Strathclyde Institute of Pharmacy and Biomedical Sciences

Comparative Genomics to Understand Specialised Metabolism in *Micromonospora*

David R. Mark

A thesis submitted in fulfilment of the requirement for the degree of Doctor of Philosophy

December 2022

Contents

I. Abstract

Hyphal, sporulating members of the *Actinomycetota* (actinomycetes) – genera such as *Micromonospora* and *Streptomyces* – are notable for production of multiple useful antibiotic natural products. The aim of this project was to screen actinomycete isolates from the Atacama Desert and Scottish soil for production of natural product antibiotics which inhibit the growth of *Pseudomonas aeruginosa,* a frequently multidrug resistant pathogen, and to sequence the genomes of producing organisms to identify biosynthetic gene clusters (BGCs) which drove that activity.

Here we report the sequence of four actinomycetes – one *Streptomyces* isolate from Glasgow Botanic Gardens (*S.* sp. GBG 9AT) and three *Micromonospora* (*Micromonospora* sp. O3, *Micromonospora* sp. O5, and *Micromonospora* sp. PH63) isolated from Atacama Desert Soil. Multiple Sequence Alignment based phylogeny and average nucleotide identity (ANI) calculations identified *S.* sp. GBG 9AT as *Streptomyces xanthophaeus*, a well characterized streptomycete. The same methodology identified the three *Micromonospora* isolates as three novel genospecies. All organisms were BGC rich.

We also sought out to examine possible regulation of these BGCs. Of well-known developmental regulators conserved within *Streptomyces*, we identify the master regulator BldD as the only universally conserved regulator within the three *Micromonospora.* In addition, the three strains are predicted to lack functional BldA, a developmentally regulated tRNA which controls morphological differentiation and antibiotic production. Finally, we analysed the loci of BGCs in single contig assemblies of *Micromonospora* and identify two distinct populations of BGCs occupying the *oriC*-proximal and *oriC*-distal regions of the chromosome, a distribution distinct from that of *Streptomyces*.

This work shows a genomic landscape in *Micromonospora* which is distinct from that of previously studied actinomycetes and lays the groundwork for future studies of the evolution of antibiotic production in these organisms.

II. Declaration

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.

Signed: David Mark

Date: 28/12/2022

III. List of Figures

IV. List of Tables

V. Previously Published Work

Zin, N.M., et al., *Adaptation to Endophytic Lifestyle Through Genome Reduction by Kitasatospora* sp. *SUK 42.* Front Bioeng Biotechnol, 2021. **9**: p. 740722.

Contributed to mining of the SUK 42 genome for Biosynthetic Gene Clusters

Algora-Gallardo, L., et al., *Bilateral symmetry of linear streptomycete chromosomes.* Microb Genom, 2021. **7**(11).

Calculated and plotted heatmap of ANI values for the linear chromosomes analysed

VI. Acknowledgements

This thesis exists, in no small part, due to the kindness, patience, and intellectual support of a huge number of people. To them I owe a massive debt of gratitude.

First, a massive thank you to the past and present members of the University of Strathclyde Microbiology Group. When I started this project, I was a wee idiot who could barely streak a plate. From the effort of every scientist on Floor six who put up with daft questions and mistakes, I now leave as a wee idiot who can mostly streak plates. You've all made it a fantastic and fun few years and if I could go back I would not choose anywhere else to do my PhD. Thanks for the good chats, for conference nights out and subsequent nine am plenary talks, and for repeatedly astounding scientific insight. If I listed everyone on Floor six(and five, Liam) who has been critical to the completion of this work, I would run out of both word count and reader patience.

To the members of the Herron group – Jan, Jana, Gaetan, Gillian, Lis, Tiago, and Elmira, cheers for being phenomenal bay buddies and putting up with the stacks of crispy agar plates. Thanks also to Ally, Emily, and Gordon for listening – time and again – about how a failed experiment was definitely the last straw and how I'm dropping out for real this time. What's a breakdown between friends?

Thanks to Dr Paul Herron, for taking a chance by offering this project to me. It's been rocky at times. Thanks for both pulling me back from the edge, kicking me up the arse when I needed it, and guiding me towards the really interesting questions. We made it, in the end. Thanks as well to Prof. Iain Hunter, Prof. Paul Hoskisson, Dr Morgan Feeney, Dr Kate Duncan, and Dr Arnaud Javelle for always having an interesting perspective on this work whenever they had a chance to see it.

Outwith bench experimentation, this work relied heavily on publicly available sequence data. Thanks to the researchers who no doubt spent a great many hours preparing that data for the sake of their own research. The accumulation of genomic data no doubt stands to further our understanding of biology and I am thankful to have dipped my toes in it.

To the boys – Adam, Chris, Ewan, Jack, and Jamie – thanks for putting up with unhinged rants about how incredible actinobacterial genomics is. I can't think of a better squad, and I'm not sure one exists.

