigation. The BGC could then be heterologously expressed in a host organism for increased productivity (Malpartida & Hopwood, 1984). Many strains of *Streptomyces* have been adopted as heterologous hosts, including *S. coelicolor* CH999 (R. McDaniel, Ebert-Khosla, Hopwood, & Khosla, 1993) and *S. lividans* TK24 (Ahmed *et al*., 2020). Recently, the BGC of a novel polyene macrolactam from *Streptomyces rochei* IFO12908 was heterologously expressed in a *Streptomyces avermitilis* derivative leading to the isolation, structure determination and bioactivity analysis of the metabolite (Hashimoto *et al*., 2020). Such methods would be an ideal approach for the continuation of this work.

In Chapter 4 of this study, one of the aims was to ecologically align conditions within the soil microcosm systems to the soil environment in which *Streptomyces* exist in nature by utilising coculture. However, this work was limited to the coculture of two Actinomycetota at any one time, whereas in nature bacteria exist in phenomenal diversity. It has been estimated that a

single gram of soil harbours between 4×10^3 and 5×10^4 bacterial species (Raynaud & Nunan, 2014). There can be no doubt that this scale is not reproducible in a laboratory setting, but when investigating coculture for the elicitation of specialised metabolites in future, more diverse microbial communities should be utilised. Cross-kingdom coculture of Actinomycetota and fungi have previously proved fruitful. *Streptomyces rochei* MB037 with *Rhinocladiella similis* 35 resulted in the elicitation of two metabolites that exhibited potent bioactivity against MRSA (M. Yu *et al.*, 2019) and in a study linking a bioactivity-targeted approach coupled with coculture, antibacterial activity against *B. subtilis* was induced in two *Streptomyces* isolates from soil when cocultured with the fungus *Schizophyllum commune* (Nicault *et al.*, 2021). Coculture conditions such as these should be applied to natural and artificial soil microcosm systems for greater elicitation of specialised metabolites.

6.3 Conclusions

In summary, NP Linker accurately inferred links between 13 known microbial metabolites and their corresponding BGCs. These links were validated by mining raw LC-MS/MS data and identifying the fragmentation patterns of each metabolite. Furthermore, an assay was developed whereby high enough crude metabolite yields for high quality LC-MS/MS analysis could be generated at a miniature scale. In these cases, the hypotheses were confirmed. Following these results, bioactive specialised metabolites were elicited from *Streptomyces* in coculture in model soil microcosm systems. This was again confirmed by NP Linker accurately predicting links which were validated by analysis of metabolite fragmentation patterns. Finally, replication of metabolite production from *Streptomyces* in soil microcosm systems was not achieved within artificial soil microcosm systems however, unique and diverse chemistry was observed as well as the elicitation of bioactive specialised metabolites. The work encompassed in this study provides a baseline for many future studies into the acceleration and rejuvenation of the field of antibiotic discovery from *Streptomyces.*

Bibliography

1. Abraham, E. P., & Chain, E. (1940). An enzyme from bacteria able to destroy penicillin. 1988. *Rev Infect Dis, 10*(4), 677-678.

2. Adnani, N., Vazquez-Rivera, E., Adibhatla, S. N., Ellis, G. A., Braun, D. R., & Bugni, T. S. (2015). Investigation of Interspecies Interactions within Marine Micromonosporaceae Using an Improved Co-Culture Approach. *Mar Drugs, 13*(10), 6082-6098. doi:10.3390/md13106082

3. 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), 5. doi:10.1186/s12934-020-1277-8

4. Akbar, N., Kawish, M., Khan, N. A., Shah, M. R., Alharbi, A. M., Alfahemi, H., & Siddiqui, R. (2022). Hesperidin-, Curcumin-, and Amphotericin B- Based Nano-Formulations as Potential Antibacterials. *Antibiotics (Basel), 11*(5). doi:10.3390/antibiotics11050696

5. Alam, K., Mazumder, A., Sikdar, S., Zhao, Y.-M., Hao, J., Song, C., . . . Li, A. (2022). Streptomyces: The biofactory of secondary metabolites. *Frontiers in Microbiology, 13*. doi:10.3389/fmicb.2022.968053

6. Alanjary, M., Steinke, K., & Ziemert, N. (2019). AutoMLST: an automated web server for generating multi-locus species trees highlighting natural product potential. *Nucleic Acids Research, 47*(W1), W276-W282. doi:10.1093/nar/gkz282

7. Alarcon-Barrera, J. C., Kostidis, S., Ondo-Mendez, A., & Giera, M. (2022). Recent advances in metabolomics analysis for early drug development. *Drug Discovery Today, 27*(6), 1763-1773. doi:https://doi.org/10.1016/j.drudis.2022.02.018

8. Álvarez-Álvarez, R., Botas, A., Albillos, S. M., Rumbero, A., Martín, J. F., & Liras, P. (2015). Molecular genetics of naringeninbiosynthesis, a typical plant secondary metabolite produced by Streptomyces clavuligerus. *Microbial Cell Factories, 14*(1), 178. doi:10.1186/s12934-015-0373-7

9. Amin, D. H., Abdallah, N. A., Abolmaaty, A., Tolba, S., & Wellington, E. M. H. (2020). Microbiological and molecular insights on rare Actinobacteria harboring bioactive prospective. *Bulletin of the National Research Centre, 44*(1), 5. doi:10.1186/s42269-019-0266-8

10. Amiri Moghaddam, J., Crüsemann, M., Alanjary, M., Harms, H., Dávila-Céspedes, A., Blom, J., . . . Schäberle, T. F. (2018). Analysis of the Genome and Metabolome of

Marine Myxobacteria Reveals High Potential for Biosynthesis of Novel Specialized Metabolites. *Scientific Reports, 8*(1), 16600. doi:10.1038/s41598-018-34954-y

11. Anand, S., Prasad, M. V., Yadav, G., Kumar, N., Shehara, J., Ansari, M. Z., & Mohanty, D. (2010). SBSPKS: structure based sequence analysis of polyketide synthases. *Nucleic Acids Res, 38*(Web Server issue), 487-496. doi:10.1093/nar/gkq340

12. Antoraz, S., Santamaría, R. I., Díaz, M., Sanz, D., & Rodríguez, H. (2015). Toward a new focus in antibiotic and drug discovery from the Streptomyces arsenal. *Front Microbiol, 6*, 461. doi:10.3389/fmicb.2015.00461

13. Arnison, P. G., Bibb, M. J., Bierbaum, G., Bowers, A. A., Bugni, T. S., Bulaj, G., . . . van der Donk, W. A. (2013). Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat Prod Rep, 30*(1), 108-160. doi:10.1039/c2np20085f

14. Aston, F. W. (1920). Isotopes and atomic weights. In: Nature Publishing Group UK London.

15. Atanasov, A. G., Zotchev, S. B., Dirsch, V. M., Orhan, I. E., Banach, M., Rollinger, J. M., . . . the International Natural Product Sciences, T. (2021). Natural products in drug discovery: advances and opportunities. *Nature Reviews Drug Discovery, 20*(3), 200-216. doi:10.1038/s41573-020-00114-z

16. Avillan, J. J., Ahmadvand, P., Lu, S.-Y., Horton, J., Liu, J., Lofgren, E., . . . Call, D. R. (2022). Excreted Antibiotics May Be Key to Emergence of Increasingly Efficient Antibiotic Resistance in Food Animal Production. *Applied and Environmental Microbiology, 88*(15), e00791-00722. doi:doi:10.1128/aem.00791-22

17. Baltz, R. H. (2017). Gifted microbes for genome mining and natural product discovery. *J Ind Microbiol Biotechnol, 44*(4-5), 573-588. doi:10.1007/s10295-016-1815-x

18. Banerjee, S., & Mazumdar, S. (2012). Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte. *Int J Anal Chem, 2012*, 282574. doi:10.1155/2012/282574

19. Baral, B., Akhgari, A., & Metsa-Ketela, M. (2018). Activation of microbial secondary metabolic pathways: Avenues and challenges. *Synth Syst Biotechnol, 3*(3), 163-178. doi:10.1016/j.synbio.2018.09.001

20. Barka, E. A., Vatsa, P., Sanchez, L., Gaveau-Vaillant, N., Jacquard, C., Meier-Kolthoff, J. P., . . . van Wezel, G. P. (2016). Taxonomy, Physiology, and Natural Products of Actinobacteria. *Microbiol Mol Biol Rev, 80*(1), 1-43. doi:10.1128/MMBR.00019-15

21. Barker, A. K., Brown, K., Ahsan, M., Sengupta, S., & Safdar, N. (2017). Social determinants of antibiotic misuse: a qualitative study of community members in Haryana, India. *BMC Public Health, 17*(1), 333. doi:10.1186/s12889-017-4261-4

22. 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 coelicolor M145. *Journal of the American Chemical Society, 126*(50), 16282-16283. doi:10.1021/ja045774k

23. Bassel, G. W., Glaab, E., Marquez, J., Holdsworth, M. J., & Bacardit, J. (2011). Functional network construction in Arabidopsis using rule-based machine learning on large-scale data sets. *Plant Cell, 23*(9), 3101-3116. doi:10.1105/tpc.111.088153

24. Becher, P. G., Verschut, V., Bibb, M. J., Bush, M. J., Molnár, B. P., Barane, E., . . . Flärdh, K. (2020). Developmentally regulated volatiles geosmin and 2-methylisoborneol attract a soil arthropod to Streptomyces bacteria promoting spore dispersal. *Nature Microbiology, 5*(6), 821-829. doi:10.1038/s41564-020-0697-x

25. Behie, S. W., Bonet, B., Zacharia, V. M., McClung, D. J., & Traxler, M. F. (2017). Molecules to Ecosystems: Actinomycete Natural Products In situ. *Frontiers in Microbiology, 7*. doi:10.3389/fmicb.2016.02149

26. Belknap, K. C., Park, C. J., Barth, B. M., & Andam, C. P. (2020). Genome mining of biosynthetic and chemotherapeutic gene clusters in Streptomyces bacteria. *Scientific Reports, 10*(1), 2003. doi:10.1038/s41598-020-58904-9

27. Bentley, S. D., Chater, K. F., Cerdeño-Tárraga, A. M., Challis, G. L., Thomson, N. R., James, K. D., . . . Hopwood, D. A. (2002). Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). *Nature, 417*(6885), 141-147. doi:10.1038/417141a

28. Bérdy, J. (2012). Thoughts and facts about antibiotics: Where we are now and where we are heading. *The Journal of Antibiotics, 65*(8), 385-395. doi:10.1038/ja.2012.27

29. Berger, H., Bacher, M., Labuda, R., Eppel, I. M., Bayer, F., Sulyok, M., . . . Strauss, J. (2022). Polaramycin B, and not physical interaction, is the signal that rewires fungal metabolism in the Streptomyces–Aspergillus interaction. *Environmental Microbiology, 24*(10), 4899-4914. doi:https://doi.org/10.1111/1462-2920.16118

30. Bisang, C., Long, P. F., Corte´s, J., Westcott, J., Crosby, J., Matharu, A.-L., . . . Leadlay, P. F. (1999). A chain initiation factor common to both modular and aromatic polyketide synthases. *Nature, 401*(6752), 502-505. doi:10.1038/46829

31. Blin, K., Shaw, S., Kloosterman, A. M., Charlop-Powers, Z., van Wezel, G. P., Medema, M. H., & Weber, T. (2021). antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res, 49*(W1), W29-W35. doi:10.1093/nar/gkab335

32. Blin, K., Wolf, T., Chevrette, M. G., Lu, X., Schwalen, C. J., Kautsar, S. A., . . . Medema, M. H. (2017). antiSMASH 4.0—improvements in chemistry prediction and gene cluster

boundary identification. *Nucleic Acids Research, 45*(W1), W36-W41. doi:10.1093/nar/gkx319

33. Bobek, J., Smidova, K., & Cihak, M. (2017). A Waking Review: Old and Novel Insights into the Spore Germination in Streptomyces. *Front Microbiol, 8*, 2205. doi:10.3389/fmicb.2017.02205

34. Bode, H. B., Bethe, B., Hofs, R., & Zeeck, A. (2002). Big effects from small changes: possible ways to explore nature's chemical diversity. *Chembiochem, 3*(7), 619-627. doi:10.1002/1439-7633(20020703)3:7<619::AID-CBIC619>3.0.CO;2-9

35. Bode, H. B., Bethe, B., Höfs, R., & Zeeck, A. (2002). Big effects from small changes: possible ways to explore nature's chemical diversity. *Chembiochem, 3*(7), 619-627. doi:10.1002/1439-7633(20020703)3:7<619::Aid-cbic619>3.0.Co;2-9

36. Bose, U., Hewavitharana, A. K., Ng, Y. K., Shaw, P. N., Fuerst, J. A., & Hodson, M. P. (2015). LC-MS-based metabolomics study of marine bacterial secondary metabolite and antibiotic production in Salinispora arenicola. *Mar Drugs, 13*(1), 249-266. doi:10.3390/md13010249

37. Boucher, H. W., Talbot, G. H., Bradley, J. S., Edwards, J. E., Gilbert, D., Rice, L. B., . . . Bartlett, J. (2009). Bad bugs, no drugs: no ESKAPE! An update from the Infectious Diseases Society of America. *Clin Infect Dis, 48*(1), 1-12. doi:10.1086/595011

38. Bratchkova, A., & Ivanova, V. (2017). ISOLATION, IDENTIFICATION AND BIOLOGICAL ACTIVITIES OF CYCLIC DIPEPTIDES FROM ANTARCTIC MICROORGANISMS. *Comptes rendus de l'AcadÈmie bulgare des Sciences, 70*, 1389+. Retrieved from https://link.gale.com/apps/doc/A570688577/AONE?u=anon~446f77c2&sid=googleScholar&xid=8a71cf63

39. Broadhurst, R. W., Nietlispach, D., Wheatcroft, M. P., Leadlay, P. F., & Weissman, K. J. (2003). The structure of docking domains in modular polyketide synthases. *Chem Biol, 10*(8), 723-731. doi:10.1016/s1074-5521(03)00156-x

40. Büchs, J. (2001). Introduction to advantages and problems of shaken cultures. *Biochemical Engineering Journal, 7*(2), 91-98.

41. Bursy, J., Kuhlmann, A. U., Pittelkow, M., Hartmann, H., Jebbar, M., Pierik, A. J., & Bremer, E. (2008). Synthesis and uptake of the compatible solutes ectoine and 5-hydroxyectoine by Streptomyces coelicolor A3(2) in response to salt and heat stresses. *Appl Environ Microbiol, 74*(23), 7286-7296. doi:10.1128/aem.00768-08

42. Caesar, L. K., Butun, F. A., Robey, M. T., Ayon, N. J., Gupta, R., Dainko, D., . . . Kelleher, N. L. (2023). Correlative metabologenomics of 110 fungi reveals metabolite-gene cluster pairs. *Nat Chem Biol*. doi:10.1038/s41589-023-01276-8

43. Caicedo-Montoya, C., Manzo-Ruiz, M., & Ríos-Estepa, R. (2021). Pan-Genome of the Genus Streptomyces and Prioritization of Biosynthetic Gene Clusters With Potential to Produce Antibiotic Compounds. *Frontiers in Microbiology, 12*. doi:10.3389/fmicb.2021.677558

44. Caicedo-Montoya, C., Manzo-Ruiz, M., & Ríos-Estepa, R. (2021). Pan-Genome of the Genus Streptomyces and Prioritization of Biosynthetic Gene Clusters With Potential to Produce Antibiotic Compounds. *Front Microbiol, 12*, 677558. doi:10.3389/fmicb.2021.677558

45. Castro, J. F., Razmilic, V., Gomez-Escribano, J. P., Andrews, B., Asenjo, J. A., & Bibb, M. J. (2015). Identification and Heterologous Expression of the Chaxamycin Biosynthesis Gene Cluster from Streptomyces leeuwenhoekii. *Appl Environ Microbiol, 81*(17), 5820-5831. doi:10.1128/aem.01039-15

46. Chang, S., Sievert, D. M., Hageman, J. C., Boulton, M. L., Tenover, F. C., Downes, F. P., . . . Fridkin, S. K. (2003). Infection with vancomycin-resistant Staphylococcus aureus containing the vanA resistance gene. *N Engl J Med, 348*(14), 1342-1347. doi:10.1056/NEJMoa025025

47. Chater, K. F. (2016). Recent advances in understanding Streptomyces. *F1000Res, 5*, 2795. doi:10.12688/f1000research.9534.1

48. Chater, K. F., & Chandra, G. (2006). The evolution of development in Streptomyces analysed by genome comparisons. *FEMS Microbiol Rev, 30*(5), 651-672. doi:10.1111/j.1574-6976.2006.00033.x

49. Chaudhary, A. K., Singh, B., Maharjan, S., Jha, A. K., Kim, B. G., & Sohng, J. K. (2014). Switching antibiotics production on and off in actinomycetes by an IclR family transcriptional regulator from Streptomyces peucetius ATCC 27952. *J Microbiol Biotechnol, 24*(8), 1065-1072. doi:10.4014/jmb.1403.03026

50. Chen, H., Cui, J., Wang, P., Wang, X., & Wen, J. (2020). Enhancement of bleomycin production in Streptomyces verticillus through global metabolic regulation of N-acetylglucosamine and assisted metabolic profiling analysis. *Microb Cell Fact, 19*(1), 32. doi:10.1186/s12934-020-01301-8

51. Chen, H., & Du, L. (2016). Iterative polyketide biosynthesis by modular polyketide synthases in bacteria. *Appl Microbiol Biotechnol, 100*(2), 541-557. doi:10.1007/s00253-015-7093-0

52. Chen, R., Wong, H. L., & Burns, B. P. (2019). New Approaches to Detect Biosynthetic Gene Clusters in the Environment. *Medicines (Basel), 6*(1). doi:10.3390/medicines6010032

53. Cheng, Y. Q., Coughlin, J. M., Lim, S. K., & Shen, B. (2009). Type I polyketide synthases that require discrete acyltransferases. *Methods Enzymol, 459*, 165-186. doi:10.1016/s0076-6879(09)04608-4

54. Chevrette, M. G., Gutierrez-Garcia, K., Selem-Mojica, N., Aguilar-Martinez, C., Yanez-Olvera, A., Ramos-Aboites, H. E., . . . Barona-Gomez, F. (2020). Evolutionary dynamics of natural product biosynthesis in bacteria. *Nat Prod Rep, 37*(4), 566-599. doi:10.1039/c9np00048h

55. Chung, Y. H., Kim, H., Ji, C. H., Je, H. W., Lee, D., Shim, S. H., . . . Kang, H. S. (2021). Comparative Genomics Reveals a Remarkable Biosynthetic Potential of the Streptomyces Phylogenetic Lineage Associated with Rugose-Ornamented Spores. *mSystems, 6*(4), e0048921. doi:10.1128/mSystems.00489-21

56. Čihák, M., Kameník, Z., Šmídová, K., Bergman, N., Benada, O., Kofroňová, O., . . . Bobek, J. (2017). Secondary Metabolites Produced during the Germination of Streptomyces coelicolor. *Frontiers in Microbiology, 8*. doi:10.3389/fmicb.2017.02495

57. Cimermancic, P., Medema, M. H., Claesen, J., Kurita, K., Wieland Brown, L. C., Mavrommatis, K., . . . Fischbach, M. A. (2014). Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell, 158*(2), 412-421. doi:10.1016/j.cell.2014.06.034

58. Cimmino, A., Andolfi, A., & Evidente, A. (2014). Phytotoxic Terpenes Produced by Phytopathogenic Fungi and Allelopathic Plants. *Natural Product Communications, 9*(3), 1934578X1400900330. doi:10.1177/1934578x1400900330

59. Claydon, M. A., Davey, S. N., Edwards-Jones, V., & Gordon, D. B. (1996). The rapid identification of intact microorganisms using mass spectrometry. *Nat Biotechnol, 14*(11), 1584-1586. doi:10.1038/nbt1196-1584

60. Clish, C. B. (2015). Metabolomics: an emerging but powerful tool for precision medicine. *Cold Spring Harb Mol Case Stud, 1*(1), a000588. doi:10.1101/mcs.a000588

61. Coates, A. R., Halls, G., & Hu, Y. (2011). Novel classes of antibiotics or more of the same? *Br J Pharmacol, 163*(1), 184-194. doi:10.1111/j.1476-5381.2011.01250.x

62. Corbaz, R., Ettlinger, L., Gäumann, E., Kalvoda, J., Keller-Schierlein, W., Kradolfer, F., . . . Zähner, H. (1957). Stoffwechselprodukte von Actinomyceten. 9. Mitteilung. Granaticin. *Helvetica Chimica Acta, 40*(5), 1262-1269. doi:https://doi.org/10.1002/hlca.19570400518

63. Covington, B. C., McLean, J. A., & Bachmann, B. O. (2017). Comparative mass spectrometry-based metabolomics strategies for the investigation of microbial secondary metabolites. *Nat Prod Rep, 34*(1), 6-24. doi:10.1039/c6np00048g

64. Cuadrat, R. R. C., Ionescu, D., Davila, A. M. R., & Grossart, H. P. (2018). Recovering Genomics Clusters of Secondary Metabolites from Lakes Using Genome-Resolved Metagenomics. *Front Microbiol, 9*, 251. doi:10.3389/fmicb.2018.00251

65. Currie, C. R., Scott, J. A., Summerbell, R. C., & Malloch, D. (1999). Fungus-growing ants use antibiotic-producing bacteria to control garden parasites. *Nature, 398*(6729), 701-704.

66. da Silva, K. E., Tanmoy, A. M., Pragasam, A. K., Iqbal, J., Sajib, M. S. I., Mutreja, A., . . . Andrews, J. R. (2022). The international and intercontinental spread and expansion of antimicrobial-resistant Salmonella Typhi: a genomic epidemiology study. *Lancet Microbe, 3*(8), e567-e577. doi:10.1016/S2666-5247(22)00093-3

67. Das, S., Lecours Boucher, X., Rogers, C., Makowski, C., Chouinard-Decorte, F., Oros Klein, K., . . . Evans, A. C. (2018). Integration of "omics" Data and Phenotypic Data Within a Unified Extensible Multimodal Framework. *Front Neuroinform, 12*, 91. doi:10.3389/fninf.2018.00091

68. Davies, J. (2006). Where have All the Antibiotics Gone? *The Canadian journal of infectious diseases & medical microbiology = Journal canadien des maladies infectieuses et de la microbiologie medicale, 17*(5), 287-290. doi:10.1155/2006/707296

69. Davies, J., & Davies, D. (2010). Origins and evolution of antibiotic resistance. *Microbiol Mol Biol Rev, 74*(3), 417-433. doi:10.1128/MMBR.00016-10

70. Davies, J., Spiegelman, G. B., & Yim, G. (2006). The world of subinhibitory antibiotic concentrations. *Current Opinion in Microbiology, 9*(5), 445-453. doi:https://doi.org/10.1016/j.mib.2006.08.006

71. Davila-Cespedes, A., Hufendiek, P., Crusemann, M., Schaberle, T. F., & Konig, G. M. (2016). Marine-derived myxobacteria of the suborder Nannocystineae: An underexplored source of structurally intriguing and biologically active metabolites. *Beilstein J Org Chem, 12*, 969-984. doi:10.3762/bjoc.12.96

72. Davison, J. R., Lohith, K. M., Wang, X., Bobyk, K., Mandadapu, S. R., Lee, S. L., . . . Bewley, C. A. (2017). A New Natural Product Analog of Blasticidin S Reveals Cellular Uptake Facilitated by the NorA Multidrug Transporter. *Antimicrob Agents Chemother, 61*(6). doi:10.1128/AAC.02635-16

73. de Carpentier, F., Le Peillet, J., Boisset, N. D., Crozet, P., Lemaire, S. D., & Danon, A. (2020). Blasticidin S Deaminase: A New Efficient Selectable Marker for Chlamydomonas reinhardtii. *Front Plant Sci, 11*, 242. doi:10.3389/fpls.2020.00242

74. DeBoer, C., Meulman, P. A., Wnuk, R. J., & Peterson, D. H. (1970). Geldanamycin, a new antibiotic. *J Antibiot (Tokyo), 23*(9), 442-447. doi:10.7164/antibiotics.23.442

75. Del Carratore, F., Zych, K., Cummings, M., Takano, E., Medema, M. H., & Breitling, R. (2019). Computational identification of co-evolving multi-gene modules in microbial biosynthetic gene clusters. *Commun Biol, 2*, 83. doi:10.1038/s42003-019-0333-6

76. Demain, A. L., & Fang, A. (2000). The natural functions of secondary metabolites. *Adv Biochem Eng Biotechnol, 69*, 1-39. doi:10.1007/3-540-44964-7_1

77. Dettmer, K., Nurnberger, N., Kaspar, H., Gruber, M. A., Almstetter, M. F., & Oefner, P. J. (2011). Metabolite extraction from adherently growing mammalian cells for metabolomics studies: optimization of harvesting and extraction protocols. *Anal Bioanal Chem, 399*(3), 1127-1139. doi:10.1007/s00216-010-4425-x

78. Devine, R., McDonald, H. P., Qin, Z., Arnold, C. J., Noble, K., Chandra, G., . . . Hutchings, M. I. (2021). Re-wiring the regulation of the formicamycin biosynthetic gene cluster to enable the development of promising antibacterial compounds. *Cell Chem Biol, 28*(4), 515-523.e515. doi:10.1016/j.chembiol.2020.12.011

79. Dilika, F., Bremner, P. D., & Meyer, J. J. M. (2000). Antibacterial activity of linoleic and oleic acids isolated from Helichrysum pedunculatum: a plant used during circumcision rites. *Fitoterapia, 71*(4), 450-452. doi:https://doi.org/10.1016/S0367-326X(00)00150-7

80. Dominguez-Moñino, I., Diaz-Herraiz, M., Jurado, V., Laiz, L., Miller, A. Z., Santos, J. L., . . . Saiz-Jimenez, C. (2017). Nature and origin of the violet stains on the walls of a Roman tomb. *Science of The Total Environment, 598*, 889-899. doi:https://doi.org/10.1016/j.scitotenv.2017.04.017

81. 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. *Nat Chem Biol, 10*(11), 963-968. doi:10.1038/nchembio.1659

82. Downie, H., Holden, N., Otten, W., Spiers, A. J., Valentine, T. A., & Dupuy, L. X. (2012). Transparent soil for imaging the rhizosphere. *PLoS One, 7*(9), e44276. doi:10.1371/journal.pone.0044276

83. Dundore-Arias, J. P., Castle, S. C., Felice, L., Dill-Macky, R., & Kinkel, L. L. (2019). Carbon Amendments Influence Composition and Functional Capacities of Indigenous Soil Microbiomes. *Front Mol Biosci, 6*, 151. doi:10.3389/fmolb.2019.00151

84. Dundore-Arias, J. P., Felice, L., Dill-Macky, R., & Kinkel, L. L. (2019). Carbon Amendments Induce Shifts in Nutrient Use, Inhibitory, and Resistance Phenotypes Among Soilborne Streptomyces. *Front Microbiol, 10*, 498. doi:10.3389/fmicb.2019.00498

85. Dunkel, M., Fullbeck, M., Neumann, S., & Preissner, R. (2006). SuperNatural: a searchable database of available natural compounds. *Nucleic Acids Res, 34*(Database issue), D678-683. doi:10.1093/nar/gkj132

86. Ehrlich, K. C., & Cotty, P. J. (2002). Variability in nitrogen regulation of aflatoxin production by Aspergillus flavus strains. *Appl Microbiol Biotechnol, 60*(1-2), 174-178. doi:10.1007/s00253-002-1094-5

87. Elliot, M. A., Bibb, M. J., Buttner, M. J., & Leskiw, B. K. (2001). BldD is a direct regulator of key developmental genes in Streptomyces coelicolor A3(2). *Mol Microbiol, 40*(1), 257-269. doi:10.1046/j.1365-2958.2001.02387.x

88. Fang, Y., Wang, J., Tang, Y., Guo, Z., Bai, J., Wu, L., . . . Zhang, D. (2023). Geninthiocins E and F, two new cyclic thiopeptides with antiviral activities from soil-derived Streptomyces sp. CPCC 200267 using OSMAC strategy. *The Journal of Antibiotics, 76*(2), 101-104. doi:10.1038/s41429-022-00580-0

89. Felder, S., Kehraus, S., Neu, E., Bierbaum, G., Schaberle, T. F., & Konig, G. M. (2013). Salimyxins and enhygrolides: antibiotic, sponge-related metabolites from the obligate marine myxobacterium Enhygromyxa salina. *Chembiochem, 14*(11), 1363-1371. doi:10.1002/cbic.201300268

90. Feng, Y., Wang, J., Zhang, J., Qi, X., Long, W., Ding, Y., & Liu, L. (2023). Soil microbes support Janzen's mountain passes hypothesis: The role of local-scale climate variability along a tropical montane gradient. *Front Microbiol, 14*, 1135116. doi:10.3389/fmicb.2023.1135116

91. Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., . . . *et al.* (1995). Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science, 269*(5223), 496-512. doi:10.1126/science.7542800

92. Fleming, A. (1944). THE DISCOVERY OF PENICILLIN. *British Medical Bulletin, 2*(1), 4-5. doi:10.1093/oxfordjournals.bmb.a071032

93. Forseth, R. R., & Schroeder, F. C. (2011). NMR-spectroscopic analysis of mixtures: from structure to function. *Curr Opin Chem Biol, 15*(1), 38-47. doi:10.1016/j.cbpa.2010.10.010

94. Galinski, E. A., Pfeiffer, H. P., & Trüper, H. G. (1985). 1,4,5,6-Tetrahydro-2-methyl-4-pyrimidinecarboxylic acid. A novel cyclic amino acid from halophilic phototrophic bacteria of the genus Ectothiorhodospira. *Eur J Biochem, 149*(1), 135-139. doi:10.1111/j.1432-1033.1985.tb08903.x

95. 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. doi:10.1038/s41564-022-01110-2

96. Gaynes, R. (2017). The Discovery of Penicillin—New Insights After More Than 75 Years of Clinical Use. *Emerging Infectious Diseases, 23*(5), 849-853. doi:10.3201/eid2305.161556

97. Geraci, J. E., Heilman, F. R., Nichols, D. R., Wellman, E. W., & Ross, G. T. (1956). Some laboratory and clinical experiences with a new antibiotic, vancomycin. *Antibiot Annu*, 90-106.

98. Gerber, N. N., & Lechevalier, H. A. (1965). Geosmin, an earthly-smelling substance isolated from actinomycetes. *Applied microbiology, 13*(6), 935-938. doi:10.1128/am.13.6.935-938.1965

99. Goering, A. W., McClure, R. A., Doroghazi, J. R., Albright, J. C., Haverland, N. A., Zhang, Y., . . . Kelleher, N. L. (2016). Metabologenomics: Correlation of Microbial Gene Clusters with Metabolites Drives Discovery of a Nonribosomal Peptide with an Unusual Amino Acid Monomer. *ACS Cent Sci, 2*(2), 99-108. doi:10.1021/acscentsci.5b00331

100. Gohlke, R. S. (1959). Time-of-Flight Mass Spectrometry and Gas-Liquid Partition Chromatography. *Analytical Chemistry, 31*(4), 535-541. doi:10.1021/ac50164a024

101. Gokhale, R. S., Tsuji, S. Y., Cane, D. E., & Khosla, C. (1999). Dissecting and exploiting intermodular communication in polyketide synthases. *Science, 284*(5413), 482-485. doi:10.1126/science.284.5413.482

102. Graham, E., Lee, J., Price, M., Tarailo-Graovac, M., Matthews, A., Engelke, U., . . . Mostafavi, S. (2018). Integration of genomics and metabolomics for prioritization of rare disease variants: a 2018 literature review. *J Inherit Metab Dis, 41*(3), 435-445. doi:10.1007/s10545-018-0139-6

103. Graham-Taylor, C., Kamphuis, L. G., & Derbyshire, M. C. (2020). A detailed in silico analysis of secondary metabolite biosynthesis clusters in the genome of the broad host range plant pathogenic fungus Sclerotinia sclerotiorum. *BMC Genomics, 21*(1), 7. doi:10.1186/s12864-019-6424-4

104. Gramajo, H. C., Takano, E., & Bibb, M. J. (1993). Stationary-phase production of the antibiotic actinorhodin in Streptomyces coelicolor A3(2) is transcriptionally regulated. *Mol Microbiol, 7*(6), 837-845. doi:10.1111/j.1365-2958.1993.tb01174.x

105. Grebe, S. K., & Singh, R. J. (2011). LC-MS/MS in the Clinical Laboratory - Where to From Here? *Clin Biochem Rev, 32*(1), 5-31. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/21451775

106. Gupta, A., Bedre, R., Thapa, S. S., Sabrin, A., Wang, G., Dassanayake, M., & Grove, A. (2017). Global Awakening of Cryptic Biosynthetic Gene Clusters in Burkholderia thailandensis. *ACS Chem Biol, 12*(12), 3012-3021. doi:10.1021/acschembio.7b00681

107.   Gürtler, H., Pedersen, R., Anthoni, U., Christophersen, C., Nielsen, P. H., Wellington, E. M., . . . Bock, K. (1994). Albaflavenone, a sesquiterpene ketone with a zizaene skeleton produced by a streptomycete with a new rope morphology. *J Antibiot (Tokyo), 47*(4), 434-439. doi:10.7164/antibiotics.47.434

108.   Haeder, S., Wirth, R., Herz, H., & Spiteller, D. (2009). Candicidin-producing <i>Streptomyces</i> support leaf-cutting ants to protect their fungus garden against the pathogenic fungus <i>Escovopsis</i>. *Proceedings of the National Academy of Sciences, 106*(12), 4742-4746. doi:doi:10.1073/pnas.0812082106

109.   Haft, D. H., Selengut, J. D., Richter, R. A., Harkins, D., Basu, M. K., & Beck, E. (2013). TIGRFAMs and Genome Properties in 2013. *Nucleic Acids Res, 41*(Database issue), D387-395. doi:10.1093/nar/gks1234

110.   Haginaka, K., Asamizu, S., Ozaki, T., Igarashi, Y., Furumai, T., & Onaka, H. (2014). Genetic approaches to generate hyper-producing strains of goadsporin: the relationships between productivity and gene duplication in secondary metabolite biosynthesis. *Bioscience, Biotechnology, and Biochemistry, 78*(3), 394-399. doi:10.1080/09168451.2014.885824

111.   Hannigan, G. D., Prihoda, D., Palicka, A., Soukup, J., Klempir, O., Rampula, L., . . . Bitton, D. A. (2019). A deep learning genome-mining strategy for biosynthetic gene cluster prediction. *Nucleic Acids Res, 47*(18), e110. doi:10.1093/nar/gkz654

112.   Hashimoto, T., Kozone, I., Hashimoto, J., Ueoka, R., Kagaya, N., Fujie, M., . . . Shin-ya, K. (2020). Novel macrolactam compound produced by the heterologous expression of a large cryptic biosynthetic gene cluster of Streptomyces rochei IFO12908. *The Journal of Antibiotics, 73*(3), 171-174. doi:10.1038/s41429-019-0265-x

113.   Hasin, Y., Seldin, M., & Lusis, A. (2017). Multi-omics approaches to disease. *Genome Biology, 18*(1), 83. doi:10.1186/s13059-017-1215-1

114.   Hazra, M., Joshi, H., Williams, J. B., & Watts, J. E. M. (2022). Antibiotics and antibiotic resistant bacteria/genes in urban wastewater: A comparison of their fate in conventional treatment systems and constructed wetlands. *Chemosphere, 303*(Pt 2), 135148. doi:10.1016/j.chemosphere.2022.135148

115.   Helfrich, E. J., & Piel, J. (2016). Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat Prod Rep, 33*(2), 231-316. doi:10.1039/c5np00125k

116.   Hellwig, N., Gómez-Brandón, M., Ascher-Jenull, J., Bardelli, T., Anschlag, K., Fornasier, F., . . . Broll, G. (2018). Humus Forms and Soil Microbiological Parameters in a Mountain Forest: Upscaling to the Slope Scale. *Soil Systems, 2*(1), 12. Retrieved from https://www.mdpi.com/2571-8789/2/1/12

117.	Hertweck, C., Luzhetskyy, A., Rebets, Y., & Bechthold, A. (2007). Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Natural Product Reports, 24*(1), 162-190. doi:10.1039/B507395M

118.	Hillis, D. M., & Bull, J. J. (1993). An Empirical Test of Bootstrapping as a Method for Assessing Confidence in Phylogenetic Analysis. *Systematic Biology, 42*(2), 182-192. doi:10.1093/sysbio/42.2.182

119.	Hjörleifsson Eldjárn, G., Ramsay, A., van der Hooft, J. J. J., Duncan, K. R., Soldatou, S., Rousu, J., . . . Rogers, S. (2021). Ranking microbial metabolomic and genomic links in the NPLinker framework using complementary scoring functions. *PLOS Computational Biology, 17*(5), e1008920. doi:10.1371/journal.pcbi.1008920

120.	Ho, C. S., Lam, C. W., Chan, M. H., Cheung, R. C., Law, L. K., Lit, L. C., . . . Tai, H. L. (2003). Electrospray ionisation mass spectrometry: principles and clinical applications. *Clin Biochem Rev, 24*(1), 3-12.

121.	Hojati, Z., Milne, C., Harvey, B., Gordon, L., Borg, M., Flett, F., . . . Micklefield, J. (2002). Structure, Biosynthetic Origin, and Engineered Biosynthesis of Calcium-Dependent Antibiotics from Streptomyces coelicolor. *Chemistry & Biology, 9*(11), 1175-1187. doi:https://doi.org/10.1016/S1074-5521(02)00252-1

122.	Hollstein, U. (1974). Actinomycin. Chemistry and mechanism of action. *Chemical Reviews, 74*(6), 625-652. doi:10.1021/cr60292a002

123.	Hong, H., Landauer, M. R., Foriska, M. A., & Ledney, G. D. (2006). Antibacterial activity of the soy isoflavone genistein. *J Basic Microbiol, 46*(4), 329-335. doi:10.1002/jobm.200510073

124.	Hoshino, S., Okada, M., Onaka, H., & Abe, I. (2016). Effective Production of Aromatic Polyketides in Streptomyces using a Combined-Culture Method. *Nat Prod Commun, 11*(7), 979-981.

125.	Hoskisson, P. A., & Seipke, R. F. (2020). Cryptic or Silent? The Known Unknowns, Unknown Knowns, and Unknown Unknowns of Secondary Metabolism. *mBio, 11*(5). doi:10.1128/mBio.02642-20

126.	Huang, H., Zheng, G., Jiang, W., Hu, H., & Lu, Y. (2015). One-step high-efficiency CRISPR/Cas9-mediated genome editing in <italic>Streptomyces</italic>. *Acta Biochimica et Biophysica Sinica, 47*(4), 231-243. doi:https://doi.org/10.1093/abbs/gmv007

127.	Huang, K. T., Misato, T., & Asuyama, H. (1964). SELECTIVE TOXICITY OF BLASTICIDIN S TO PIRICULARIA ORYZAE AND PELLICULARIA SASAKII. *J Antibiot (Tokyo), 17*, 71-74.

128.   Hutchings, M. I., Truman, A. W., & Wilkinson, B. (2019). Antibiotics: past, present and future. *Current Opinion in Microbiology, 51*, 72-80. doi:https://doi.org/10.1016/j.mib.2019.10.008

129.   Hutchings, M. I., Truman, A. W., & Wilkinson, B. (2019). Antibiotics: past, present and future. *Curr Opin Microbiol, 51*, 72-80. doi:10.1016/j.mib.2019.10.008

130.   Hwang, G. J., Jang, M., Son, S., Kim, G. S., Lee, B., Heo, K. T., . . . Jang, J.-H. (2022). Ulleungdolin, a Polyketide–Peptide Hybrid Bearing a 2,4-Di-O-methyl-β-d-antiarose from Streptomyces sp. 13F051 Co-cultured with Leohumicola minima 15S071. *Journal of Natural Products, 85*(10), 2445-2453. doi:10.1021/acs.jnatprod.2c00682

131.   Iliĉ, S. B., Konstantinoviĉ, S. S., Todoroviĉ, Z. B., Laziĉ, M. L., Veljkoviĉ, V. B., Jokoviĉ, N., & Radovanoviĉ, B. C. (2007). [Characterization and antimicrobial activity of the bioactive metabolites in streptomycete isolates]. *Mikrobiologiia, 76*(4), 480-487.

132.   Iskander, M., Bathurst, R., & Omidvar, M. (2015). Past, present, and future of transparent soils. *Geotechnical Testing Journal, 38*(5), 557-573.