To my partner, Holly, thank you for putting up with me over the writing up process. I look forward to putting up with you when you're doing the same and seeing what the future holds for us when we're both Doctors. Thanks as well Bobbie, Sarah, and Lewis for enduring me working from yours on the odd occasion. JP, thanks to you too.

Mum and Ali, thank you for supporting me through the past eight years of chasing biology. Dale, welcome to the family. Thea, if you're reading this – dear god, don't become a scientist.

Dad, you're the reason I got into this biology schtick. Thank you for showing me how to be a good man in the worst of circumstances. I hope this work makes you proud.

When you enter the University's James Weir building from Montrose Street, you are greeted by a quote from one of Isaac Newton's letters to Robert Hooke: "If I have seen further, it is by standing on the shoulders of Giants.". I'm proud to have spent the past four years of my life amongst giants.

Chapter 1 Literature Review & Introduction

1.1. Antibiotic Resistance

1.1.1. Antimicrobial Resistance

Antimicrobial resistance (AMR) has arisen due to the development of traits within microorganisms that allow them to resist agents used for their control. This is a major clinical problem, with patients developing bacterial infections that persist in the face of previously effective therapies. It is estimated that AMR bacteria are responsible for 700,000 deaths per year, and this is projected to increase to ~ 10 million per year if resistance develops at its current rate [1]. Although AMR is ancient [2], the explosion in rates of resistance are an anthropogenic effect and a combined result of increased consumption of antibacterial agents in varied settings [3]. The effects of increased consumption are amplified by clinical, industrial, and agricultural mishandling of antibacterial agents such as improper prescribing or environmental dumping [4]. This large-scale exposure of bacteria to antibacterial agents creates a selection pressure through which bacteria are forced to develop or acquire mechanisms of resistance.

1.1.2. Mechanisms of Resistance

Despite the enormous variety of resistance genes, the functions they exert are categorised as the exclusion of a bioactive molecule from the cell or through preventing it from interacting with its target ligand. The latter is achieved by employing enzymes which modify either the compound or its target [5] or through the selection of strains that are naturally resistant through having evolved target ligands with a lower affinity for the antimicrobial molecules (e.g. PBP2a, encoded by *mecA* from methicillin resistant *Staphylococcus aureus* (MRSA) *,* has a low affinity for beta-lactam antibiotics and thus confers resistance to them) [6]. Understanding the underlying biochemistry of these resistance mechanisms is of great importance, as it can guide the search for molecules that counteract their effects and thus restore antibiotic sensitivity – such as clavulanic acid from *Streptomyces clavuligerus,* which acts as a β-lactamase inhibitor. Discussed below are some of the common mechanisms that contribute to AMR.

1.1.3. Impermeability

As many antimicrobial compounds depend on interrupting intracellular processes, a viable resistance strategy is for bacterial cells to exclude them from the cell. This can happen as a result of the cell's physiology, as is the case for the members of the genus *Mycobacterium*, whose characteristic rigid cell walls and membranes, and

sparsely distributed porin channels limit antimicrobial entry to the cell [7]. Another example of permeability mediated resistance can be seen in *Pseudomonas aeruginosa,* whose sensitivity to the carbapenem antibiotic imipenem depends on the outer membrane porin, OprD [8] (Fig. 1). Interestingly, alterations in *P. aeruginosa* sensitivity to imipenem can be measured either through mutation of *oprD* or through negative regulation of its expression [9, 10]. The latter demonstrates that permeability is not necessarily static and predetermined by cell membrane and wall composition, but rather complex and regulated by the conditions of the environment that the bacteria inhabit. Loss of antibiotic susceptibility as a result of mutations in porin genes has been noted in other species of bacteria, such as in the *por* gene of *Neisseria gonorrhoeae,* associated with reduced susceptibility to β-lactam antibiotics [11].

Another mechanism to exclude antibacterial compounds is the formation of biofilms by bacterial communities. Composed of a mixture of extracellular polymers including carbohydrates, nucleic acids, and proteins, biofilms can mediate pathogenicity by providing an immune evasion mechanism alongside providing protection from antibacterial agents [12]. One contribution that the biofilm makes towards inhibition is through direct blocking of compounds. However, this effect varies greatly between species and strains of bacteria and is affected by the physiochemical properties of the biofilm's extracellular components [13]. A recurring observation when investigating the effects of impermeability on antibiotic resistance is that, although slightly protective, they only delay the accumulation of molecules [7, 14]. Therefore, prevention of antibacterial toxicity requires other protective mechanisms that take advantage of the delayed intracellular accumulation.

1.1.4. Inactivation Mechanisms

Chemical inactivation of active antibacterial molecules is a central mechanism to preventing their activity, by inhibiting their ability to bind to their ligand and exert their toxic effects. This is commonly achieved by the recruitment of enzymes which either hydrolyse the molecules, such as the β-lactamases; or through a wide range of modifications including phosphorylation, acetylation, and nucleotidylation [5] Alternatively, the target of the antibiotic can be modified, such as with ribosomal RNA methyltransferases which confers resistance to translation inhibiting genes, such as the aminoglycosides [15].

The β-lactam family of antibiotics include the penicillin, cephalosporin, and carbapenem antibiotics which inhibit peptidoglycan synthesis by mimicking the (D)-

Ala-(D)-Ala of nascent peptidoglycan and binding to critical transpeptidases, or penicillin binding proteins [16]. β -lactamases, in turn, hydrolyse β-lactams to prevent this. The Bush and Jacoby classification system introduced in 1995 [17] is commonly used to describe β-lactamase enzymes according to function and has been updated in the last decade to reflect progress in the understanding of β-lactamase diversity [18].

1.2. *Pseudomonas aeruginosa*

1.2.1. *Pseudomonas*

Model organisms for the study of AMR, the genus *Pseudomonas* is comprised of Gram-negative, rod shaped cells that inhabit a wide range of environments [19] and have established themselves as important opportunistic pathogens of both plants [20] and animals. The most important pathogenic member of this genus is *P. aeruginosa* which colonises compromised epithelial surfaces such as burn wounds and the lungs of cystic fibrosis patients [21]. *P. aeruginosa* possesses a swathe of mechanisms that enable it to be an extremely effective pathogen and resist both immunological and chemical killing whilst establishing persistent infections within the host. Discussed below are some of the better understood mechanisms by which *P. aeruginosa* establishes a niche within the host – including the production of toxic virulence factors, biofilm construction, and the regulation of these factors through quorum sensing.

1.2.2. Virulence factors of *P. aeruginosa*

In order to establish and maintain an infection in a host, *P. aeruginosa* must liberate and sequester nutrients from host tissues. This is made possible by the secretion of a variety of factors into the surrounding environment which kill surrounding host cells, protect the bacteria from the hostile conditions present in the host, and aid in the uptake of nutrients. Factors that can lead to the induction of virulence include quorum sensing [22], mechanical attachment to host tissue, [23] and pressure from host defences [24].

Once established in the host, *P. aeruginosa* cells secrete toxic effectors to secure their niche. One method of secretion is the type III secretion system, a needle like multi-protein complex which pierces host cell membranes to inject toxic molecules [25]. These include the production of hydrogen cyanide at toxic levels [26, 27], and elastases that digest the host extracellular matrix and play a role in alginate secretion [28]. In order to sequester scarce iron from the environment, *P. aeruginosa* secretes

siderophores such as pyoverdine and pyochelin which enable it to scavenge both Fe^{3+} and Fe^{2+} ions [29].

1.2.3. Biofilm production by *P. aeruginosa*

Biofilm formation is a common phenomenon whereby bacteria form adherent colonies to a solid substrate, secreting and becoming encased in extracellular polymeric substance (EPS) to protect the colonies from environmental hazards [12, 13]. *P. aeruginosa* biofilms are commonly associated with chronic infections in cystic fibrosis patients, resulting from mutations to the *muc* gene family which encodes negative regulators of alginate production [30].

The *P. aeruginosa* biofilm incorporates different molecules into its structure, including rhamnolipids, extracellular DNA, and other polysaccharides in addition to alginate [31], whilst the cells of the community undergo a number of physiological and metabolic changes as they attach to the substrate and become sessile. As the biofilm matures, planktonic cells are eventually released into the surrounding environment [31].

1.2.4. Quorum sensing in *P. aeruginosa*

Quorum sensing is the process by which bacteria sense population density and then elicit physiological changes in response to this density. There are two key factors responsible for this: the autoinducing molecule (AI) and response regulator (RR), the latter detecting the concentration of the former and then inducing a transcriptional response when AI concentration – as a proxy of population density – is at a high enough level [32]. In *P. aeruginosa*, quorum sensing regulates a variety of responses – biofilm production, host virulence, and antibiotic resistance – through the Las, Rhl, and PQS systems [22].

Quorum sensing can have a key role in antibiotic resistance and tolerance. The PQS system is implicated in the expression of efflux pumps and resistance to both fluoroquinolone and aminoglycoside antibiotics [24]. Alongside this, quorum sensing is also implicated in biofilm development and maturation, as well as controlling starvation responses [33], all of which can play a role in infection. Interestingly, quorum sensing, despite its complexity, does not always confer an advantage to a population, and *P. aeruginosa* deficient in the *lasR* regulator of the Las system are commonly isolated from patients with cystic fibrosis [34, 35].

1.2.5. Clinical *P. aeruginosa*

An important factor to consider when evaluating new antimicrobial agents is that the genomes of laboratory strains can differ significantly from clinically isolated strains, and alongside this there can be significant variation between different clinical isolates. This results in a large pangenome and, importantly, varying resistance profiles between isolates [36]. In parallel, evolution and adaptation to infection sites is observed over the course of a chronic infection, such as those often established in cystic fibrosis patients [37]. These adaptations include resisting the host immune system [38], and evolving a distinct, less acutely virulent phenotype over the course of a chronic infection [39]. Strains of *P. aeruginosa* that colonise a cystic fibrosis airway frequently adopt a mucoid phenotype on agar plates, associated with mutations of the *mucA* gene and its regulation of alginate production [40].