133.   Ivanisevic, J., Zhu, Z. J., Plate, L., Tautenhahn, R., Chen, S., O'Brien, P. J., . . . Siuzdak, G. (2013). Toward 'omic scale metabolite profiling: a dual separation-mass spectrometry approach for coverage of lipid and central carbon metabolism. *Anal Chem, 85*(14), 6876-6884. doi:10.1021/ac401140h

134.   Izawa, M., Kawasaki, T., & Hayakawa, Y. (2013). Cloning and heterologous expression of the thioviridamide biosynthesis gene cluster from Streptomyces olivoviridis. *Appl Environ Microbiol, 79*(22), 7110-7113. doi:10.1128/AEM.01978-13

135.   Izoré, T., Candace Ho, Y. T., Kaczmarski, J. A., Gavriilidou, A., Chow, K. H., Steer, D. L., . . . Cryle, M. J. (2021). Structures of a non-ribosomal peptide synthetase condensation domain suggest the basis of substrate selectivity. *Nature Communications, 12*(1), 2511. doi:10.1038/s41467-021-22623-0

136.   Jabila Mary, T. R., Kannan, R. R., Iniyan, A. M., Ramachandran, D., & Prakash Vincent, S. G. (2021). Cell wall distraction and biofilm inhibition of marine Streptomyces derived angucycline in methicillin resistant Staphylococcus aureus. *Microb Pathog, 150*, 104712. doi:10.1016/j.micpath.2020.104712

137.   Jakimowicz, D., & van Wezel, G. P. (2012). Cell division and DNA segregation in Streptomyces: how to build a septum in the middle of nowhere? *Mol Microbiol, 85*(3), 393-404. doi:10.1111/j.1365-2958.2012.08107.x

138.   James, A. T., & Martin, A. J. (1952). Gas-liquid partition chromatography; the separation and micro-estimation of volatile fatty acids from formic acid to dodecanoic acid. *Biochem J, 50*(5), 679-690. doi:10.1042/bj0500679

139.   Jia, Q., Chen, X., Köllner, T. G., Rinkel, J., Fu, J., Labbé, J., . . . Chen, F. (2019). Terpene Synthase Genes Originated from Bacteria through Horizontal Gene Transfer Contribute to Terpenoid Diversity in Fungi. *Sci Rep, 9*(1), 9223. doi:10.1038/s41598-019-45532-1

140.   Jiang, J., He, X., & Cane, D. E. (2006). Geosmin biosynthesis. Streptomyces coelicolor germacradienol/germacrene D synthase converts farnesyl diphosphate to geosmin. *J Am Chem Soc, 128*(25), 8128-8129. doi:10.1021/ja062669x

141.   Jiang, J., He, X., & Cane, D. E. (2007). Biosynthesis of the earthy odorant geosmin by a bifunctional Streptomyces coelicolor enzyme. *Nat Chem Biol, 3*(11), 711-715. doi:10.1038/nchembio.2007.29

142.   Jones, M. B., Nierman, W. C., Shan, Y., Frank, B. C., Spoering, A., Ling, L., . . . Nelson, K. E. (2017). Reducing the Bottleneck in Discovery of Novel Antibiotics. *Microb Ecol, 73*(3), 658-667. doi:10.1007/s00248-016-0889-3

143.   Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., . . . Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature, 596*(7873), 583-589. doi:10.1038/s41586-021-03819-2

144.   Kalaitzis, J. A., Hamano, Y., Nilsen, G., & Moore, B. S. (2003). Biosynthesis and structural revision of neomarinone. *Organic Letters, 5*(23), 4449-4452.

145.   Kaltenpoth, M. (2009). Actinobacteria as mutualists: general healthcare for insects? *Trends in Microbiology, 17*(12), 529-535. doi:https://doi.org/10.1016/j.tim.2009.09.006

146.   Kang, D. D., Froula, J., Egan, R., & Wang, Z. (2015). MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ, 3*, e1165. doi:10.7717/peerj.1165

147.   Kannaste, A., Copolovici, L., & Niinemets, U. (2014). Gas chromatography-mass spectrometry method for determination of biogenic volatile organic compounds emitted by plants. *Methods Mol Biol, 1153*, 161-169. doi:10.1007/978-1-4939-0606-2_11

148.   Katsifas, E. A., Koraki, T. G., & Karagouni, A. D. (2000). Determination of metabolic activity of streptomycetes in soil microcosms. *Journal of Applied Microbiology, 89*(1), 178-184. doi:https://doi.org/10.1046/j.1365-2672.2000.01094.x

149.   Kautsar, S. A., Blin, K., Shaw, S., Navarro-Munoz, J. C., Terlouw, B. R., van der Hooft, J. J. J., . . . Medema, M. H. (2020). MIBiG 2.0: a repository for biosynthetic gene clusters of known function. *Nucleic Acids Res, 48*(D1), D454-D458. doi:10.1093/nar/gkz882

150.   Kawahara, T., Izumikawa, M., Kozone, I., Hashimoto, J., Kagaya, N., Koiwai, H., . . . Shin-Ya, K. (2018). Neothioviridamide, a Polythioamide Compound Produced

by Heterologous Expression of a Streptomyces sp. Cryptic RiPP Biosynthetic Gene Cluster. *J Nat Prod, 81*(2), 264-269. doi:10.1021/acs.jnatprod.7b00607

151.   Kaweewan, I., Hemmi, H., Komaki, H., & Kodani, S. (2020). Isolation and structure determination of a new antibacterial peptide pentaminomycin C from Streptomyces cacaoi subsp. cacaoi. *J Antibiot (Tokyo), 73*(4), 224-229. doi:10.1038/s41429-019-0272-y

152.   Kelemen, G. H., Brian, P., Flärdh, K., Chamberlin, L., Chater, K. F., & Buttner, M. J. (1998). Developmental regulation of transcription of whiE, a locus specifying the polyketide spore pigment in Streptomyces coelicolor A3 (2). *J Bacteriol, 180*(9), 2515-2521. doi:10.1128/jb.180.9.2515-2521.1998

153.   Keller, N. P., Turner, G., & Bennett, J. W. (2005). Fungal secondary metabolism - from biochemistry to genomics. *Nat Rev Microbiol, 3*(12), 937-947. doi:10.1038/nrmicro1286

154.   Kenig, M., & Reading, C. (1979). Holomycin and an antibiotic (MM 19290) related to tunicamycin, metabolites of Streptomyces clavuligerus. *J Antibiot (Tokyo), 32*(6), 549-554. doi:10.7164/antibiotics.32.549

155.   Kersten, R. D., Yang, Y. L., Xu, Y., Cimermancic, P., Nam, S. J., Fenical, W., . . . Dorrestein, P. C. (2011). A mass spectrometry-guided genome mining approach for natural product peptidogenomics. *Nat Chem Biol, 7*(11), 794-802. doi:10.1038/nchembio.684

156.   Keys, C. J., Dare, D. J., Sutton, H., Wells, G., Lunt, M., McKenna, T., . . . Shah, H. N. (2004). Compilation of a MALDI-TOF mass spectral database for the rapid screening and characterisation of bacteria implicated in human infectious diseases. *Infection, Genetics and Evolution, 4*(3), 221-242. doi:https://doi.org/10.1016/j.meegid.2004.02.004

157.   Kim, K. R., Kim, T. J., & Suh, J. W. (2008). The gene cluster for spectinomycin biosynthesis and the aminoglycoside-resistance function of spcM in Streptomyces spectabilis. *Curr Microbiol, 57*(4), 371-374. doi:10.1007/s00284-008-9204-y

158.   Kim, T. K., Hewavitharana, A. K., Shaw, P. N., & Fuerst, J. A. (2006). Discovery of a new source of rifamycin antibiotics in marine sponge actinobacteria by phylogenetic prediction. *Appl Environ Microbiol, 72*(3), 2118-2125. doi:10.1128/aem.72.3.2118-2125.2006

159.   Kleeff, J., Kornmann, M., Sawhney, H., & Korc, M. (2000). Actinomycin D induces apoptosis and inhibits growth of pancreatic cancer cells. *Int J Cancer, 86*(3), 399-407. doi:10.1002/(sici)1097-0215(20000501)86:3<399::aid-ijc15>3.0.co;2-g

160.   Klemm, E. J., Wong, V. K., & Dougan, G. (2018). Emergence of dominant multidrug-resistant bacterial clades: Lessons from history and whole-genome

sequencing. *Proceedings of the National Academy of Sciences, 115*(51), 12872-12877. doi:10.1073/pnas.1717162115

161.   Kloosterman, A. M., Medema, M. H., & van Wezel, G. P. (2021). Omics-based strategies to discover novel classes of RiPP natural products. *Current Opinion in Biotechnology, 69*, 60-67. doi:https://doi.org/10.1016/j.copbio.2020.12.008

162.   Knappe, T. A., Linne, U., Zirah, S., Rebuffat, S., Xie, X., & Marahiel, M. A. (2008). Isolation and Structural Characterization of Capistruin, a Lasso Peptide Predicted from the Genome Sequence of Burkholderia thailandensis E264. *Journal of the American Chemical Society, 130*(34), 11446-11454. doi:10.1021/ja802966g

163.   Kontro, M., Lignell, U., Hirvonen, M. R., & Nevalainen, A. (2005). pH effects on 10 Streptomyces spp. growth and sporulation depend on nutrients. *Lett Appl Microbiol, 41*(1), 32-38. doi:10.1111/j.1472-765X.2005.01727.x

164.   Köster, W. (2001). ABC transporter-mediated uptake of iron, siderophores, heme and vitamin B12. *Research in Microbiology, 152*(3), 291-301. doi:https://doi.org/10.1016/S0923-2508(01)01200-1

165.   Kumar, R., & Jadeja, V. (2016). Isolation of Actinomycetes: A Complete Approach. *International Journal of Current Microbiology and Applied Sciences, 5*, 606-618. doi:10.20546/ijcmas.2016.505.062

166.   Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol, 35*(6), 1547-1549. doi:10.1093/molbev/msy096

167.   Kurt-Kızıldoğan, A., Akarsu, N., Otur, Ç., Kivrak, A., Aslan-Ertas, N., Arslan, S., . . . Şahin, N. (2022). A Novel 4H-Chromen-4-One Derivative from Marine Streptomyces ovatisporus S4702(T) as Potential Antibacterial and Anti-Cancer Agent. *Anticancer Agents Med Chem, 22*(2), 362-370. doi:10.2174/1871520621666210311085748

168.   Lee, N., Kim, W., Chung, J., Lee, Y., Cho, S., Jang, K. S., . . . Cho, B. K. (2020). Iron competition triggers antibiotic biosynthesis in Streptomyces coelicolor during coculture with Myxococcus xanthus. *ISME J, 14*(5), 1111-1124. doi:10.1038/s41396-020-0594-6

169.   Letunic, I., & Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res, 47*(W1), W256-W259. doi:10.1093/nar/gkz239

170.   Levine, D. P. (2006). Vancomycin: A History. *Clinical Infectious Diseases, 42*(Supplement_1), S5-S12. doi:10.1086/491709

171.   Lewis, K. (2013). Platforms for antibiotic discovery. *Nature Reviews Drug Discovery, 12*(5), 371-387. doi:10.1038/nrd3975

172.   Li, B., & Walsh, C. T. (2010). Identification of the gene cluster for the dithiolopyrrolone antibiotic holomycin in Streptomyces clavuligerus. *Proc Natl Acad Sci U S A, 107*(46), 19731-19735. doi:10.1073/pnas.1014140107

173.   Li, L., Hao, Y., Zheng, Z., Wang, W., Biederman, J. A., Wang, Y., . . . Xu, Z. (2022). Heavy rainfall in peak growing season had larger effects on soil nitrogen flux and pool than in the late season in a semiarid grassland. *Agriculture, Ecosystems & Environment, 326*, 107785. doi:https://doi.org/10.1016/j.agee.2021.107785

174.   Li, M. H., Ung, P. M., Zajkowski, J., Garneau-Tsodikova, S., & Sherman, D. H. (2009). Automated genome mining for natural products. *BMC Bioinformatics, 10*, 185. doi:10.1186/1471-2105-10-185

175.   Li, S., Hu, X., Li, L., Liu, H., Yu, L., You, X., . . . Wu, L. (2019). Geninthiocins C and D from Streptomyces as 35-membered macrocyclic thiopeptides with modified tail moiety. *The Journal of Antibiotics, 72*(2), 106-110. doi:10.1038/s41429-018-0127-y

176.   Lianza, M., Leroy, R., Machado Rodrigues, C., Borie, N., Sayagh, C., Remy, S., . . . Nuzillard, J. M. (2021). The Three Pillars of Natural Product Dereplication. Alkaloids from the Bulbs of Urceolina peruviana (C. Presl) J.F. Macbr. as a Preliminary Test Case. *Molecules, 26*(3). doi:10.3390/molecules26030637

177.   Lioy, V. S., Lorenzi, J. N., Najah, S., Poinsignon, T., Leh, H., Saulnier, C., . . . Bury-Mone, S. (2021). Dynamics of the compartmentalized Streptomyces chromosome during metabolic differentiation. *Nat Commun, 12*(1), 5221. doi:10.1038/s41467-021-25462-1

178.   Liu, Y., He, N., Wen, X., Xu, L., Sun, X., Yu, G., . . . Schipper, L. A. (2018). The optimum temperature of soil microbial respiration: Patterns and controls. *Soil Biology and Biochemistry, 121*, 35-42. doi:https://doi.org/10.1016/j.soilbio.2018.02.019

179.   Liu, Y. Y., Wang, Y., Walsh, T. R., Yi, L. X., Zhang, R., Spencer, J., . . . Shen, J. (2016). Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study. *Lancet Infect Dis, 16*(2), 161-168. doi:10.1016/S1473-3099(15)00424-7

180.   Liu, Z., Zhao, Y., Huang, C., & Luo, Y. (2021). Recent Advances in Silent Gene Cluster Activation in Streptomyces. *Frontiers in Bioengineering and Biotechnology, 9*. doi:10.3389/fbioe.2021.632230

181.   Llor, C., & Bjerrum, L. (2014). Antimicrobial resistance: risk associated with antibiotic overuse and initiatives to reduce the problem. *Ther Adv Drug Saf, 5*(6), 229-241. doi:10.1177/2042098614554919

182.   Loucif, L., Bendjama, E., Gacemi-Kirane, D., & Rolain, J. M. (2014). Rapid identification of Streptomyces isolates by MALDI-TOF MS. *Microbiol Res, 169*(12), 940-947. doi:10.1016/j.micres.2014.04.004

183.   Lozano Andrade, C., Nogueira, C. A., Wibowo, M., & Kovács, Á. (2022). *Establishment of a transparent soil system to study Bacillus subtilis chemical ecology.*

184.   Lu, D. F., Wang, Y. S., Li, C., Wei, G. J., Chen, R., Dong, D. M., & Yao, M. (2015). Actinomycin D inhibits cell proliferations and promotes apoptosis in osteosarcoma cells. *Int J Clin Exp Med, 8*(2), 1904-1911.

185.   Ludwig, W., Euzéby, J., Schumann, P., Busse, H.-J., Trujillo, M. E., Kämpfer, P., & Whitman, W. B. (2012). Road map of the phylum Actinobacteria. In *Bergey's Manual of Systematics of Archaea and Bacteria* (pp. 1-37).

186.   Ma, L., Shi, Y., Siemianowski, O., Yuan, B., Egner, T. K., Mirnezami, S. V., . . . Cademartiri, L. (2019). Hydrogel-based transparent soils for root phenotyping in vivo. *Proceedings of the National Academy of Sciences, 116*(22), 11063-11068. doi:doi:10.1073/pnas.1820334116

187.   Maansson, M., Vynne, N. G., Klitgaard, A., Nybo, J. L., Melchiorsen, J., Nguyen, D. D., . . . Gram, L. (2016). An Integrated Metabolomic and Genomic Mining Workflow To Uncover the Biosynthetic Potential of Bacteria. *mSystems, 1*(3). doi:10.1128/mSystems.00028-15

188.   Machushynets, N. V., Wu, C., Elsayed, S. S., Hankemeier, T., & van Wezel, G. P. (2019). Discovery of novel glycerolated quinazolinones from Streptomyces sp. MBT27. *J Ind Microbiol Biotechnol, 46*(3-4), 483-492. doi:10.1007/s10295-019-02140-2

189.   MacPherson, D. W., Gushulak, B. D., Baine, W. B., Bala, S., Gubbins, P. O., Holtom, P., & Segarra-Newnham, M. (2009). Population mobility, globalization, and antimicrobial drug resistance. *Emerg Infect Dis, 15*(11), 1727-1732. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/19891858

190.   Malpartida, F., & Hopwood, D. A. (1984). Molecular cloning of the whole biosynthetic pathway of a Streptomyces antibiotic and its expression in a heterologous host. *Nature, 309*(5967), 462-464. doi:10.1038/309462a0

191.   Maltezou, H. C., Theodoridou, M., & Daikos, G. L. (2017). Antimicrobial resistance and the current refugee crisis. *J Glob Antimicrob Resist, 10*, 75-79. doi:10.1016/j.jgar.2017.03.013

192.   Manteca, A., Fernandez, M., & Sanchez, J. (2005). Mycelium development in Streptomyces antibioticus ATCC11891 occurs in an orderly pattern which determines multiphase growth curves. *BMC Microbiol, 5*, 51. doi:10.1186/1471-2180-5-51

193.   Manteca, A., & Yague, P. (2018). Streptomyces Differentiation in Liquid Cultures as a Trigger of Secondary Metabolism. *Antibiotics (Basel), 7*(2). doi:10.3390/antibiotics7020041

194.   Martin, M. J., Thottathil, S. E., & Newman, T. B. (2015). Antibiotics Overuse in Animal Agriculture: A Call to Action for Health Care Providers. *Am J Public Health, 105*(12), 2409-2410. doi:10.2105/AJPH.2015.302870

195.   Martinez, A. F. C., Mello, F. M. P., Zucchi, T. D., Melo, I. S., & Moraes, L. A. B. (2020). Tandem mass spectrometry methods to accelerate the identification of phytotoxic metabolites produced by Streptomyces sp. 39 PL. *Nat Prod Res, 34*(2), 210-216. doi:10.1080/14786419.2018.1525713

196.   Martínez-Núñez, M. A., & López, V. E. L. y. (2016). Nonribosomal peptides synthetases and their applications in industry. *Sustainable Chemical Processes, 4*(1), 13. doi:10.1186/s40508-016-0057-6

197.   Massung, R. F., Esposito, J. J., Liu, L. I., Qi, J., Utterback, T. R., Knight, J. C., . . . *et al.* (1993). Potential virulence determinants in terminal regions of variola smallpox virus genome. *Nature, 366*(6457), 748-751. doi:10.1038/366748a0

198.   Matarrita-Carranza, B., Murillo-Cruz, C., Avendano, R., Rios, M. I., Chavarria, M., Gomez-Calvo, M. L., . . . Pinto-Tomas, A. A. (2021). Streptomyces sp. M54: an actinobacteria associated with a neotropical social wasp with high potential for antibiotic production. *Antonie Van Leeuwenhoek, 114*(4), 379-398. doi:10.1007/s10482-021-01520-y

199.   McClure, R. A., Goering, A. W., Ju, K. S., Baccile, J. A., Schroeder, F. C., Metcalf, W. W., . . . Kelleher, N. L. (2016). Elucidating the Rimosamide-Detoxin Natural Product Families and Their Biosynthesis Using Metabolite/Gene Cluster Correlations. *ACS Chem Biol, 11*(12), 3452-3460. doi:10.1021/acschembio.6b00779

200.   McCormick, M. H., McGuire, J. M., Pittenger, G. E., Pittenger, R. C., & Stark, W. M. (1955). Vancomycin, a new antibiotic. I. Chemical and biologic properties. *Antibiot Annu, 3*, 606-611. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/13355336

201.   McDaniel, R., Ebert-Khosla, S., Fu, H., Hopwood, D. A., & Khosla, C. (1994). Engineered biosynthesis of novel polyketides: influence of a downstream enzyme on the catalytic specificity of a minimal aromatic polyketide synthase. *Proceedings of the National Academy of Sciences, 91*(24), 11542-11546. doi:doi:10.1073/pnas.91.24.11542

202.   McDaniel, R., Ebert-Khosla, S., Hopwood, D. A., & Khosla, C. (1993). Engineered biosynthesis of novel polyketides. *Science, 262*(5139), 1546-1550. doi:10.1126/science.8248802

203.   McDonald, B. R., & Currie, C. R. (2017). Lateral Gene Transfer Dynamics in the Ancient Bacterial Genus Streptomyces. *mBio, 8*(3). doi:10.1128/mBio.00644-17