1.2.6. Antibiotic Resistance in *P. aeruginosa*

The three classes of antibiotic most effective at treating *P. aeruginosa* infections are the β-lactams, aminoglycosides, and fluoroquinolones [41]. *P. aeruginosa* has a variety of mechanisms that allow it to tolerate and resist the toxic effects of antibiotic treatment. These include the employment of drug degrading and modifying enzymes, multi-drug efflux pumps which maintain a low intracellular drug concentration, and mechanisms which prevent the entry of the drug into *P. aeruginosa*.

The β-lactam antibiotics exert bactericidal effects by interfering with the synthesis of cell wall peptidoglycan, resulting in the lysis of actively growing bacteria [16]. The current effective β-lactam treatments against *P. aeruginosa* include members of the penicillins, cephalosporins, and carbapenems [42]. *P. aeruginosa* possess a chromosomally encoded cephalosporinase, AmpC, which inactivates most β-lactam antibiotics with the exception of carbapenems [41]. AmpC is generally expressed at low levels and works in conjunction with efflux pumps to help protect *P. aeruginosa* from β-lactam mediated toxicity. Finally, *P. aeruginosa* has a remarkable capacity for excluding small molecules from their cytosol through tripartite efflux pumps such as the MexAB OprM system [43] (Fig. 1). These efflux pumps extend beyond just antibiotic efflux, exporting molecules such as organic solvents [44].

Antimicrobial resistance in *P. aeruginosa* is acknowledged as being a serious threat, with the World Health Organisation placing carbapenem resistant *P. aeruginosa* on its list of organisms against whom new treatments are a critical priority [45].

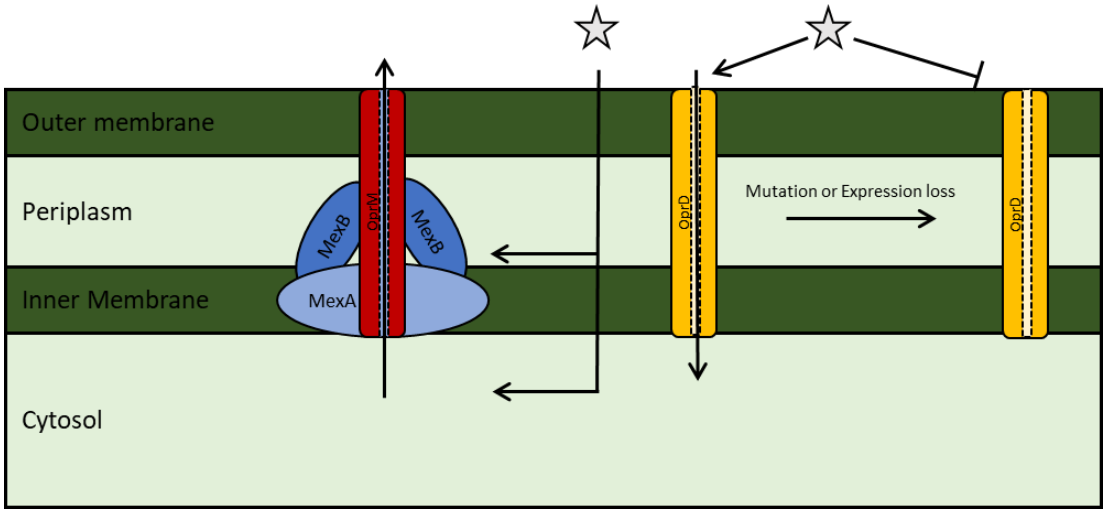


Figure 1-1 Maintenance of low intracellular antibiotic concentration in *Pseudomonas*.

Efflux pumps such as the RND pumps (e.g. MexAB OprM) can maintain a low intracellular concentration of antibiotic molecules (grey stars) by expelling them from both the periplasm and the cytosol. The loss of nutrient uptake systems has also been seen when the uptake system is associated with the import of antibiotic molecules (e.g. OprD).

1.3. Actinobacteria: Nature's Smallest Pharmacists

To combat antimicrobial resistance, new drugs must be discovered. Of major importance in the search for new antibiotics are the actinomycetes (used here in reference to members of the Actinomycetota that form mycelium and grow through hyphal extension). Surpassing fungi and other bacteria in their ability to biosynthesise natural products [46], the actinomycetes account for the production of 60-80% of clinically used antibiotics and anticancer drugs, and contributed hugely to the golden age of antibiotic discovery [47]. Owing to a combined ease of isolation, prodigious capacity for the synthesis of useful natural products, and an unusual physiology including a complex multi-stage life cycle and linear chromosomes, the actinobacterial genus *Streptomyces* gained traction as a model organism for research into actinomycete biology – resulting in the development of an abundance of resources for the research of these organisms [48]. One consequence of this, however, has been the neglect of rarer actinomycetes.

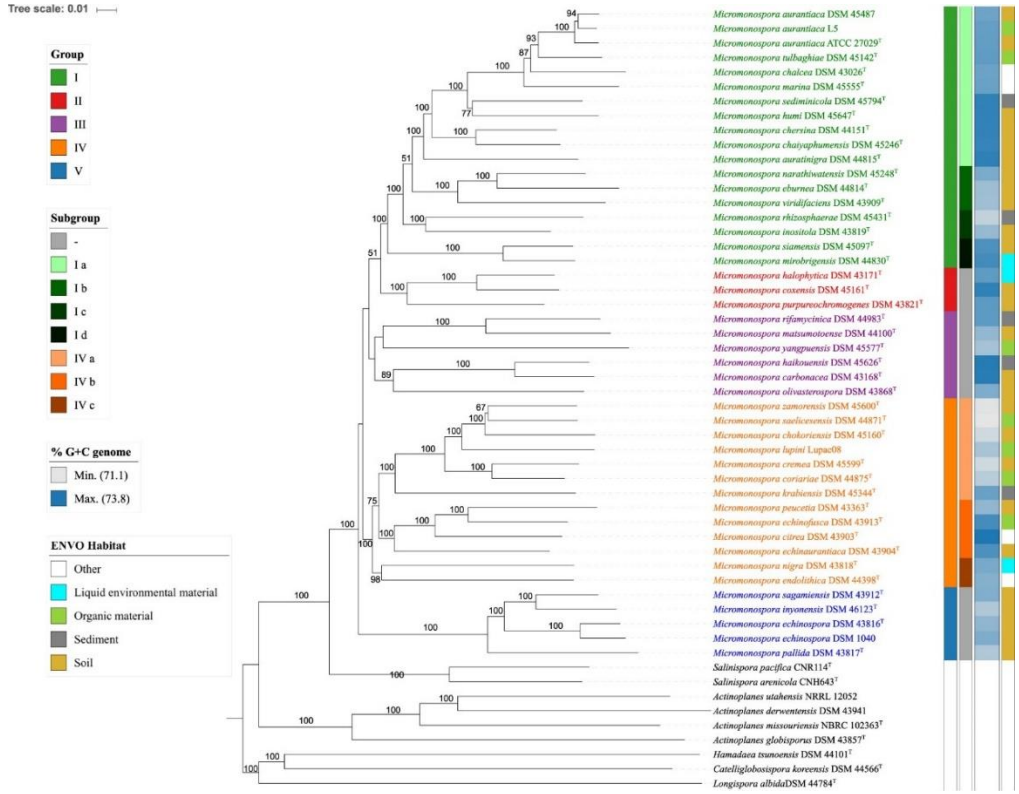
1.3.1. *Micromonospora*

The genus *Micromonospora* occupy an interesting position amongst the rare actinobacteria. Despite being second to *Streptomyces* for the numbers of antibiotics they produce [49] their biology in terms of development and physiology is still poorly understood. They are predicted to have diverged ~250 million years ago from an ancestor common with the genus *Salinispora* [50, 51], and fall within a monophyletic group [52]. The genus was once thought to possess a linear chromosome in line with *Streptomyces* following pulsed-field gel electrophoresis [53], however whole genome sequencing of isolates showed that they possess circular chromosomes instead [54, 55]. In addition to this, both plasmids [54, 56] and phages have been isolated from members of the genus [57].

Having been isolated from samples ranging from the Atacama desert soil [58] to Mariana Trench sediment obtained at depths >10 km [59], the discovery of *Micromonospora* can be described as a process of highs and lows. In addition, members of the genus have been discovered in association with marine sponges [60], mangrove soil [61], and from root-nodules [55, 62], a reflection of the ability of these organisms to survive in a wide variety of environments. They are the source of a variety of important molecules including the aminoglycosides gentamicin and G418, [63] and potent anti-cancer drugs such as calicheamicin [64], highlighting their industrial importance. They are frequently isolated from the rhizosphere environment

and possess a number of enzymes associated with both hydrolysis and nitrogen fixation [65] and have also been identified as potential bioremediation agents [66]. For example, *M. saelicensis* GAR05 colonise *Arabidopsis thaliana* roots and protect against metal toxicity [67]. Despite this, research into members of the genus has been neglected – a PubMed search for *Micromonospora* returns nearly 30x fewer results than one for *Streptomyces* (Fig. 1.3.).