204.   McKusick, V. A., & Ruddle, F. H. (1987). A new discipline, a new name, a new journal. *Genomics, 1*, 1-2.

205.   McLafferty, F. W. (2011). A Century of Progress in Molecular Mass Spectrometry. *Annual Review of Analytical Chemistry, 4*(1), 1-22. doi:10.1146/annurev-anchem-061010-114018

206.   McLean, T. C., Lo, R., Tschowri, N., Hoskisson, P. A., Al Bassam, M. M., Hutchings, M. I., & Som, N. F. (2019). Sensing and responding to diverse extracellular signals: an updated analysis of the sensor kinases and response regulators of Streptomyces species. *Microbiology (Reading), 165*(9), 929-952. doi:10.1099/mic.0.000817

207.   Medema, M. H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M. A., . . . Breitling, R. (2011). antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res, 39*(Web Server issue), W339-346. doi:10.1093/nar/gkr466

208.   Medema, M. H., Cimermancic, P., Sali, A., Takano, E., & Fischbach, M. A. (2014). A systematic computational analysis of biosynthetic gene cluster evolution: lessons for engineering biosynthesis. *PLoS Comput Biol, 10*(12), e1004016. doi:10.1371/journal.pcbi.1004016

209.   Medema, M. H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J. B., Blin, K., . . . Glockner, F. O. (2015). Minimum Information about a Biosynthetic Gene cluster. *Nat Chem Biol, 11*(9), 625-631. doi:10.1038/nchembio.1890

210.   Mehetre, G. T., J. S, V., Burkul, B. B., Desai, D., B, S., Dharne, M. S., & Dastager, S. G. (2019). Bioactivities and molecular networking-based elucidation of metabolites of potent actinobacterial strains isolated from the Unkeshwar geothermal springs in India. *RSC Advances, 9*(17), 9850-9859. doi:10.1039/C8RA09449G

211.   Melinda, Y. N., Widada, J., Wahyuningsih, T. D., Febriansah, R., Damayanti, E., & Mustofa, M. (2021). Metabologenomics approach to the discovery of novel compounds from Streptomyces sp. GMR22 as anti-SARS-CoV-2 drugs. *Heliyon, 7*(11), e08308. doi:10.1016/j.heliyon.2021.e08308

212.   Meng-Xi, L. I., Hui-Bin, H., Jie-Yun, L., Jing-Xiao, C. A. O., & Zhen-Wang, Z. (2021). Antibacterial Performance of a Streptomyces spectabilis Strain Producing Metacycloprodigiosin. *Curr Microbiol, 78*(7), 2569-2576. doi:10.1007/s00284-021-02513-w

213.   Miller, B. R., & Gulick, A. M. (2016). Structural Biology of Nonribosomal Peptide Synthetases. *Methods Mol Biol, 1401*, 3-29. doi:10.1007/978-1-4939-3375-4_1

214.	Min Jou, W., Haegeman, G., Ysebaert, M., & Fiers, W. (1972). Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein. *Nature, 237*(5350), 82-88. doi:10.1038/237082a0

215.	Minasny, B., McBratney, A. B., Brough, D. M., & Jacquier, D. (2011). Models relating soil pH measurements in water and calcium chloride that incorporate electrolyte concentration. *European Journal of Soil Science, 62*(5), 728-732. doi:https://doi.org/10.1111/j.1365-2389.2011.01386.x

216.	Mohimani, H., Gurevich, A., Mikheenko, A., Garg, N., Nothias, L. F., Ninomiya, A., . . . Pevzner, P. A. (2017). Dereplication of peptidic natural products through database search of mass spectra. *Nat Chem Biol, 13*(1), 30-37. doi:10.1038/nchembio.2219

217.	Mohimani, H., Gurevich, A., Shlemov, A., Mikheenko, A., Korobeynikov, A., Cao, L., . . . Pevzner, P. A. (2018). Dereplication of microbial metabolites through database search of mass spectra. *Nat Commun, 9*(1), 4035. doi:10.1038/s41467-018-06082-8

218.	Montaser, R., & Kelleher, N. L. (2020). Discovery of the Biosynthetic Machinery for Stravidins, Biotin Antimetabolites. *ACS Chem Biol, 15*(5), 1134-1140. doi:10.1021/acschembio.9b00890

219.	Mudalungu, C. M., von Torne, W. J., Voigt, K., Ruckert, C., Schmitz, S., Sekurova, O. N., . . . Sussmuth, R. D. (2019). Noursamycins, Chlorinated Cyclohexapeptides Identified from Molecular Networking of Streptomyces noursei NTR-SR4. *J Nat Prod, 82*(6), 1478-1486. doi:10.1021/acs.jnatprod.8b00967

220.	Mulder, E., & Antheunisse, J. (1963). *Morphologie, physiologie et ecologie des Arthrobacter*: Masson.

221.	Muok, A. R., Claessen, D., & Briegel, A. (2021). Microbial hitchhiking: how Streptomyces spores are transported by motile soil bacteria. *ISME J, 15*(9), 2591-2600. doi:10.1038/s41396-021-00952-8

222.	Muok, A. R., Claessen, D., & Briegel, A. (2021). Microbial hitchhiking: how Streptomyces spores are transported by motile soil bacteria. *The ISME Journal, 15*(9), 2591-2600. doi:10.1038/s41396-021-00952-8

223.	Murray, C. J. L., Ikuta, K. S., Sharara, F., Swetschinski, L., Robles Aguilar, G., Gray, A., . . . Naghavi, M. (2022). Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet, 399*(10325), 629-655. doi:10.1016/S0140-6736(21)02724-0

224.	Navarro-Munoz, J. C., Selem-Mojica, N., Mullowney, M. W., Kautsar, S. A., Tryon, J. H., Parkinson, E. I., . . . Medema, M. H. (2020). A computational framework

to explore large-scale biosynthetic diversity. *Nat Chem Biol, 16*(1), 60-68. doi:10.1038/s41589-019-0400-9

225.   Neckers, L. (2002). Hsp90 inhibitors as novel cancer chemotherapeutic agents. *Trends Mol Med, 8*(4 Suppl), S55-61. doi:10.1016/s1471-4914(02)02316-x

226.   Nguyen, J. T., Riebschleger, K. K., Brown, K. V., Gorgijevska, N. M., & Nybo, S. E. (2021). A BioBricks toolbox for metabolic engineering of the tetracenomycin pathway. *Biotechnol J*, e2100371. doi:10.1002/biot.202100371

227.   Ni, Y., Jian, Z., Zeng, L., Liu, J., Lei, L., Zhu, J., . . . Xiao, W. (2022). Climate, soil nutrients, and stand characteristics jointly determine large-scale patterns of biomass growth rates and allocation in Pinus massoniana plantations. *Forest Ecology and Management, 504*, 119839. doi:https://doi.org/10.1016/j.foreco.2021.119839

228.   Nicault, M., Zaiter, A., Dumarcay, S., Chaimbault, P., Gelhaye, E., Leblond, P., & Bontemps, C. (2021). Elicitation of Antimicrobial Active Compounds by Streptomyces-Fungus Co-Cultures. *Microorganisms, 9*(1). doi:10.3390/microorganisms9010178

229.   Niu, G., Zheng, J., & Tan, H. (2017). Biosynthesis and combinatorial biosynthesis of antifungal nucleoside antibiotics. *Science China Life Sciences, 60*(9), 939-947. doi:10.1007/s11427-017-9116-0

230.   Nwokeji, P., Enodiana, O. I., Ezenweani, R., Osasere, O.-I., & Abiola, A. (2016). The Chemistry Of Natural Product: Plant Secondary Metabolites. *INTERNATIONAL JOURNAL OF TECHNOLOGY ENHANCEMENTS AND EMERGING ENGINEERING RESEARCH*.

231.   O'Neill, J. (2016). Tackling Drug-Resistant Infections Globally: The Review on Antimicrobial Resistance. *HM Government, Medical*. Retrieved from https://amr-review.org/sites/default/files/160518_Final%20paper_with%20cover.pdf

232.   Ochel, H. J., Eichhorn, K., & Gademann, G. (2001). Geldanamycin: the prototype of a class of antitumor drugs targeting the heat shock protein 90 family of molecular chaperones. *Cell Stress Chaperones, 6*(2), 105-112. doi:10.1379/1466-1268(2001)006<0105:gtpoac>2.0.co;2

233.   Olanrewaju, O. S., & Babalola, O. O. (2019). Streptomyces: implications and interactions in plant growth promotion. *Appl Microbiol Biotechnol, 103*(3), 1179-1188. doi:10.1007/s00253-018-09577-y

234.   Ordóñez-Robles, M., Rodríguez-García, A., & Martín, J. F. (2018). Genome-wide transcriptome response of Streptomyces tsukubaensis to N-acetylglucosamine: effect on tacrolimus biosynthesis. *Microbiological Research, 217*, 14-22. doi:https://doi.org/10.1016/j.micres.2018.08.014

235.   Palazzotto, E., & Weber, T. (2018). Omics and multi-omics approaches to study the biosynthesis of secondary metabolites in microorganisms. *Curr Opin Microbiol, 45*, 109-116. doi:10.1016/j.mib.2018.03.004

236.   Pan, R., Bai, X., Chen, J., Zhang, H., & Wang, H. (2019). Exploring Structural Diversity of Microbe Secondary Metabolites Using OSMAC Strategy: A Literature Review. *Front Microbiol, 10*, 294. doi:10.3389/fmicb.2019.00294

237.   Parkinson, E. I., Tryon, J. H., Goering, A. W., Ju, K. S., McClure, R. A., Kemball, J. D., . . . Metcalf, W. W. (2018). Discovery of the Tyrobetaine Natural Products and Their Biosynthetic Gene Cluster via Metabologenomics. *ACS Chem Biol, 13*(4), 1029-1037. doi:10.1021/acschembio.7b01089

238.   Peirson. (2012). Wilhelm Johannsen's Genotype-Phenotype Distinction. Retrieved from https://embryo.asu.edu/handle/10776/4206

239.   Perez, E. R., Knapp, J. A., Horn, C. K., Stillman, S. L., Evans, J. E., & Arfsten, D. P. (2016). Comparison of LC-MS-MS and GC-MS Analysis of Benzodiazepine Compounds Included in the Drug Demand Reduction Urinalysis Program. *J Anal Toxicol, 40*(3), 201-207. doi:10.1093/jat/bkv140

240.   Pitt, J. J. (2009). Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry. *Clin Biochem Rev, 30*(1), 19-34.

241.   Ponge, J.-F., Jabiol, B., & Gégout, J.-C. (2011). Geology and climate conditions affect more humus forms than forest canopies at large scale in temperate forests. *Geoderma, 162*(1), 187-195. doi:https://doi.org/10.1016/j.geoderma.2011.02.003

242.   Ponge, Z. S. J. (2010). Terrestrial humus forms: ecological relevance and classification. *HAL Open Science*. doi:10.13140/RG.2.1.3713.5521

243.   Pope, M. K., Green, B. D., & Westpheling, J. (1996). The bld mutants of Streptomyces coelicolor are defective in the regulation of carbon utilization, morphogenesis and cell--cell signalling. *Mol Microbiol, 19*(4), 747-756. doi:10.1046/j.1365-2958.1996.414933.x

244.   Prabhu, G. R. D., Williams, E. R., Wilm, M., & Urban, P. L. (2023). Mass spectrometry using electrospray ionization. *Nature Reviews Methods Primers, 3*(1), 23. doi:10.1038/s43586-023-00203-4

245.   Pringle, P. (2012). Experiment Eleven: Dark Secrets Behind the Discovery of a Wonder Drug. In (pp. 269): Walker and Company.

246.   Procopio, R. E., Silva, I. R., Martins, M. K., Azevedo, J. L., & Araujo, J. M. (2012). Antibiotics produced by Streptomyces. *Braz J Infect Dis, 16*(5), 466-471. doi:10.1016/j.bjid.2012.08.014

247.   Radjasa, Oedjijono, & Ryandini. (2021). Bioactive compounds derived from Streptomyces sp. SA32:

248.   antibacterial activity, chemical profile, and their related genes. *IOP Conference Series: Earth and Environmental Science, 948*(1), 1-10.

249.   Ramachandran, P., Rachuri, N. K., Martha, S., Shakthivel, R., Gundala, A., & Battu, T. S. (2019). Implications of Overprescription of Antibiotics: A Cross-Sectional Study. *J Pharm Bioallied Sci, 11*(Suppl 2), S434-S437. doi:10.4103/JPBS.JPBS_62_19

250.   Ranieri, M. R. M., Chan, D. C. K., Yaeger, L. N., Rudolph, M., Karabelas-Pittman, S., Abdo, H., . . . Burrows, L. L. (2019). Thiostrepton Hijacks Pyoverdine Receptors To Inhibit Growth of Pseudomonas aeruginosa. *Antimicrobial Agents and Chemotherapy, 63*(9), e00472-00419. doi:doi:10.1128/AAC.00472-19

251.   Rascher, A., Hu, Z., Viswanathan, N., Schirmer, A., Reid, R., Nierman, W. C., . . . Hutchinson, C. R. (2003). Cloning and characterization of a gene cluster for geldanamycin production in Streptomyces hygroscopicus NRRL 3602. *FEMS Microbiol Lett, 218*(2), 223-230. doi:10.1016/S0378-1097(02)01148-5

252.   Rateb, M. E., Houssen, W. E., Harrison, W. T. A., Deng, H., Okoro, C. K., Asenjo, J. A., . . . Jaspars, M. (2011). Diverse Metabolic Profiles of a Streptomyces Strain Isolated from a Hyper-arid Environment. *Journal of Natural Products, 74*(9), 1965-1971. doi:10.1021/np200470u

253.   Rateb, M. E., Zhai, Y., Ehrner, E., Rath, C. M., Wang, X., Tabudravu, J., . . . Deng, H. (2015). Legonaridin, a new member of linaridin RiPP from a Ghanaian Streptomyces isolate. *Org Biomol Chem, 13*(37), 9585-9592. doi:10.1039/c5ob01269d

254.   Rawlins. (2012). The disputed discovery of streptomycin. *The Lancet, 380*(9838), 207. doi:https://doi.org/10.1016/S0140-6736(12)61202-1

255.   Raynaud, X., & Nunan, N. (2014). Spatial ecology of bacteria at the microscale in soil. *PLoS One, 9*(1), e87217. doi:10.1371/journal.pone.0087217

256.   Reynolds, W. F. (2017). Chapter 29 - Natural Product Structure Elucidation by NMR Spectroscopy. In S. Badal & R. Delgoda (Eds.), *Pharmacognosy* (pp. 567-596). Boston: Academic Press.

257.   Riekeberg, E., & Powers, R. (2017). New frontiers in metabolomics: from measurement to insight. *F1000Res, 6*, 1148. doi:10.12688/f1000research.11495.1

258.   Rigali, S., Nothaft, H., Noens, E. E., Schlicht, M., Colson, S., Muller, M., . . . van Wezel, G. P. (2006). The sugar phosphotransferase system of Streptomyces coelicolor is regulated by the GntR-family regulator DasR and links N-acetylglucosamine metabolism to the control of development. *Mol Microbiol, 61*(5), 1237-1251. doi:10.1111/j.1365-2958.2006.05319.x

259.   Ripa, F. A., Nikkon, F., Zaman, S., & Khondkar, P. (2009). Optimal Conditions for Antimicrobial Metabolites Production from a New Streptomyces sp. RUPA-08PR

Isolated from Bangladeshi Soil. *Mycobiology, 37*(3), 211-214. doi:10.4489/myco.2009.37.3.211

260. Rushing, B. R., & Selim, M. I. (2019). Aflatoxin B1: A review on metabolism, toxicity, occurrence in food, occupational exposure, and detoxification methods. *Food Chem Toxicol, 124*, 81-100. doi:10.1016/j.fct.2018.11.047

261. Russell, A. H., & Truman, A. W. (2020). Genome mining strategies for ribosomally synthesised and post-translationally modified peptides. *Comput Struct Biotechnol J, 18*, 1838-1851. doi:10.1016/j.csbj.2020.06.032

262. Sabatini, M., Comba, S., Altabe, S., Recio-Balsells, A. I., Labadie, G. R., Takano, E., . . . Arabolaza, A. (2018). Biochemical characterization of the minimal domains of an iterative eukaryotic polyketide synthase. *Febs j, 285*(23), 4494-4511. doi:10.1111/febs.14675

263. Salo, O., Guzman-Chavez, F., Ries, M. I., Lankhorst, P. P., Bovenberg, R. A. L., Vreeken, R. J., & Driessen, A. J. M. (2016). Identification of a Polyketide Synthase Involved in Sorbicillin Biosynthesis by Penicillium chrysogenum. *Appl Environ Microbiol, 82*(13), 3971-3978. doi:10.1128/AEM.00350-16

264. Sanchez, S., & Demain, A. L. (2008). Metabolic regulation and overproduction of primary metabolites. *Microb Biotechnol, 1*(4), 283-319. doi:10.1111/j.1751-7915.2007.00015.x

265. Sánchez-Hidalgo, M., Martín, J., & Genilloud, O. (2020). Identification and Heterologous Expression of the Biosynthetic Gene Cluster Encoding the Lasso Peptide Humidimycin, a Caspofungin Activity Potentiator. *Antibiotics, 9*(2), 67. Retrieved from https://www.mdpi.com/2079-6382/9/2/67

266. Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, C. A., . . . Smith, M. (1977). Nucleotide sequence of bacteriophage phi X174 DNA. *Nature, 265*(5596), 687-695. doi:10.1038/265687a0

267. Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A, 74*(12), 5463-5467. doi:10.1073/pnas.74.12.5463

268. Santos-Aberturas, J., Chandra, G., Frattaruolo, L., Lacret, R., Pham, T. H., Vior, N. M., . . . Truman, A. W. (2019). Uncovering the unexplored diversity of thioamidated ribosomal peptides in Actinobacteria using the RiPPER genome mining tool. *Nucleic Acids Research, 47*(9), 4624-4637. doi:10.1093/nar/gkz192

269. Sarma, U. P., Bhetaria, P. J., Devi, P., & Varma, A. (2017). Aflatoxins: Implications on Health. *Indian J Clin Biochem, 32*(2), 124-133. doi:10.1007/s12291-017-0649-2

270.   Sattely, E. S., Fischbach, M. A., & Walsh, C. T. (2008). Total biosynthesis: in vitro reconstitution of polyketide and nonribosomal peptide pathways. *Natural Product Reports, 25*(4), 757-793. doi:10.1039/B801747F

271.   Schaberle, T. F., Orland, A., & Konig, G. M. (2014). Enhanced production of undecylprodigiosin in Streptomyces coelicolor by co-cultivation with the corallopyronin A-producing myxobacterium, Corallococcus coralloides. *Biotechnol Lett, 36*(3), 641-648. doi:10.1007/s10529-013-1406-0

272.   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.