A



B

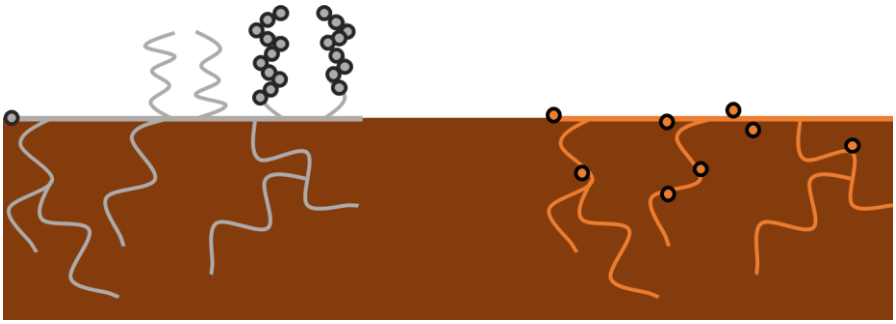


Figure 1-2 The complex life cycle of *Micromonospora*.

(A) Genome-based taxonomy of the genus *Micromonospora* - *Micromonospora* form a well defined group within the *Micromonosporaceae*. Reproduced from [52]. (B) Life cycle of *Micromonospora* - spores germinate, forming a dense multicellular hyphal mycelium in order to scavenge nutrients from the surrounding environment. Differentiation occurs in response to environmental and cellular signals. *Streptomyces* species (grey) erect aerial hyphae which eventually septate and develop into pigmented spores. *Micromonospora* species (orange) do not erect aerial hyphae, instead forming pigmented monospores at hyphal branch points.

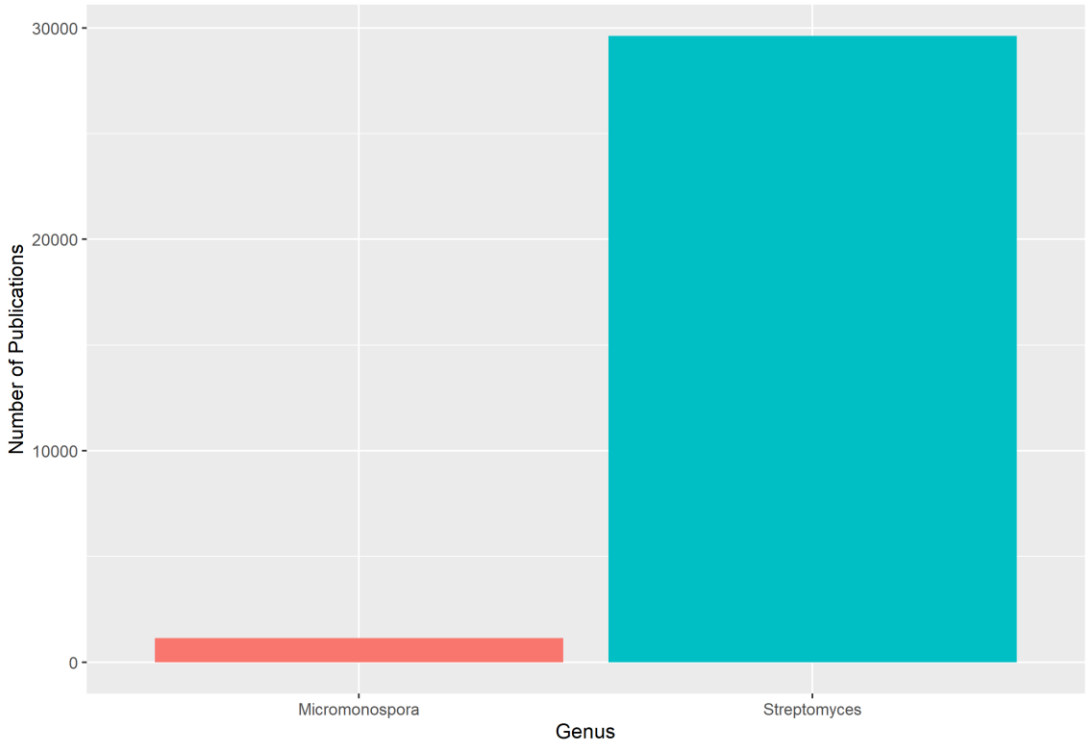


Figure 1-3 *Micromonospora* are a neglected genus compared to *Streptomyces*:

Comparison of number of returned PubMed results when "*Micromonospora*" is used as a search term compared to "*Streptomyces*".

1.3.2. Life Cycle and Development of *Micromonospora*

Despite being clinically and industrially important organisms, the *Micromonospora* life cycle is also poorly understood. *Micromonospora* bear hallmarks of the classic actinomycete life cycle. They begin their life cycle as a spore which germinates to form a vegetative mycelial matt growing into the substrate media (Fig. 1.2), with one of the characteristic features of *Micromonospora* being the formation of individual monospores to enable dispersal [49], in contrast to the spore-chains of *Streptomyces*. Traditionally, *Micromonospora* were considered to not form aerial mycelium, however the formation of aerial structures by colonies has been observed [68, 69]. Although, in the absence of simultaneous chromosome segregation and cell division as in *Streptomyces* it could be argued that these elevated structures are not “true” aerial hyphae.

Although development in *Micromonospora* is understudied, we can make inferences as to how it progresses from research on closely related organisms. Development in actinomycetes is primarily studied in the model streptomycetes *S. coelicolor* A3(2) [70] and *S. venezuelae* [71] because of their diverse and easily observable developmental phenotypes: production the of stark blue and red pigments actinorhodin and undecylprodigiosin; “bald” mutants which are unable to erect aerial hyphae; and “white” mutants which erect aerial hyphae that fail to complete septation and sporulation [72].

From *S. coelicolor* we know that polarised hyphal growth in actinomycetes is driven by the polarisome, a collection of proteins nucleated around DivIVA which drives the growth of and supports the extending tip [73] and promotes chromosome segregation through the ParABS system [74]. From *S. venezuelae*, another model *Streptomyces* which is distinguished from *S. coelicolor* by its ability to complete its life cycle in liquid culture we are beginning to understand the processes involved in cell division in spore-bearing and vegetative mycelia [75, 76].

As well as an understanding of the morphology and growth in *Micromonospora*, we can also use research into *Streptomyces* to infer how development is regulated. The master regulator of development in *S. coelicolor* is BldD, which when deleted results in a bald (hence Bld) colony phenotype [72]. Dimerised by the binding of a cyclic-di-GMP tetramer [77], BldD represses a number of genes essential for development, including itself [72, 78]. As the proteins studied while investigating *Streptomyces* development are often distributed across actinomycetes as a whole [75-77] we can

predict that they also have a conserved function, although these hypotheses are yet to be tested experimentally.

Looking beyond *Streptomyces* to more closely related actinomycetes supports the hypothesis of conserved function in conserved proteins. The genus *Actinoplanes* is closely related to *Micromonospora*, with both being members of the family *Micromonosporaceae* [49]. Clinically famous for the glycopeptide antibiotic teicoplanin produced by *A. teichomyceticus* [79], *Actinoplanes* have garnered interest from developmental biologists for their remarkable sporulation strategy. Instead of forming spore chains like *Streptomyces* or monospores like *Micromonospora*, *Actinoplanes* develop sporangia at hyphal branches. Enclosed within these sporangia are flagellate, motile zoospores which are released in the presence of water [80]. In *A. missouriensis,* a BldD orthologue has been shown to also act as a master repressor of sporangium development in a role analogous to spore chain development in *Streptomyces*, however outwith the control of regulatory genes, the *A. missouriensis* and *S. coelicolor* BldD proteins have markedly different regulons [81]. *Actinoplanes* also have unique methods of controlling their development: *A. missouriensis* for example possesses a two-component system consisting of the HhkA sensor kinase and TcrA response regulator, which are required for proper sporangia formation and dormancy. HhkA homologues are also found in other sporangia forming *Micromonosporaceae* [82].

Thus, although work done in other actinomycetes can shed some light on how *Micromonospora* grow and develop, it is unlikely to provide the entire picture. Just as how *Actinoplanes* and other sporangia forming *Micromonosporaceae* possess their own tools for regulating their unique lifestyle, it is likely *Micromonospora* possess unique factors that are as-yet-unknown.

1.3.3. *Micromonospora* Genomics

Although there is a lack of interest in the physiology of *Micromonospora* when it comes to growth and development, their proclivity for natural product production means that there is a wealth of genome sequence available for analysis [52]. As of 2022, there are 225 *Micromonospora* RefSeq assemblies available from NCBI [83]. This enables easy placement of novel strains in their taxonomic context and for comparisons to be made between strains' abilities to produce natural products.

Work has been done to understand the composition of the *Micromonospora* core and pangenome – the genes shared between members of a set of *Micromonospora* and unique to members of that set, respectively [84]. 81.3% of genes in *M. saelicesensis* and 88.05% of genes in *M. noduli* belonged to their core genomes, dropping to 74.72% of genes shared between the two species [85]. However, we should keep in mind that this study compared only two species of *Micromonospora*, both of which were endophytic – this may have introduced bias to the core genome assessment as endophytic niche specialisation is associated with genome shrinkage [86]. At the genus level, the *Micromonospora* core genome is estimated to be 40-50% of the pangenome [52] As secondary metabolism – responsible for the biosynthesis of natural products – varies even between closely related strains [87] it is important that we gain an understanding of how the *Micromonospora* genomic landscape is laid out and in particular how their specialised metabolism has evolved.

1.3.4. Organisation of the *Micromonospora* chromosome

The industrial utility and clinical significance of actinomycetes has also driven interest in how they organise the biosynthetic gene clusters (BGCs) involved in production of their natural products. Despite how well studied they are, the linear chromosomes of *Streptomyces* limit what we can apply from them to *Micromonospora*. What we do know about *Streptomyces* is that their linear chromosomes are marked by a central core containing the origin of replication, flanked by two arms where BGCs tend to accumulate [88, 89]. The core of *Streptomyces* chromosomes serves as an axis around which recombination occurs [90]. This pattern has been mirrored in *Amycolatopsis* [91] and *Salinispora* [92] chromosomes, although these observations were made from short-read sequenced genomes which may have introduced bias in BGC prediction and analysis [93]. Work in better studied organisms has identified molecular drivers of chromosome architecture, such as the FtsK orienting polar sequences involved in replication termination [94] and architecture imparting sequences which restrict horizontal gene transfer [95].