273.   Schatz, A., & Waksman, S. A. (1944). Effect of Streptomycin and Other Antibiotic Substances upon Mycobacterium tuberculosis and Related Organisms. *Proceedings of the Society for Experimental Biology and Medicine, 57*(2), 244-248. doi:10.3181/00379727-57-14769

274.   Schatz, A., & Waksman, S. A. (1945). Strain Specificity and Production of Antibiotic Substances: IV. Variations Among Actionomycetes, with Special Reference to Actinomyces Griseus. *Proc Natl Acad Sci U S A, 31*(5), 129-137. doi:10.1073/pnas.31.5.129

275.   Schiewe, H. J., & Zeeck, A. (1999). Cineromycins, gamma-butyrolactones and ansamycins by analysis of the secondary metabolite pattern created by a single strain of Streptomyces. *J Antibiot (Tokyo), 52*(7), 635-642. doi:10.7164/antibiotics.52.635

276.   Schindl, K., Sharma, D., & Spiteller, D. (2020). Deacylation of Calcium-Dependent Antibiotics from Streptomyces violaceoruber in Co-culture with Streptomyces sp. MG7-G1. *Chembiochem, 21*(21), 3151-3157. doi:https://doi.org/10.1002/cbic.202000404

277.   Schlatter, D., Fubuh, A., Xiao, K., Hernandez, D., Hobbie, S., & Kinkel, L. (2009). Resource amendments influence density and competitive phenotypes of Streptomyces in soil. *Microb Ecol, 57*(3), 413-420. doi:10.1007/s00248-008-9433-4

278.   Schneider, O., Simic, N., Aachmann, F. L., Ruckert, C., Kristiansen, K. A., Kalinowski, J., . . . Zotchev, S. B. (2018). Genome Mining of Streptomyces sp. YIM 130001 Isolated From Lichen Affords New Thiopeptide Antibiotic. *Front Microbiol, 9*, 3139. doi:10.3389/fmicb.2018.03139

279.   Schoenian, I., Spiteller, M., Ghaste, M., Wirth, R., Herz, H., & Spiteller, D. (2011). Chemical basis of the synergism and antagonism in microbial communities in the nests of leaf-cutting ants. *Proceedings of the National Academy of Sciences of the United States of America, 108*, 1955-1960. doi:10.1073/pnas.1008441108

280.   Schorn, M. A., Verhoeven, S., Ridder, L., Huber, F., Acharya, D. D., Aksenov, A. A., . . . van der Hooft, J. J. J. (2021). A community resource for paired genomic and metabolomic data mining. *Nature Chemical Biology, 17*(4), 363-368. doi:10.1038/s41589-020-00724-z

281.   Schroeckh, V., Scherlach, K., Nützmann, H.-W., Shelest, E., Schmidt-Heck, W., Schuemann, J., . . . Brakhage, A. A. (2009). Intimate bacterial–fungal interaction triggers biosynthesis of archetypal polyketides in *Aspergillus nidulans*. *Proceedings of the National Academy of Sciences, 106*(34), 14558-14563. doi:doi:10.1073/pnas.0901870106

282.   Schymanski, E. L., Meringer, M., & Brack, W. (2009). Matching structures to mass spectra using fragmentation patterns: are the results as good as they look? *Anal Chem, 81*(9), 3608-3617. doi:10.1021/ac802715e

283.   Seipke, R. F., Kaltenpoth, M., & Hutchings, M. I. (2012). Streptomyces as symbionts: an emerging and widespread theme? *FEMS Microbiol Rev, 36*(4), 862-876. doi:10.1111/j.1574-6976.2011.00313.x

284.   Ser, Z., Liu, X., Tang, N. N., & Locasale, J. W. (2015). Extraction parameters for metabolomics from cultured cells. *Anal Biochem, 475*, 22-28. doi:10.1016/j.ab.2015.01.003

285.   Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., . . . Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res, 13*(11), 2498-2504. doi:10.1101/gr.1239303

286.   Shao, S., Burns, D. A., Shen, H., Chen, Y., Russell, A. G., & Driscoll, C. T. (2021). The response of streams in the Adirondack region of New York to projected changes in sulfur and nitrogen deposition under changing climate. *Sci Total Environ, 800*, 149626. doi:10.1016/j.scitotenv.2021.149626

287.   Sharma, K., Palatinszky, M., Nikolov, G., Berry, D., & Shank, E. A. (2020). Transparent soil microcosms for live-cell imaging and non-destructive stable isotope probing of soil microorganisms. *eLife, 9*, e56275. doi:10.7554/eLife.56275

288.   Shen, B. (2000). Biosynthesis of Aromatic Polyketides. In F. J. Leeper & J. C. Vederas (Eds.), *Biosynthesis: Aromatic Polyketides, Isoprenoids, Alkaloids* (pp. 1-51). Berlin, Heidelberg: Springer Berlin Heidelberg.

289.   Shepherd, M. D., Kharel, M. K., Bosserman, M. A., & Rohr, J. (2010). Laboratory maintenance of Streptomyces species. *Curr Protoc Microbiol, Chapter 10*, Unit 10E 11. doi:10.1002/9780471729259.mc10e01s18

290. Sheth, R. U., Cabral, V., Chen, S. P., & Wang, H. H. (2016). Manipulating Bacterial Communities by in situ Microbiome Engineering. *Trends Genet, 32*(4), 189-200. doi:10.1016/j.tig.2016.01.005

291. Shin, J. C., Na, Z., Lee, D. H., Kim, W. C., Lee, K., Shen, Y. M., . . . Lee, J. J. (2008). Characterization of tailoring genes involved in the modification of geldanamycin polyketide in Streptomyces hygroscopicus JCM4427. *J Microbiol Biotechnol, 18*(6), 1101-1108. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/18600054

292. Siebenberg, S., Bapat, P. M., Lantz, A. E., Gust, B., & Heide, L. (2010). Reducing the variability of antibiotic production in Streptomyces by cultivation in 24-square deepwell plates. *J Biosci Bioeng, 109*(3), 230-234. doi:10.1016/j.jbiosc.2009.08.479

293. Siebenberg, S., Bapat, P. M., Lantz, A. E., Gust, B., & Heide, L. (2010). Reducing the variability of antibiotic production in Streptomyces by cultivation in 24-square deepwell plates. *Journal of Bioscience and Bioengineering, 109*(3), 230-234. doi:https://doi.org/10.1016/j.jbiosc.2009.08.479

294. Singh, R., Kumar, M., Mittal, A., & Mehta, P. K. (2017). Microbial metabolites in nutrition, healthcare and agriculture. *3 Biotech, 7*(1), 15. doi:10.1007/s13205-016-0586-4

295. Singhal, N., Kumar, M., Kanaujia, P. K., & Virdi, J. S. (2015). MALDI-TOF mass spectrometry: an emerging technology for microbial identification and diagnosis. *Frontiers in Microbiology, 6.* doi:10.3389/fmicb.2015.00791

296. Skinnider, M. A., Johnston, C. W., Gunabalasingam, M., Merwin, N. J., Kieliszek, A. M., MacLellan, R. J., . . . Magarvey, N. A. (2020). Comprehensive prediction of secondary metabolite structure and biological activity from microbial genome sequences. *Nat Commun, 11*(1), 6058. doi:10.1038/s41467-020-19986-1

297. 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 Res, 45*(W1), W49-W54. doi:10.1093/nar/gkx320

298. SMG. (1997). A sample of Penicillium mould presented by Alexander Fleming to Douglas Macleod, 1935.: The Board of Trustees of the Science Museum.

299. Soldatou, S., Eldjarn, G. H., Huerta-Uribe, A., Rogers, S., & Duncan, K. R. (2019). Linking biosynthetic and chemical space to accelerate microbial secondary metabolite discovery. *FEMS Microbiol Lett, 366*(13). doi:10.1093/femsle/fnz142

300. Song, L., Barona-Gomez, F., Corre, C., Xiang, L., Udwary, D. W., Austin, M. B., . . . Challis, G. L. (2006). Type III Polyketide Synthase β-Ketoacyl-ACP Starter Unit and Ethylmalonyl-CoA Extender Unit Selectivity Discovered by Streptomyces

coelicolor Genome Mining. *Journal of the American Chemical Society, 128*(46), 14754-14755. doi:10.1021/ja065247w

301.   Spagnuolo, C., Russo, G. L., Orhan, I. E., Habtemariam, S., Daglia, M., Sureda, A., . . . Nabavi, S. M. (2015). Genistein and Cancer: Current Status, Challenges, and Future Directions. *Advances in Nutrition, 6*(4), 408-419. doi:10.3945/an.114.008052

302.   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 Res, 36*(21), 6882-6892. doi:10.1093/nar/gkn685

303.   Staunton, J., & Weissman, K. J. (2001). Polyketide biosynthesis: a millennium review. *Natural Product Reports, 18*(4), 380-416. doi:10.1039/A909079G

304.   Stogios, P. J., & Savchenko, A. (2020). Molecular mechanisms of vancomycin resistance. *Protein Sci, 29*(3), 654-669. doi:10.1002/pro.3819

305.   Subramanian, I., Verma, S., Kumar, S., Jere, A., & Anamika, K. (2020). Multi-omics Data Integration, Interpretation, and Its Application. *Bioinform Biol Insights, 14*, 1177932219899051. doi:10.1177/1177932219899051

306.   Sun, C., Yang, Z., Zhang, C., Liu, Z., He, J., Liu, Q., . . . Ma, J. (2019). Genome Mining of Streptomyces atratus SCSIO ZH16: Discovery of Atratumycin and Identification of Its Biosynthetic Gene Cluster. *Organic Letters, 21*(5), 1453-1457. doi:10.1021/acs.orglett.9b00208

307.   Swiatek, M. A., Tenconi, E., Rigali, S., & van Wezel, G. P. (2012). Functional analysis of the N-acetylglucosamine metabolic genes of Streptomyces coelicolor and role in control of development and antibiotic production. *J Bacteriol, 194*(5), 1136-1144. doi:10.1128/JB.06370-11

308.   Swiatek, M. A., Urem, M., Tenconi, E., Rigali, S., & van Wezel, G. P. (2012). Engineering of N-acetylglucosamine metabolism for improved antibiotic production in Streptomyces coelicolor A3(2) and an unsuspected role of NagA in glucosamine metabolism. *Bioengineered, 3*(5), 280-285. doi:10.4161/bioe.21371

309.   Takeuchi, S., Hirayama, K., Ueda, K., Sakai, H., & Yonehara, H. (1958). Blasticidin S, a new antibiotic. *J Antibiot (Tokyo), 11*(1), 1-5.

310.   Tamura, K., & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol, 10*(3), 512-526. doi:10.1093/oxfordjournals.molbev.a040023

311.   Tao, W., Yang, A., Deng, Z., & Sun, Y. (2018). CRISPR/Cas9-Based Editing of Streptomyces for Discovery, Characterization, and Production of Natural Products. *Frontiers in Microbiology, 9*. doi:10.3389/fmicb.2018.01660

312.   Tarkka, M., & Hampp, R. (2008). Secondary Metabolites of Soil Streptomycetes in Biotic Interactions. In P. Karlovsky (Ed.), *Secondary Metabolites in Soil Ecology* (pp. 107-126). Berlin, Heidelberg: Springer Berlin Heidelberg.

313.   Terra, L., Ratcliffe, N., Castro, H. C., Vicente, A. C. P., & Dyson, P. (2021). Biotechnological Potential of Streptomyces Siderophores as New Antibiotics. *Curr Med Chem, 28*(7), 1407-1421. doi:10.2174/0929867327666200510235512

314.   Thomas, M. G., Chan, Y. A., & Ozanick, S. G. (2003). Deciphering tuberactinomycin biosynthesis: isolation, sequencing, and annotation of the viomycin biosynthetic gene cluster. *Antimicrob Agents Chemother, 47*(9), 2823-2830. doi:10.1128/AAC.47.9.2823-2830.2003

315.   Thomson, J. J. (1913). *Rays of positive electricity and their application to chemical analyses*: Longmans, Green and Company.

316.   Tran, P. N., Yen, M.-R., Chiang, C.-Y., Lin, H.-C., & Chen, P.-Y. (2019). Detecting and prioritizing biosynthetic gene clusters for bioactive compounds in bacteria and fungi. *Applied Microbiology and Biotechnology, 103*(8), 3277-3287. doi:10.1007/s00253-019-09708-z

317.   Tran, P. N., Yen, M. R., Chiang, C. Y., Lin, H. C., & Chen, P. Y. (2019). Detecting and prioritizing biosynthetic gene clusters for bioactive compounds in bacteria and fungi. *Appl Microbiol Biotechnol, 103*(8), 3277-3287. doi:10.1007/s00253-019-09708-z

318.   Tribalat, L., Paisse, O., Dessalces, G., & Grenier-Loustalot, M. F. (2006). Advantages of LC-MS-MS compared to LC-MS for the determination of nitrofuran residues in honey. *Anal Bioanal Chem, 386*(7-8), 2161-2168. doi:10.1007/s00216-006-0878-3

319.   Uguru, G. C., Mondhe, M., Goh, S., Hesketh, A., Bibb, M. J., Good, L., & Stach, J. E. (2013). Synthetic RNA Silencing of Actinorhodin Biosynthesis in Streptomyces coelicolor A3(2). *PLoS One, 8*(6), e67509. doi:10.1371/journal.pone.0067509

320.   Ulén, B. (2020). Nutrient leaching driven by rainfall on a vermiculite clay soil under altered management and monitored with high-frequency time resolution. *Acta Agriculturae Scandinavica, Section B — Soil & Plant Science, 70*(5), 392-403. doi:10.1080/09064710.2020.1750686

321.   USDA. (2011). Carbon to Nitrogen Ratios in Cropping Systems [Press release]. Retrieved from https://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/nrcseprd331820.pdf

322.   Vailati-Riboni, M., Palombo, V., & Loor, J. J. (2017). What Are Omics Sciences? In B. N. Ametaj (Ed.), *Periparturient Diseases of Dairy Cows: A Systems Biology Approach* (pp. 1-7). Cham: Springer International Publishing.

323.   Valiquette, L., & Laupland, K. B. (2015). Digging for new solutions. *Can J Infect Dis Med Microbiol, 26*(6), 289-290. doi:10.1155/2015/971858

324.   van der Hooft, J. J. J., Mohimani, H., Bauermeister, A., Dorrestein, P. C., Duncan, K. R., & Medema, M. H. (2020). Linking genomics and metabolomics to chart specialized metabolic diversity. *Chemical Society Reviews, 49*(11), 3297-3314. doi:10.1039/D0CS00162G

325.   van Wezel, G. P., & McDowall, K. J. (2011). The regulation of the secondary metabolism of Streptomyces: new links and experimental advances. *Nat Prod Rep, 28*(7), 1311-1333. doi:10.1039/c1np00003a

326.   Ventola, C. L. (2015). The antibiotic resistance crisis: part 1: causes and threats. *P T, 40*(4), 277-283. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/25859123

327.   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. *Microbiol Mol Biol Rev, 71*(3), 495-548. doi:10.1128/mmbr.00005-07

328.   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. doi:10.3181/00379727-45-11768

329.   Waksman, S. A., & Woodruff, H. B. (1942). Selective Antibiotic Action of Various Substances of Microbial Origin. *J Bacteriol, 44*(3), 373-384. doi:10.1128/jb.44.3.373-384.1942

330.   Waldman, A. J., Pechersky, Y., Wang, P., Wang, J. X., & Balskus, E. P. (2015). The Cremeomycin Biosynthetic Gene Cluster Encodes a Pathway for Diazo Formation. *Chembiochem, 16*(15), 2172-2175. doi:https://doi.org/10.1002/cbic.201500407

331.   Wang, B., Guo, F., Huang, C., & Zhao, H. (2020). Unraveling the iterative type I polyketide synthases hidden in <i>Streptomyces</i>. *Proceedings of the National Academy of Sciences, 117*(15), 8449-8454. doi:doi:10.1073/pnas.1917664117

332.   Wang, M., Carver, J. J., Phelan, V. V., Sanchez, L. M., Garg, N., Peng, Y., . . . Bandeira, N. (2016). Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol, 34*(8), 828-837. doi:10.1038/nbt.3597

333.   Watson, D. G. (2013). A rough guide to metabolite identification using high resolution liquid chromatography mass spectrometry in metabolomic profiling in metazoans. *Comput Struct Biotechnol J, 4*, e201301005. doi:10.5936/csbj.201301005

334. Wei, H., Lin, Z., Li, D., Gu, Q., & Zhu, T. (2010). [OSMAC (one strain many compounds) approach in the research of microbial metabolites--a review]. *Wei Sheng Wu Xue Bao, 50*(6), 701-709.

335. Weissenbach, J. (2016). The rise of genomics. *C R Biol, 339*(7-8), 231-239. doi:10.1016/j.crvi.2016.05.002

336. Wellington, E. M., Cresswell, N., & Saunders, V. A. (1990). Growth and survival of streptomycete inoculants and extent of plasmid transfer in sterile and nonsterile soil. *Appl Environ Microbiol, 56*(5), 1413-1419. doi:10.1128/aem.56.5.1413-1419.1990

337. Wendlandt, S., Li, B., Lozano, C., Ma, Z., Torres, C., & Schwarz, S. (2013). Identification of the novel spectinomycin resistance gene spw in methicillin-resistant and methicillin-susceptible Staphylococcus aureus of human and animal origin. *J Antimicrob Chemother, 68*(7), 1679-1680. doi:10.1093/jac/dkt081

338. Wiliams, S. T. (1990). Actinomycetes — The ray fungi. *Mycologist, 4*(3), 110-114. doi:https://doi.org/10.1016/S0269-915X(09)80036-0

339. Wishart, D. S. (2016). Emerging applications of metabolomics in drug discovery and precision medicine. *Nature Reviews Drug Discovery, 15*(7), 473-484. doi:10.1038/nrd.2016.32

340. Woudt, S. H. S., de Greeff, S. C., Schoffelen, A. F., Vlek, A. L. M., Bonten, M. J. M., & Group, I. D. S. I. S. A. R. S. (2017). Antibiotic Resistance and the Risk of Recurrent Bacteremia. *Clinical Infectious Diseases, 66*(11), 1651-1657. doi:10.1093/cid/cix1076

341. Xu, Y., Zhong, L.-L., Srinivas, S., Sun, J., Huang, M., Paterson, D. L., . . . Feng, Y. (2018). Spread of MCR-3 Colistin Resistance in China: An Epidemiological, Genomic and Mechanistic Study. *EBioMedicine, 34*, 139-157. doi:https://doi.org/10.1016/j.ebiom.2018.07.027

342. 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. doi:10.1111/1574-6968.12128

343. Yamada, R., Okada, D., Wang, J., Basak, T., & Koyama, S. (2021). Interpretation of omics data analyses. *Journal of Human Genetics, 66*(1), 93-102. doi:10.1038/s10038-020-0763-5

344. Yamada, Y., Kuzuyama, T., Komatsu, M., Shin-ya, K., Omura, S., Cane, D. E., & Ikeda, H. (2015). Terpene synthases are widely distributed in bacteria. *Proceedings of the National Academy of Sciences, 112*(3), 857-862. doi:doi:10.1073/pnas.1422108112

345.   Yamanaka, K., Oikawa, H., Ogawa, H. O., Hosono, K., Shinmachi, F., Takano, H., . . . Ueda, K. (2005). Desferrioxamine E produced by Streptomyces griseus stimulates growth and development of Streptomyces tanashiensis. *Microbiology (Reading), 151*(Pt 9), 2899-2905. doi:10.1099/mic.0.28139-0

346.   Yang, J. Y., Sanchez, L. M., Rath, C. M., Liu, X., Boudreau, P. D., Bruns, N., . . . Dorrestein, P. C. (2013). Molecular Networking as a Dereplication Strategy. *Journal of Natural Products, 76*(9), 1686-1699. doi:10.1021/np400413s

347.   Yates III, J. R. (1998). Mass spectrometry and the age of the proteome. *Journal of Mass Spectrometry, 33*(1), 1-19. doi:https://doi.org/10.1002/(SICI)1096-9888(199801)33:1<1::AID-JMS624>3.0.CO;2-9

348.   Yehia, R., Hathout, R. M., Attia, D. A., Elmazar, M. M., & Mortada, N. D. (2017). Anti-tumor efficacy of an integrated methyl dihydrojasmonate transdermal microemulsion system targeting breast cancer cells: In vitro and in vivo studies. *Colloids Surf B Biointerfaces, 155*, 512-521. doi:10.1016/j.colsurfb.2017.04.031

349.   Yergey, A. L., & Yergey, A. K. (1997). Preparative Scale Mass Spectrometry: A Brief History of the Calutron. *Journal of the American Society for Mass Spectrometry, 8*(9), 943-953. doi:https://doi.org/10.1016/S1044-0305(97)00123-2

350.   Yu, D., Xu, F., Zeng, J., & Zhan, J. (2012). Type III polyketide synthases in natural product biosynthesis. *IUBMB Life, 64*(4), 285-295. doi:https://doi.org/10.1002/iub.1005

351.   Yu, F., & Utsumi, R. (2009). Diversity, regulation, and genetic manipulation of plant mono- and sesquiterpenoid biosynthesis. *Cell Mol Life Sci, 66*(18), 3043-3052. doi:10.1007/s00018-009-0066-7

352.   Yu, M., Li, Y., Banakar, S. P., Liu, L., Shao, C., Li, Z., & Wang, C. (2019). New Metabolites From the Co-culture of Marine-Derived Actinomycete Streptomyces rochei MB037 and Fungus Rhinocladiella similis 35. *Frontiers in Microbiology, 10.* doi:10.3389/fmicb.2019.00915

353.   Zarins-Tutt, J. S., Barberi, T. T., Gao, H., Mearns-Spragg, A., Zhang, L., Newman, D. J., & Goss, R. J. M. (2016). Prospecting for new bacterial metabolites: a glossary of approaches for inducing, activating and upregulating the biosynthesis of bacterial cryptic or silent natural products. *Natural Product Reports, 33*(1), 54-72. doi:10.1039/C5NP00111K

354.   Zhang, D., Wang, J., Qiao, Y., Lin, B., Deng, Z., Kong, L., & You, D. (2022). Genome Mining and Metabolic Profiling Reveal Cytotoxic Cyclodipeptides in Streptomyces hygrospinosus var. Beijingensis. *Antibiotics (Basel), 11*(11). doi:10.3390/antibiotics11111463

355.   Zhang, Z., Du, C., Barsy, F. d., Liem, M., Liakopoulos, A., Wezel, G. P. v., . . . Rozen, D. E. (2020). Antibiotic production in <i>Streptomyces</i> is organized by a division of labor through terminal genomic differentiation. *Science Advances, 6*(3), eaay5781. doi:doi:10.1126/sciadv.aay5781

356.   Zorzet, A. (2014). Overcoming scientific and structural bottlenecks in antibacterial discovery and development. *Ups J Med Sci, 119*(2), 170-175. doi:10.3109/03009734.2014.897277

Appendices

Supplementary Table 3.1 – Genome size and BGC content as predicted by antiSMASH v6.0.1 for the nine strains encompassed in the study.

Strain	NCBI Accession No.	Genome Size (Mbp)	No. BGCs
Micromonospora chalcea	MAGP01000001.1	7	22
Micromonospora echinospora	LT607413.1	7.78	28
Nocardia farcinica	CP031418.1	6.43	21
Streptomyces lividans 1326	NZ_CM001889.1	8.5	28
Streptomyces coelicolor M145	AL645882.2	9.05	27
Streptomyces scabiei 87.22	FN554889.1	10.15	34
Streptomyces collinus Tu 365	CP006259.1	8.38	32
Streptomyces venezuelae	NC_018750.1	8.23	30
Streptomyces clavuligerus	CP027858.1	8.54	25

Supplementary Table 3.2 – Percentage sequence similarity within the nine strains. Sequence homology to BGCs that have been experimentally verified and submitted to the MIBiG repository (Kautsar *et al*., 2020) as predicted by antiSMASH v6.0.1 (K. Blin *et al*., 2021).

Strain	NCBI Accession No.	Sequence Similarity to BGCs Submitted to MIBiG	
		0%	1% - 100%
Micromonospora chalcea	MAGP01000001.1	5	17
Micromonospora echinospora	LT607413.1	7	21
Nocardia farcinica	CP031418.1	4	17
Streptomyces lividans 1326	NZ_CM001889.1	5	18
Streptomyces coelicolor M145	AL645882.2	4	24
Streptomyces scabiei 87.22	FN554889.1	6	27
Streptomyces collinus Tu 365	CP006259.1	4	28
Streptomyces venezuelae	NC_018750.1	2	27
Streptomyces clavuligerus	CP027858.1	7	18
	Total	44	196

Supplementary Table 3.3 – BGC suites of selected strains as predicted by antiSMASH v6.0.1 (K. Blin *et al*., 2021). 'MSKC' stands for Most Similar Known Cluster, as predicted by comparison to MIBiG database (Medema *et al*., 2015).

Supplementary Table 3.3.1: *S. platensis* ATCC 23948

Region	Type	From	To	MSKC	MSKC type	Similarity
Region 1	terpene	18,292	37,716	cinnamycin	RiPP:Lanthipeptide	14%
Region 2	arylpolyene	51,750	93,045	chloramphenicol	NRP	11%
Region 3	CDPS,T1PKS,NRPS	347,542	406,917	jomthonic acid A / jomthonic acid B / jomthonic acid C	NRP	24%
Region 4	terpene	431,321	451,285	blasticidin S	Other	10%
Region 5	NRPS,thiopeptide,LAP	458,863	502,048	muraymycin C1	NRP + Polyketide	23%
Region 6	NRPS,betalactone	517,778	569,059	blasticidin S	Other	7%
Region 7	T3PKS,lanthipeptide-class-iii	656,362	706,046	SapB	RiPP:Lanthipeptide	100%
Region 8	NAPAA	795,527	831,637	stenothricin	NRP:Cyclic depsipeptide	13%
Region 9	terpene,RiPP-like	849,959	875,240	herbimycin A	Polyketide	16%
Region 10	T2PKS	903,653	974,275	spore pigment	Polyketide	83%
Region 11	melanin,T1PKS,NRPS	1,002,358	1,144,981	lydicamycin	NRP + Polyketide:Modular type I	96%
Region 12	siderophore	1,291,827	1,303,525			
Region 13	siderophore	2,061,841	2,070,635	desferrioxamine E	Other	100%
Region 14	ectoine	2,156,241	2,166,657	ectoine	Other	100%
Region 15	T2PKS	2,616,248	2,687,488	JBIR-76 / JBIR-77	Polyketide	64%
Region 16	T1PKS	3,148,327	3,194,374	hormaomycin / hormaomycin A1 / hormaomycin A2 / hormaomycin A3 / hormaomycin A4 / hormaomycin A5 / hormaomycin A6	NRP:Cyclic depsipeptide	13%
Region 17	terpene	3,262,152	3,281,441	ebelactone	Polyketide	5%
Region 18	terpene	4,697,319	4,717,974	salinomycin	Polyketide:Modular type I	6%
Region 19	NRPS,lassopeptide,NRPS-like	4,869,768	4,939,963	A54145	NRP	15%
Region 20	siderophore	6,451,197	6,465,909	ficellomycin	NRP	3%
Region 21	terpene	6,625,459	6,644,599	arixanthomycin A /	Polyketide:Type II + Saccharide:Hybrid/tailoring	5%

Region		From	To			
				arixanthomycin B / arixanthomycin C		
Region 22	butyrolactone	6,657,273	6,667,287			
Region 23	RiPP-like	6,717,081	6,727,015			
Region 24	butyrolactone	7,012,161	7,021,843			
Region 25	terpene	7,328,335	7,353,315	hopene	Terpene	61%
Region 26	hglE-KS,T1PKS	7,450,400	7,502,208	herboxidiene	Polyketide	4%
Region 27	NRPS-like,NRPS	7,568,650	7,616,981	daptomycin	NRP	6%
Region 28	lanthipeptide-class-i	7,666,330	7,690,879	oxalomycin B	NRP + Polyketide	9%
Region 29	RiPP-like	7,769,126	7,781,060	ansamitocin P-3	Polyketide	4%
Region 30	T2PKS	7,788,097	7,860,654	oxytetracycline	Polyketide	95%
Region 31	NRPS-like	7,961,199	8,004,232	Sch-47554 / Sch-47555	Polyketide	5%
Region 32	butyrolactone	8,039,125	8,050,108	coelimycin P1	Polyketide:Modular type I	12%
Region 33	NRPS-like,NRPS,other	8,111,908	8,167,628	deimino-antipain	NRP	66%
Region 34	terpene	8,197,932	8,223,582	isorenieratene	Terpene	100%
Region 35	NRPS,T1PKS	8,372,638	8,428,652	JBIR-06	NRP + Polyketide	66%

Supplementary Table 3.3.2: *S. noursei* ATCC 11455

Region	Type	From	To	MSKC	MSKC type	Similarity
Region 1	T1PKS,NRPS-like,terpene	78,726	192,583	natamycin	Polyketide	68%
Region 2	NRPS-like	200,625	241,100	tiacumicin B	Polyketide:Modular type I	9%
Region 3	lanthipeptide-class-iii,NRPS	253,400	306,411	s56-p1	NRP	17%
Region 4	T1PKS	331,336	375,420			
Region 5	terpene,T3PKS,NRPS,T1PKS, NRPS-like	383,541	464,447	xiamycin A	Terpene	9%
Region 6	hglE-KS,T1PKS	721,734	769,431			
Region 7	NRPS	1,202,574	1,263,050	cadaside A / cadaside B	NRP	19%
Region 8	NRPS-like	1,277,166	1,319,125	caniferolide A / caniferolide B / caniferolide C / caniferolide D	Polyketide:Modular type I	3%
Region 9	NRPS-like,transAT-PKS	1,322,214	1,391,499	cycloheximide	Polyketide:Trans-AT type I	50%
Region 10	hglE-KS	1,414,678	1,456,930	zorbamycin	NRP:Glycopeptide + Polyketide:Modular type I + Saccharide:Hybrid/tailoring	4%
Region 11	T1PKS	1,493,678	1,558,816	piericidin A1	Polyketide	50%
Region 12	RiPP-like	1,823,021	1,830,465	conglobatin	NRP	10%

Region	Type	Start	End	Most similar known cluster	Class	Similarity
Region 13	T2PKS	1,863,262	1,935,777	spore pigment	Polyketide	83%
Region 14	terpene	1,977,201	1,997,364	herbimycin A	Polyketide	10%
Region 15	lanthipeptide-class-i	2,555,618	2,577,864			
Region 16	siderophore	2,817,815	2,826,271	desferrioxamine E	Other	100%
Region 17	ectoine	2,907,414	2,917,818	ectoine	Other	100%
Region 18	lanthipeptide-class-i	3,888,531	3,913,005			
Region 19	terpene	5,333,616	5,354,322	salinomycin	Polyketide:Modular type I	6%
Region 20	thiopeptide,RiPP-like,PKS-like	5,434,803	5,485,839	radamycin / globimycin	RiPP	88%
Region 21	linaridin	5,496,843	5,517,262	legonaridin	RiPP	66%
Region 22	NAPAA	5,954,858	5,988,224			
Region 23	T1PKS	6,082,295	6,187,762	aldgamycin J / aldgamycin K / aldgamycin P / aldgamycin E	Polyketide	47%
Region 24	CDPS	6,664,526	6,685,176	albonoursin	Other	83%
Region 25	siderophore	7,011,652	7,026,331	ficellomycin	NRP	3%
Region 26	NRPS,betalactone	7,033,994	7,094,730	ulleungmycin	NRP	77%
Region 27	T3PKS	7,154,221	7,195,276	naringenin	Terpene	100%
Region 28	RiPP-like	7,302,922	7,312,793			
Region 29	lanthipeptide-class-iv	7,482,842	7,505,598	toxoflavin / fervenulin	Other	7%
Region 30	arylpolyene,T1PKS	7,517,546	7,601,648	ansamitocin P-3	Polyketide	4%
Region 31	terpene	7,691,826	7,712,591	ebelactone	Polyketide	5%
Region 32	other,NRPS,lanthipeptide-class-ii	7,734,627	7,807,150	A-503083 A / A-503083 B / A-503083 E / A-503083 F	NRP	7%
Region 33	terpene	7,983,923	8,003,502			
Region 34	RRE-containing	8,003,648	8,022,882			
Region 35	T1PKS,RRE-containing	8,064,306	8,208,199	nystatin A1	Polyketide:Modular type I + Saccharide:Hybrid/tailoring	100%

Region 36	butyrolactone	8,539,920	8,550,315	coelimycin P1	Polyketide:Modular type I	8%
Region 37	RiPP-like,butyrolactone,NAPAA	8,856,149	8,899,966	Sch-47554 / Sch-47555	Polyketide	10%
Region 38	NRPS-like	8,972,099	9,015,383	abyssomicin C / atrop-abyssomicin C	Polyketide:Modular type I	10%
Region 39	terpene	9,229,738	9,256,445	hopene	Terpene	61%
Region 40	lassopeptide	9,442,074	9,464,608	citrulassin A	RiPP	50%
Region 41	linaridin	9,505,864	9,526,475	pentostatine / vidarabine	Other	9%
Region 42	hglE-KS	9,556,696	9,602,914	combamide	NRP + Polyketide	22%
Region 43	T1PKS	9,672,100	9,748,795	tetronasin	Polyketide	19%

Supplementary Table 3.3.3: *S. nodosus* ATCC 14899

Region	Type	From	To	MSKC	MSKC type	Similarity
Region 1	amglyccycl	8,307	29,554	validamycin A	Polyketide + Other:Cyclitol	44%
Region 2	nucleoside	306,881	327,585	sanglifehrin A	NRP + Polyketide	4%
Region 3	RiPP-like	366,664	375,864			
Region 4	T1PKS	566,779	709,220	nystatin A1	Polyketide:Modular type I + Saccharide:Hybrid/tailoring	77%
Region 5	NAPAA	843,702	874,965	kanamycin	Saccharide	1%
Region 6	NRPS,NAPAA	879,135	920,041	stenothricin	NRP:Cyclic depsipeptide	13%
Region 7	T3PKS	1,264,922	1,305,465	flaviolin	Other	75%
Region 8	ectoine	1,973,875	1,984,270	ectoine	Other	100%
Region 9	siderophore	2,857,059	2,867,873	desferrioxamin B / desferrioxamin e E	Other	83%
Region 10	T2PKS	3,487,421	3,558,890	spore pigment	Polyketide	83%
Region 11	terpene	4,954,743	4,974,174	julichrome Q3-3 / julichrome Q3-5	Polyketide	25%
Region 12	siderophore	5,554,171	5,564,965			

Region 13	RiPP-like	5,813,635	5,825,062			
Region 14	terpene,lanthipeptide-class-i	5,849,040	5,879,160	geosmin	Terpene	100%
Region 15	siderophore	5,959,715	5,972,976	paulomycin	Other	3%
Region 16	terpene,NRPS,arylpolyene,ladderane	6,495,185	6,605,616	RP-1776	Polyketide + NRP:Cyclic depsipeptide	57%
Region 17	T3PKS,T2PKS,oligosaccharide, NRPS, terpene	6,648,291	6,771,698	saprolmycin E	Polyketide	86%
Region 18	NRPS-like	6,859,901	6,902,300	paromomycin	Saccharide	5%
Region 19	NRPS	7,554,781	7,607,104	rimosamide	NRP	21%
Region 20	NRPS	7,655,753	7,706,669	coelichelin	NRP	100%
Region 21	lassopeptide	7,726,589	7,749,103	siamycin I	RiPP	56%
Region 22	butyrolactone	7,758,849	7,769,841	neocarzinostatin	Polyketide:Iterative type I + Polyketide:Enediyne type I	4%

Supplementary Table 3.3.4: *S. spectabilis* ATCC 27465

Region	Type	From	To	MSKC	MSKC type	Similarity
Region 1	butyrolactone	107,498	115,888			
Region 2	PKS-like,blactam	117,602	156,911	valclavam / (-)-2-(2-hydroxyethyl)clavam	Other:Non-NRP beta-lactam	57%
Region 3	T1PKS	162,483	207,224	A83543A	Polyketide	8%
Region 4	transAT-PKS,NRPS,lassopeptide	215,598	279,431	lagmysin	RiPP	80%
Region 5	transAT-PKS,PKS-like,T1PKS,NRPS	285,784	351,455			
Region 6	other	386,027	426,101	LL-D49194α1 (LLD)	Polyketide	3%
Region 7	NRPS	433,756	497,923	lysocin	NRP	14%
Region 8	NRPS,lanthipeptide-class-i,CDPS	525,671	581,503	triostin A	NRP	11%
Region 9	T1PKS,lanthipeptide-class-iii	729,103	786,857	calicheamicin	Polyketide	8%
Region 10	transAT-PKS,T3PKS,PKS-like,	869,624	969,589	kalimantacin A	NRP + Polyketide:Modular type I + Polyketide:Trans-AT type I	17%
Region 11	T1PKS	1,073,676	1,187,790	linfuranone B / linfuranone C	Polyketide	46%

Region 12	terpene	1,203,262	1,221,838	telomycin	NRP	8%
Region 13	NRPS	1,228,600	1,289,798	chondrochloren A	NRP + Polyketide:Modular type I	16%
Region 14	arylpolyene,ladderane	1,341,854	1,385,133	WS9326	NRP	5%
Region 15	T3PKS	1,601,644	1,642,333	BE-14106	Polyketide:Modular type I	17%
Region 16	NRPS	1,694,698	1,756,190	cadaside A / cadaside B	NRP	9%
Region 17	ectoine	2,555,196	2,565,600	ectoine	Other	100%
Region 18	terpene	3,125,710	3,144,567			
Region 19	T1PKS	3,252,364	3,334,738	herbimycin A	Polyketide	33%
Region 20	terpene,NRPS	3,471,376	3,526,897	abyssomicin M / abyssomicin N / abyssomicin O / abyssomicin P / abyssomicin Q / abyssomicin R / abyssomicin S / abyssomicin T / abyssomicin U / abyssomicin V / abyssomicin W / abyssomicin X	Polyketide	6%
Region 21	melanin	4,000,560	4,008,219	melanin	Other	28%
Region 22	T2PKS	4,361,674	4,434,171	julichrome Q3-3 / julichrome Q3-5	Polyketide	33%
Region 23	melanin	5,001,328	5,011,720	istamycin	Saccharide	4%
Region 24	terpene	6,274,672	6,294,044	albaflavenone	Terpene	100%
Region 25	siderophore	6,962,082	6,976,560	ficellomycin	NRP	3%
Region 26	RiPP-like	7,245,160	7,255,639			
Region 27	NRPS,RiPP-like	7,364,787	7,440,020	kutzneride 2	NRP	17%
Region 28	redox-cofactor	7,643,537	7,665,668	kanamycin	Saccharide	2%
Region 29	terpene	7,723,522	7,743,821	geosmin	Terpene	100%
Region 30	terpene	7,845,069	7,869,390	hopene	Terpene	92%
Region 31	other,T1PKS,NRPS-like,prodigiosin	7,874,787	7,942,911	undecylprodigiosin	NRP + Polyketide	90%
Region 32	T1PKS	7,957,995	8,031,435	spectinabilin / orinocin /	Polyketide:Modular type I	90%

Region	Type	From	To	MSKC	MSKC type	Similarity
				SNF4435C / SNF4435D		
Region 33	T1PKS	8,047,124	8,193,590	ibomycin	Polyketide	25%
Region 34	butyrolactone,T1PKS,NRPS-like	8,211,145	8,312,939	streptovaricin	Polyketide	95%
Region 35	NRPS-like,NRPS	8,413,778	8,465,573	zorbamycin	NRP:Glycopeptide + Polyketide:Modular type I + Saccharide:Hybrid/tailoring	8%
Region 36	NRPS	8,481,324	8,524,083	incednine	Polyketide	2%
Region 37	amglyccycl,NRPS,NRPS-like,T1PKS	8,577,339	8,686,353	actinospectacin	Saccharide	100%
Region 38	NRPS-like,NRPS	8,718,752	8,793,220	stenothricin	NRP:Cyclic depsipeptide	36%
Region 39	lanthipeptide-class-iv	9,012,718	9,035,438	blasticidin S	Other	7%
Region 40	NRPS	9,088,806	9,168,348	mirubactin	NRP	50%
Region 41	NRPS	9,229,627	9,279,536	coelichelin	NRP	100%
Region 42	RiPP-like	9,394,818	9,406,740			
Region 43	terpene	9,564,781	9,584,369	ebelactone	Polyketide	5%
Region 44	lanthipeptide-class-ii	9,591,999	9,614,878	akaeolide	Polyketide	8%
Region 45	NRPS	9,615,789	9,694,321	herboxidiene	Polyketide	3%
Region 46	T1PKS,T3PKS,terpene,butyrolactone	9,707,266	9,798,331	lavendiol	Polyketide	6%

Supplementary Table 3.3.5: *S. kanamyceticus* ATCC 12853

Region	Type	From	To	MSKC	MSKC type	Similarity
Region 1	T1PKS	50,744	144,793	lasalocid	Polyketide	72%
Region 2	PKS-like,NRPS,NRPS-like	387,310	474,885	A-201A	Other	12%
Region 3	T1PKS,NRPS-like,NRPS	522,473	620,271	FR-900520	NRP + Polyketide	68%
Region 4	T2PKS,NRPS,T3PKS	641,611	781,334	A-47934	NRP:Glycopeptide	47%
Region 5	T1PKS	850,440	977,587	streptovaricin	Polyketide	31%
Region 6	lassopeptide	1,051,670	1,073,872	anantin B1 / anantin B2	RiPP	60%
Region 7	T1PKS	1,075,231	1,121,820	rifamorpholine A / rifamorpholine B / rifamorpholine C / rifamorpholine D / rifamorpholine E	Polyketide	4%
Region 8	CDPS,terpene	1,151,786	1,192,294	hopene	Terpene	92%

Region 9	other	1,359,971	1,401,363	capreomycin IA / capreomycin IB / capreomycin IIA / capreomycin IIB	NRP	6%
Region 10	redox-cofactor	1,473,874	1,495,946	lankacidin C	NRP + Polyketide	20%
Region 11	amglyccycl	1,704,759	1,725,200	kanamycin	Saccharide	50%
Region 12	NRPS	1,773,103	1,818,411	colabomycin E	Polyketide:Type II	13%
Region 13	T2PKS,NRPS-like	1,819,922	1,922,687	formicamycins A-M	Polyketide	81%
Region 14	terpene	2,339,398	2,354,465	geosmin	Terpene	100%
Region 15	NRPS,T3PKS	2,375,587	2,489,798	CDA1b / CDA2a / CDA2b / CDA3a / CDA3b / CDA4a / CDA4b	NRP:Ca+-dependent lipopeptide	42%
Region 16	RiPP-like	2,508,625	2,519,284			
Region 17	furan,butyrolactone	2,602,020	2,625,105	methylenomycin A	Other	28%
Region 18	siderophore	2,824,712	2,838,331			
Region 19	terpene	3,533,937	3,552,892	albaflavenone	Terpene	100%
Region 20	NRPS,ladderane	4,406,445	4,458,613	heme D1	Other	11%
Region 21	lanthipeptide-class-i	4,734,203	4,757,249	kanamycin	Saccharide	1%
Region 22	LAP	4,981,022	5,003,333	granaticin	Polyketide:Type II	8%
Region 23	NRPS	5,790,061	5,836,148	streptobactin	NRP	76%
Region 24	thiopeptide,LAP	6,041,031	6,064,934			
Region 25	siderophore	6,437,261	6,449,027	desferrioxamin B / desferrioxamine E	Other	83%
Region 26	melanin	6,532,122	6,542,640	istamycin	Saccharide	5%
Region 27	lanthipeptide-class-iii	7,049,363	7,072,086	SapB	RiPP:Lanthipeptide	75%
Region 28	oligosaccharide,LAP,T1PKS,NRPS	7,098,475	7,163,570	landomycin A	Polyketide:Type II + Saccharide:Hybrid/tailoring	15%
Region 29	ectoine	7,745,330	7,755,734	ectoine	Other	100%
Region 30	T1PKS,transAT-PKS	8,366,405	8,462,361	tetrocarcin A	Polyketide	24%
Region 31	NRPS	8,789,320	8,847,690	taromycin A	NRP:Ca+-dependent lipopeptide	10%
Region 32	NRPS	8,993,483	9,062,070	CDA1b / CDA2a / CDA2b / CDA3a / CDA3b / CDA4a / CDA4b	NRP:Ca+-dependent lipopeptide	7%
Region 33	amglyccycl,T2PKS,terpene,PKS-like,other,oligosaccharide	9,075,451	9,190,481	cinerubin B	Polyketide:Type II	74%
Region 34	NRPS-like,NRPS	9,241,325	9,322,307	daptomycin	NRP	15%
Region 35	RiPP-like,T3PKS	9,456,717	9,502,870	violapyrone B	Polyketide	28%
Region 36	NRPS,lanthipeptide-class-i	9,503,075	9,586,275	avilamycin A / avilamycin C	Saccharide:Oligosaccharide	5%
Region 37	NRPS	9,594,778	9,647,952	rimosamide	NRP	14%
Region 38	lanthipeptide-class-i,T3PKS,other	9,771,334	9,844,286	A-503083 A / A-503083 B / A-503083 E / A-503083 F	NRP	5%
Region 39	blactam,T1PKS,NRPS-like	9,921,273	9,969,483	valclavam / (-)-2-(2-	Other:Non-NRP beta-lactam	50%

Region	Type	From	To	MSKC	MSKC type	Similarity
				hydroxyethyl)clav am		
Region 40	NRPS-like,T3PKS,terpene,lassope ptide	10,045,09 6	10,128,06 8	anantin C	RiPP	75%

Supplementary Table 3.3.6: *S. fradiae* ATCC 10745

Region	Type	From	To	MSKC	MSKC type	Similarity
Region 1	PKS-like	1	24,376	β-D-galactosylvalidoxylamin e-A	Saccharide	13%
Region 2	terpene	147,427	167,254	geosmin	Terpene	100%
Region 3	RRE-containing	279,965	300,150			
Region 4	terpene	310,810	330,174	ikarugamycin	NRP + Polyketide:Iterative type I	8%
Region 5	terpene	413,402	434,453			
Region 6	NRPS-like	496,251	536,868	desotamide	NRP	9%
Region 7	thioamide-NRP	542,475	594,171	auroramycin	Polyketide	5%
Region 8	lanthipeptide-class-i	614,676	640,644	akaeolide	Polyketide	12%
Region 9	indole	662,337	685,768	staurosporine	Alkaloid	60%
Region 10	lassopeptide	1,118,75 1	1,140,37 7			
Region 11	ectoine	1,300,08 1	1,310,47 9	ectoine	Other	100%
Region 12	siderophore	2,301,41 2	2,312,81 3	desferrioxamin B	Other	100%
Region 13	melanin	3,703,32 6	3,713,93 7	istamycin	Saccharide	11%
Region 14	RiPP-like	3,841,19 4	3,851,09 4			
Region 15	NRPS	5,270,07 8	5,330,29 9	dechlorocuracomycin	NRP	20%
Region 16	T1PKS,NRPS-like	5,360,12 5	5,425,74 2	bleomycin	NRP:Glycopeptide + Polyketide:Modular type I + Saccharide:Hybrid/tailori ng	9%
Region 17	LAP,thiopeptide,NRP S-like	5,505,13 2	5,570,41 6	salinomycin	Polyketide:Modular type I	6%
Region 18	RiPP-like	5,673,04 6	5,684,47 6			
Region 19	NRPS-like	5,749,73 7	5,791,52 8	indigoidine	Saccharide	24%
Region 20	siderophore	5,808,50 0	5,824,11 3	ficellomycin	NRP	10%
Region 21	CDPS	5,871,21 4	5,892,10 1	kanamycin	Saccharide	1%

Region	Type	From	To	MSKC	MSKC type	Similarity
Region 22	NRPS	5,914,126	5,959,343	rubrolone_A / rubrolone_B	Alkaloid	11%
Region 23	lassopeptide	6,051,987	6,074,319			
Region 24	terpene	6,121,727	6,147,439	hopene	Terpene	69%
Region 25	terpene	6,169,627	6,193,098	isorenieratene	Terpene	100%
Region 26	amglyccycl	6,215,716	6,250,147	neomycin	Saccharide	60%
Region 27	terpene	6,401,702	6,422,817			
Region 28	lanthipeptide-class-ii,NRPS	6,456,826	6,511,202	incednine	Polyketide	2%
Region 29	NRPS-like,butyrolactone,T3PKS	6,584,238	6,660,692	lomofungin	Other	26%
Region 30	NRPS-like,NRPS,betalactone	6,672,072	6,725,579	teleocidin B1	NRP + Terpene	75%

Supplementary Table 3.3.7: *S. vinaceus* ATCC 27476

Region	Type	From	To	MSKC	MSKC type	Similarity
Region 1	lanthipeptide-class-iv	27,216	50,137			
Region 2	terpene	295,776	315,327	avermitilol	Terpene	100%
Region 3	terpene,hglE-KS,T1PKS,CDPS	543,257	620,186	ebelactone	Polyketide	5%
Region 4	terpene	641,978	660,423	oxalomycin B	NRP + Polyketide	12%
Region 5	T3PKS	723,314	763,639	alkylresorcinol	Polyketide	100%
Region 6	siderophore	795,773	808,803			
Region 7	melanin	877,417	906,864	melanin	Other	28%
Region 8	terpene	910,830	930,437	monensin	Polyketide	5%
Region 9	lanthipeptide-class-iii	1,042,945	1,065,590	SapB	RiPP:Lanthipeptide	100%
Region 10	terpene,T3PKS	1,240,059	1,291,960	geosmin	Terpene	100%
Region 11	terpene	1,294,477	1,321,123	hopene	Terpene	61%
Region 12	terpene	1,578,488	1,599,843	toxoflavin / fervenulin	Other	14%
Region 13	NRPS,lanthipeptide-class-i,lanthipeptide-class-ii,RiPP-like	1,660,217	1,722,360	ashimides	NRP	12%
Region 14	siderophore	1,974,766	1,987,824	ficellomycin	NRP	3%
Region 15	linaridin	2,473,409	2,493,996	pentostatine / vidarabine	Other	9%
Region 16	butyrolactone,RRE-containing	2,498,874	2,528,367	granaticin	Polyketide:Type II	8%

Region 17	siderophore	4,749,446	4,761,236	desferrioxamin B	Other	100%
Region 18	NRPS,terpene	6,235,140	6,319,607	atratumycin	NRP	13%
Region 19	NRPS,NRPS-like,blactam	6,379,930	6,429,206	lipopeptide 8D1-1 / lipopeptide 8D1-2	NRP	4%
Region 20	NAPAA	6,946,799	6,980,704			
Region 21	thiopeptide,LAP	7,039,541	7,069,265	lactazole	RiPP:Thiopeptide	66%
Region 22	NRPS-like,T1PKS	7,369,613	7,415,182	miharamycin A / miharamycin B	Polyketide	88%
Region 23	lassopeptide	7,569,637	7,592,091	vazabitide A	NRP	13%

Supplementary Table 3.3.8: *S. clavuligerus* ATCC 27064

Region	Type	From	To	MSKC	MSKC type	Similarity
Region 1	siderophore	12,509	25,247			
Region 2	NRPS-like,other,T1PKS,NRPS,terpene	78,305	175,260	pactamides	NRP + Polyketide	55%
Region 3	T3PKS,lanthipeptide	278,060	319,115	naringenin	Terpene	100%
Region 4	NRPS,NRPS-like	460,004	508,289	nucleocidin	Other	47%
Region 5	NRPS	527,593	570,751	holomycin	NRP	100%
Region 6	terpene	578,333	604,898	hopene	Terpene	69%
Region 7	NRPS	685,288	726,057			
Region 8	redox-cofactor	750,603	772,973	lankacidin C	NRP + Polyketide	20%
Region 9	lanthipeptide-class-iii,T2PKS	926,046	1,019,570	spore pigment	Polyketide	83%
Region 10	RiPP-like	1,041,263	1,052,636			
Region 11	NRPS	1,165,379	1,241,466	A-201A	Other	15%
Region 12	siderophore	1,292,298	1,301,992			
Region 13	NRPS	1,544,099	1,591,419	kanamycin	Saccharide	1%
Region 14	lanthipeptide-class-i	1,655,098	1,678,971			
Region 15	nucleoside	1,782,692	1,803,654	tunicamycin B1	Other:Nucleoside	85%
Region 16	NRPS,blactam	1,861,960	1,912,502	cephamycin C	NRP:Beta-lactam	84%
Region 17	melanin	2,231,583	2,242,017	melanin	Other	100%
Region 18	blactam	3,301,354	3,322,328	alanylclavam / 2-hydroxymethylclavam / 2-formyloxymethylclava	Other:Non-NRP beta-lactam	75%

				m / clavam-2-carboxylate		
Region 19	lanthipeptide-class-i	3,834,469	3,859,566			
Region 20	butyrolactone	4,012,162	4,021,865	lactonamycin	Polyketide	3%
Region 21	NRPS	4,028,923	4,076,251			
Region 22	siderophore	4,431,537	4,443,411	desferrioxamin B	Other	100%
Region 23	ectoine	5,423,823	5,434,248	ectoine	Other	100%
Region 24	PKS-like,LAP,butyrolactone,T1PKS	6,066,295	6,148,898	4-hexadecanoyl-3-hydroxy-2-(hydroxymethyl)-2H-furan-5-one	Polyketide	63%
Region 25	terpene	6,504,864	6,527,038	geosmin	Terpene	100%
Region 26	T1PKS	6,654,213	6,748,591	bafilomycin B1	Polyketide:Modular type I	66%

Supplementary Table 3.3.9: *S. venezuelae* ATCC 10712

Region	Type	From	To	MSKC	MSKC type	Similarity
Region 1	ectoine	237,663	248,079	ectoine	Other	100%
Region 2	terpene	274,351	295,302	geosmin	Terpene	100%
Region 3	T1PKS,NRPS-like,T3PKS,NRPS	503,449	603,058	venemycin	Polyketide	100%
Region 4	lanthipeptide-class-ii,terpene	613,669	643,318	chrysomycin	Polyketide	5%
Region 5	lanthipeptide-class-iv	706,670	729,522	venezuelin	RiPP:Lanthipeptide	75%
Region 6	indole	866,720	889,926	rebeccamycin	Other:Aminocoumarin	25%
Region 7	NRPS-like	1,030,635	1,072,739	chloramphenicol	NRP	100%
Region 8	CDPS	2,070,715	2,091,455	malacidin A / malacidin B	NRP:Ca+-dependent lipopeptide	5%
Region 9	siderophore	2,799,249	2,810,183	desferrioxamin B	Other	100%
Region 10	lassopeptide	3,411,369	3,433,728	albusnodin	RiPP	100%
Region 11	NRPS-like	4,408,662	4,450,901	lactonamycin	Polyketide	10%
Region 12	butyrolactone	4,522,587	4,531,360	scleric acid	NRP	29%
Region 13	melanin	5,002,462	5,010,679	istamycin	Saccharide	8%
Region 14	other,butyrolactone	5,475,955	5,517,060	A-factor	Other	100%
Region 15	LAP,thiopeptide	5,526,152	5,559,470	BD-12	NRP	17%

Region	Type			MSKC	MSKC type	
Region 16	T3PKS	5,784,770	5,823,048	flaviolin	Other	50%
Region 17	siderophore	5,873,375	5,885,837	murayaquinone	Polyketide	6%
Region 18	siderophore	5,938,831	5,953,252	ficellomycin	NRP	3%
Region 19	RiPP-like	6,353,037	6,363,866			
Region 20	T2PKS,butyrolactone	6,474,034	6,545,829	auricin	Polyketide:Type II + Saccharide:Hybrid/tailoring	55%
Region 21	NAPAA	6,676,752	6,710,654	formicamycins A-M	Polyketide	18%
Region 22	NRPS,NRPS-like,RRE-containing,ladderane	6,732,363	6,854,491	atratumycin	NRP	34%
Region 23	terpene	7,021,150	7,046,269	hopene	Terpene	69%
Region 24	lanthipeptide-class-iii	7,062,113	7,084,740	SapB	RiPP:Lanthipeptide	100%
Region 25	RiPP-like	7,127,836	7,138,690			
Region 26	T2PKS	7,404,285	7,476,797	spore pigment	Polyketide	83%
Region 27	melanin	7,482,644	7,493,033	melanin	Other	28%
Region 28	NRPS	7,704,842	7,757,706	salinichelins	NRP	61%
Region 29	terpene	7,786,268	7,806,390	2-methylisoborneol	Terpene	100%
Region 30	T3PKS	7,943,622	7,984,779	alkylresorcinol	Polyketide	100%
Region 31	terpene,NRPS	8,186,751	8,223,505			

Supplementary Table 3.3.10: *S. rimosus* ATCC 10970

Region	Type	From	To	MSKC	MSKC type	Similarity
Region 1	NRPS-like	74,142	115,888	paromomycin	Saccharide	7%
Region 2	NRPS,T1PKS,terpene	176,663	266,622	RP-1776	Polyketide + NRP:Cyclic depsipeptide	8%
Region 3	T1PKS	306,755	365,656	sceliphrolactam	Polyketide	32%
Region 4	T1PKS,NRPS-like	388,252	507,432	nystatin A1	Polyketide:Modular type I + Saccharide:Hybrid/tailoring	72%
Region 5	NRPS	517,170	561,063	ginichelins	NRP	22%
Region 6	lassopeptide	574,136	595,276	lagmysin	RiPP	80%

Region	Type	Start	End	Most similar known cluster	Class	Similarity
Region 7	T2PKS,terpene	596,516	672,154	oxytetracycline	Polyketide	95%
Region 8	T1PKS	767,321	814,231	A54145	NRP	3%
Region 9	lanthipeptide-class-iii	892,714	914,112			
Region 10	T1PKS	914,817	959,552	spiroindimicin A / spiroindimicin B / spiroindimicin C / spiroindimicin D / indimicin A / indimicin B / indimicin C / indimicin D / lynamicin A /	Other	6%
Region 11	NRPS,NAPAA	987,179	1,077,272	rimosamide	NRP	100%
Region 12	arylpolyene,terpene	1,163,055	1,217,524	herboxidiene	Polyketide	3%
Region 13	terpene	1,383,115	1,409,061	hopene	Terpene	76%
Region 14	NRPS	1,560,269	1,628,841	isocomplestatin	NRP	93%
Region 15	terpene	1,697,580	1,718,298	kanamycin	Saccharide	5%
Region 16	melanin	1,754,696	1,766,081			
Region 17	other	1,817,455	1,858,505	A-503083 A / A-503083 B / A-503083 E / A-503083 F	NRP	7%
Region 18	oligosaccharide	2,006,992	2,039,271			
Region 19	RiPP-like	2,071,993	2,083,348			
Region 20	butyrolactone	2,158,375	2,168,205			
Region 21	lanthipeptide-class-i	2,182,485	2,206,878			
Region 22	nucleoside	2,209,177	2,229,300	tubercidin	Other	27%
Region 23	NRPS,indole,betalactone,NRPS-like	2,258,779	2,412,083	ulleungmycin	NRP	36%
Region 24	siderophore	2,500,631	2,514,132	ficellomycin	NRP	3%
Region 25	NRPS,PKS-like	3,073,210	3,129,529	tyrobetaine	NRP	100%
Region 26	NRPS	4,128,538	4,209,554	feglymycin	NRP	21%
Region 27	arylpolyene,lanthipeptide-class-iii	4,246,749	4,287,909	fusaricidin B	Polyketide + NRP:Lipopeptide	25%
Region 28	thioamide-NRP,NRPS	4,780,377	4,843,097	ishigamide	NRP + Polyketide	61%
Region 29	butyrolactone	5,457,335	5,468,056	neocarzinostatin	Polyketide:Iterative type I +	8%

					Polyketide:Enediyne type I	
Region 30	lassopeptide	5,827,989	5,850,536	moomysin	RiPP	50%
Region 31	lanthipeptide-class-iii	6,580,629	6,603,265	SAL-2242	RiPP:Lanthipeptide	77%
Region 32	terpene	6,807,557	6,829,254	geosmin	Terpene	100%
Region 33	ectoine	7,240,541	7,250,963	ectoine	Other	100%
Region 34	siderophore	7,328,813	7,336,394	desferrioxamine E	Other	100%
Region 35	siderophore	7,432,072	7,445,806			
Region 36	terpene	8,045,842	8,065,587			
Region 37	T1PKS,NRPS	8,340,903	8,393,865	marinacarboline A / marinacarboline B	Alkaloid	23%
Region 38	NAPAA	8,406,074	8,440,303			
Region 39	NRPS-like,NRPS,phosphonate	8,488,029	8,553,189	deimino-antipain	NRP	66%
Region 40	NRPS,T1PKS,terpene	8,605,582	8,732,273	tetronasin	Polyketide	9%
Region 41	other,NRPS-like	8,808,985	8,880,673	A83543A	Polyketide	8%
Region 42	butyrolactone	8,886,613	8,897,599	cyphomycin	Polyketide	11%
Region 43	T1PKS,NRPS,other,nucleoside, NRPS-like	8,969,427	9,148,676	pseudouridimycin	Other:Nucleoside	68%
Region 44	NRPS	9,237,979	9,285,256			

Supplementary Figure 3.1 – GNPS molecular network showing produced metabolites linked to BGCs. a) GNPS molecular network showing 3622 parent ion nodes produced by 15 Actinomycetota strains. Thirteen metabolites linked by NP Linker are colour coded as per key, metabolites with no link are coloured blue.
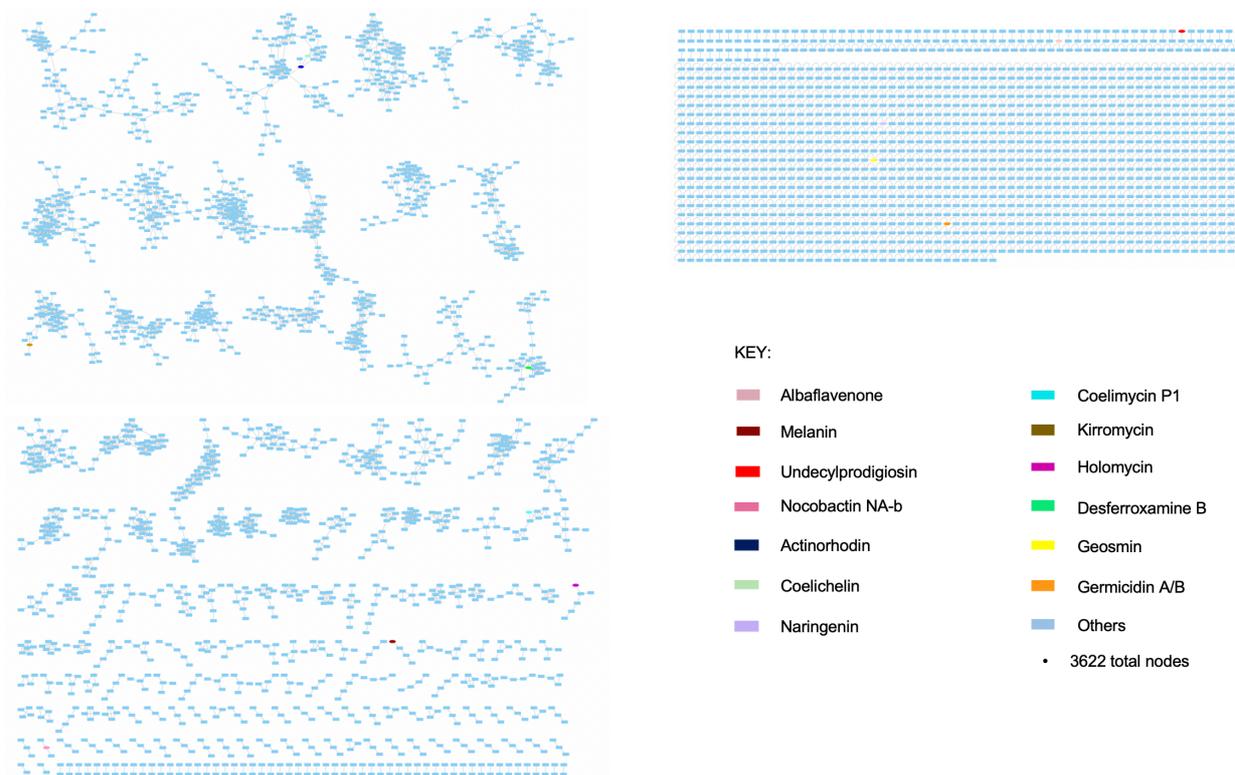
Supplementary Table 3.4 – Comparing extract dry mass from diverse culture conditions. Strains cultured in triplicate and the average weight is shown. The difference between the pre- and post- drying weight is shown in the far-right column, indicating the dry weight of metabolites within the vial.

Solid ISP2	
24-well plates	**Extract weight (mg)**
Media blank	0.4
S. spectabilis	1.4
S. fradiae	1.6

Liquid ISP2	
24-well plates	**Extract weight (mg)**
Media blank	0.3
S. spectabilis	1.7
S. fradiae	1.5

Supplementary Table 3.5 – Comparing extract dry mass between nutrient rich and nutrient poor media with scaled carbon source. The results of metabolite extractions from *S. noursei* using liquid GYM and NMMP media and a liquid/liquid solvent extraction method. Strains cultured in triplicate and the average weight is shown.

GlcNAc conc. (mM)	GYM	NMMP
0	10.45	8.8
10	4	6.53
50	6.97	9.3
Blank	1.2	1.2

		E. coli S.s/S.n	E. coli S.n/S.v	E. coli S.s/S.nod	E. coli S.s/S.v	E. coli S.n/S.nod
Soil Blank	MeOH/H2O	0	0	0	0	0
Soil Blank	MeOH	0	0	0	0	0
Soil Blank	ACN	0	0	0	0	0
Soil Blank	EtOAc	0	0	0	0	0
Prot-Mono	MeOH/H2O	1	1	0	0	1
Prot-Mono	MeOH	0	1	0	0	0
Prot-Mono	ACN	0	0	0	0	0
Prot-Mono	EtOAc	0	0	0	0	1
Ant-Mono	MeOH/H2O	1	1	0	0	1
Ant-Mono	MeOH	0	1	0	0	1
Ant-Mono	ACN	0	1	0	0	0
Ant-Mono	EtOAc	0	1	0	0	1
Co-1	MeOH/H2O	1	0	0	0	1
Co-1	MeOH	1	0	0	0	0
Co-1	ACN	0	0	0	0	0
Co-1	EtOAc	1	1	0	0	0
Co-2	MeOH/H2O	1	0	0	0	1
Co-2	MeOH	0	1	0	0	0
Co-2	ACN	0	1	0	0	0
Co-2	EtOAc	1	0	0	0	0
Co-3	MeOH/H2O	1	0	0	0	1
Co-3	MeOH	1	0	0	0	0
Co-3	ACN	0	1	0	0	0
Co-3	EtOAc	0	1	0	0	0
	Total					28

		S. aureus S.s/S.n	S. aureus S.n/S.v	S. aureus S.s/S.nod	S. aureus S.s/S.v	S. aureus S.n/S.nod
Soil Blank	MeOH/H2O	0	0	0	0	0
Soil Blank	MeOH	0	0	0	0	0
Soil Blank	ACN	0	0	0	0	0
Soil Blank	EtOAc	0	0	0	0	0
Prot-Mono	MeOH/H2O	1	1	1	0	0
Prot-Mono	MeOH	1	1	0	0	0
Prot-Mono	ACN	0	0	0	0	0
Prot-Mono	EtOAc	0	0	0	0	0
Ant-Mono	MeOH/H2O	1	1	1	0	0
Ant-Mono	MeOH	0	0	0	0	0
Ant-Mono	ACN	0	0	0	0	0
Ant-Mono	EtOAc	0	1	0	0	0
Co-1	MeOH/H2O	0	1	1	0	0
Co-1	MeOH	0	0	0	0	0
Co-1	ACN	0	0	0	0	0
Co-1	EtOAc	0	0	0	0	0
Co-2	MeOH/H2O	0	1	0	0	0
Co-2	MeOH	0	0	0	0	0
Co-2	ACN	0	0	0	0	0
Co-2	EtOAc	0	0	0	0	0
Co-3	MeOH/H2O	1	0	1	0	0
Co-3	MeOH	0	0	0	0	0
Co-3	ACN	0	0	0	0	0
Co-3	EtOAc	0	0	0	0	0
	Total					14

Supplementary Figure 4.1 – Breakdown of observed bioactivity within solvent fractions from *Streptomyces* sp. Heat map of observed bioactivity within fractions generated through SPE of solvent extracts of *Streptomyces* sp. 'Prot-' refers to monocultures of the protagonist strain, 'Ant-' refers to monocultures of the antagonist strain. Monocultures were inoculated in triplicate but have been combined into one row above.

Supplementary Figure 4.2 – NP Linker output depicting link between *S. noursei* BGC and spectrum ID 13883.

Display from NP Linker webapp output showing an inferred link between region 10 of *S. noursei* predicted BGC suite (NCBI accession no. CP070326.1) and Spectrum ID 13883 (*m/z* 1411.438 Da, MolFam 50). *S. noursei* region 10 was predicted to be zorbamycin by antiSMASH v6.0.1, which has an accurate mass of 1412.5.
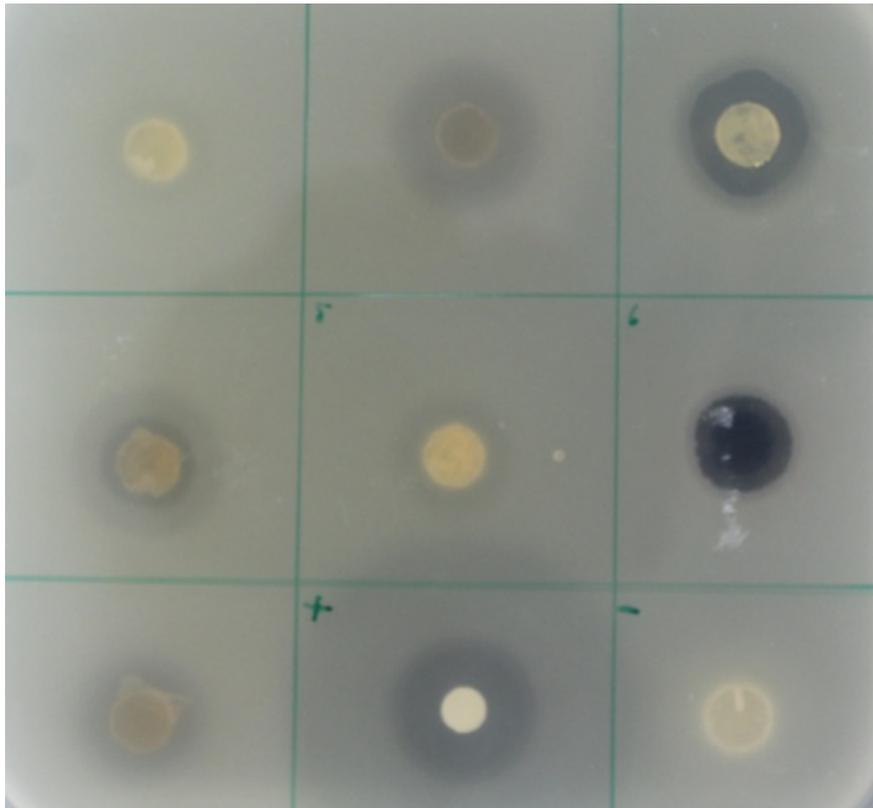
Supplementary Figure 5.1 – Bioactivity screen of seven potential *Streptomyces* sp. isolates against *B. subtilis*. Positive control is a paper disc inoculated with 1 mg/mL spectinomycin. All isolates show some level of activity against the pathogen.

Supplementary Table 5.1 - Showing the nucleic acid concentration (ng/μL) and 260/280 ratio for each soil isolate in preparation for 16S PCR analysis.

Isolate	Nucleic Acid Conc. (ng/μL)	260/280
1	220.4	1.43
2	464.8	1.63
4	625.2	1.34
5	836.5	1.47
6	293.4	1.57
7	500.6	1.60

Supplementary Figure 5.2 – Agarose gel visualised under UV light, showing bands at 1500 compared to the DNA ladder indicating the successful extraction of 16S DNA from each isolate.
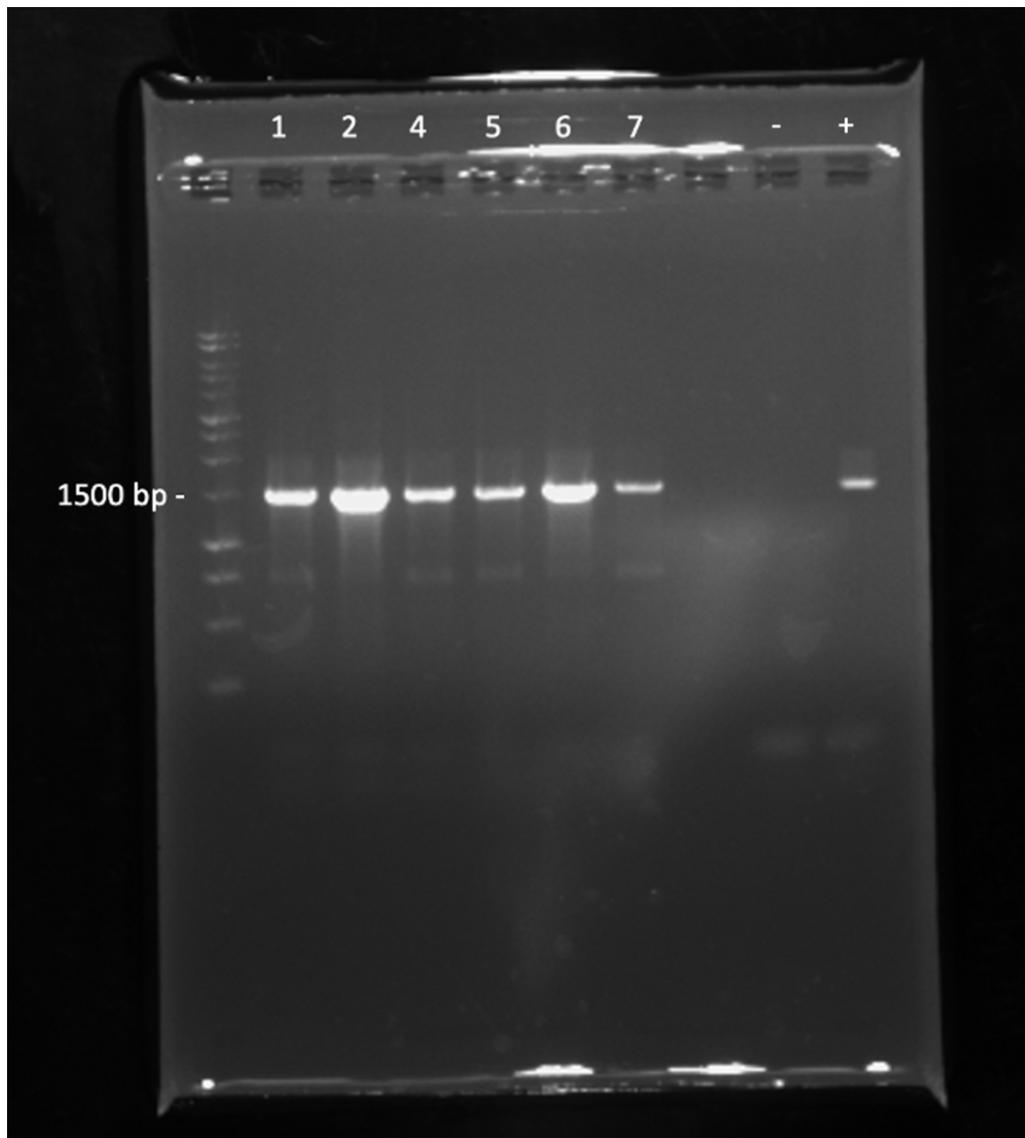
Supplementary Table 5.2 – Showing the nucleic acid concentration (ng/μL) and 260/280 ratio for each soil isolate post-PCR clean-up as preparation for sequencing.

Isolate	Nucleic Acid Conc. (ng/μL)	260/280
1	2.9	2.01
2	3.1	2.04
4	2.4	1.50
5	8.3	1.85
6	14.5	1.89
7	10.8	1.89
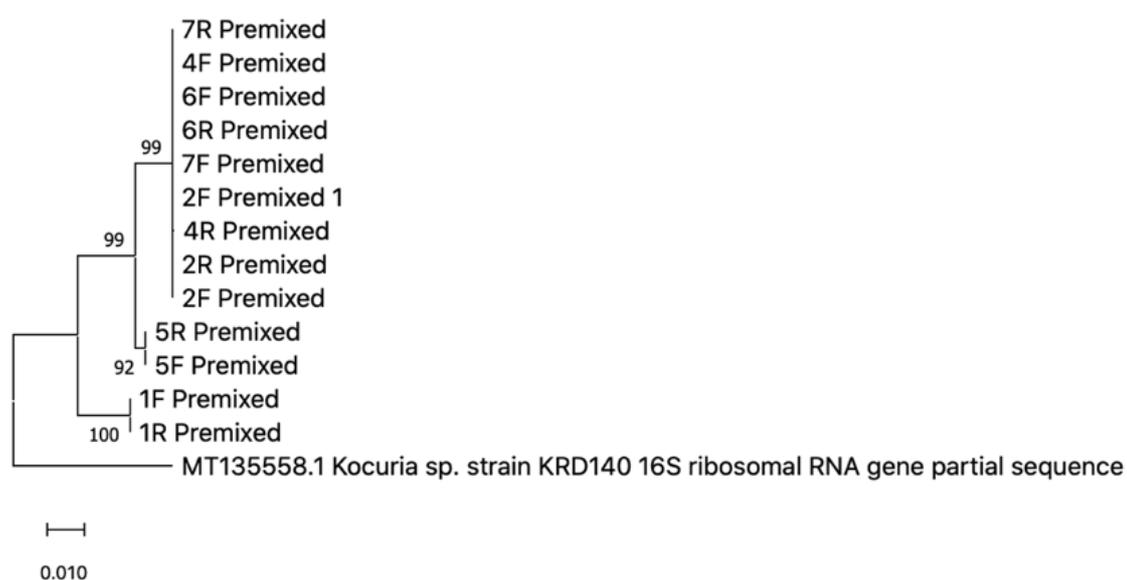
Supplementary Figure 5.3 – Phylogenetic tree showing the evolutionary relationship between 16S rRNA genes of seven soil isolates. Phylogenetic tree constructed using the Temura-Nei method with bootstrap resampling (1000 Bootstrap value) via the MEGAX package. 'F' signifies that the sample was prepared using the forward primer, 'R' signifies that the reverse primer was used. 16S sequence from Kocuria sp. MT135558.1 was included as an outgroup.
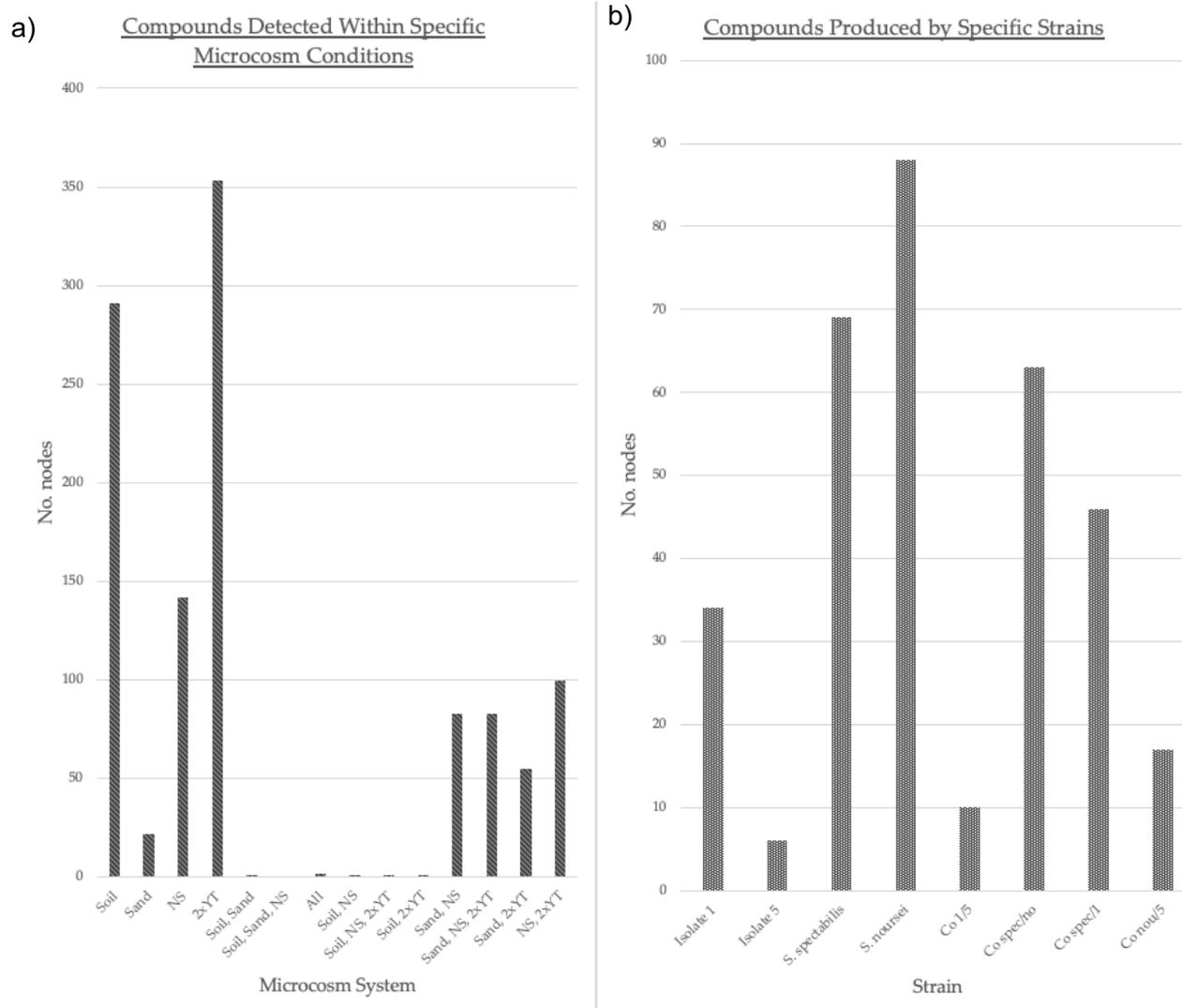
Supplementary Figure 5.4 – Bar charts showing the breakdown of individual spectra occurring a) per microcosm system condition and b) per strain/coculture combination. a) shows the highest number of compounds were detected as being produced within the 2xYT media microcosm system, followed by natural soil. The artificial soil construct which utilised an Artificial soil (NS) based on the composition of natural soil yielded the third highest number of compounds detected, with the sand microcosm system fewest. b) shows that the highest number of individual compounds produced by a strain was by *S. spectabilis*, followed by *S. noursei,* closely then followed by coculture of the two strains.