At the time of writing, there are 31 assemblies of *Micromonospora* available on NCBI which are either complete or chromosome level [83]. These assemblies present a dataset that can help us understand the organisation and evolution of these organisms.

1.3.5. Hyper-arid *Micromonospora*

One observation of *Micromonospora* is their ability to survive conditions considered hostile to life. Strains have been isolated from hyper-saline and metal rich salars [96], and organic matter-poor desert soil [97, 98]. Despite being isolated from hyper-arid environments, these strains are consistently closely related to other *Micromonospora* that have been isolated from a variety of different environments [96-98] and there appears to be little relationship between phylogeny and environmental conditions from which *Micromonospora* are isolated [52].

This work will focus on three strains of *Micromonospora* isolated from soil obtained from the Atacama Desert, specifically, the Altiplano region. This region is characterised by extreme hostility to life owing to hyper-aridity, high levels of radiation, and sparse levels of organic carbon [99]. It can be hypothesised that this extreme environment generates evolutionary pressure for the evolution of novel chemistry, and thus the basis of novel antimicrobial activity. Previous work has identified *Micromonospora* from this region that possess BGCs and thus have rich biosynthetic potential [58].

1.3.6. Natural Products of *Micromonospora*

As is often the case with actinomycetes, *Micromonospora* are recognised as being biosynthetically gifted organisms [100]. Their most important natural product is gentamicin, an aminoglycoside with potent broad-spectrum antibacterial effects [101]. As a clinically and industrially important molecule, research into gentamicin production and the production of related aminoglycosides has garnered much interest [102, 103].

Another natural product class of interest are the enediyne antibiotics. These molecules, can be subdivided into the calicheamicins, dynemicins, and neocarzistatins [104] and possess powerful anti-tumour activity which is hamstrung by extraordinary toxicity [64, 105] owed to inducing double strand breaks in DNA at 5`-TCCT-3` and 5`-TTTT-3` sites. The ability for the molecules to cleave DNA is conferred by nine or 10 membered enediyne rings present in the molecule. How these molecules are biosynthesised is a topic of interest, such as with dynemicin A, which consists of an enediyne ring coupled to an anthraquinone ring. These rings, despite

stark differences in structure, share a polyketide precursor synthesised by DynE8 [104].

Research into *Micromonospora* natural products does, however, extend beyond finding molecules with clinical utility. As previously discussed, *Micromonospora* have been shown to protect plants from the effects of heavy metals, with a potential contribution from siderophores [67]. Soluble products from an *M. lupini* isolate also promote the growth of edible fungi [106] – although the nature of the soluble product remains to be elucidated. This highlights that *Micromonospora* and their natural products offer potential benefits to agriculture as well as in the clinic.

1.3.7. Secondary & Specialised Metabolism of other Actinomycetes

The reason for the actinomycetes', including that of the *Micromonospora*, clinical and industrial importance is their versatile and diverse array of natural products. While primary metabolism describes how organisms incorporate molecules from their external environment, providing precursors essential for growth. Secondary and specialised metabolism, on the other hand, are often used interchangeably to describe the biosynthesis of molecules with more niche purposes, such as ecological interactions [49, 107, 108] and tolerance of adverse environmental conditions [109, 110]. The chemical space occupied by secondary metabolism is broad and diverse, encompassing a large variety of potential molecules including polyketides [111], ribosomally synthesised [112] and nonribosomally synthesised [113] peptides, aminoglycosides [102, 103, 114], and modified nucleosides [115, 116]. Control of these metabolically expensive processes is maintained through a suite of regulatory mechanisms, ensuring that these metabolites are only synthesised as they are needed, such as coinciding with developmental or environmental cues [117-119].

1.4. Biosynthetic Gene Clusters

Genes encoding these control mechanisms are often located adjacent to genes encoding biosynthesis of specialised metabolites, as well as genes encoding other functions such as resistance to toxic metabolites and machinery for their transport. These regions are termed "Biosynthetic Gene Clusters", or BGCs (Fig. 1.4) [120]. A large proportion of genomic space is often dedicated to housing BGCs, approaching 10% of total genome content in some organisms, such as the *Salinispora* [121].

Of particular interest to the field of drug discovery are BGCs without links to natural product in laboratory conditions. This genetic dark matter came to the forefront of

actinomycete research when the sequencing of the *S. coelicolor* A3(2) genome revealed 18 previously unaccounted for BGCs [88]. BGCs that aren't expressed under laboratory conditions – interchangeably referred to as "cryptic" or "silent" BGCs – are in fact the rule, rather than the exception, and are often identified in actinobacterial genome sequences [122].

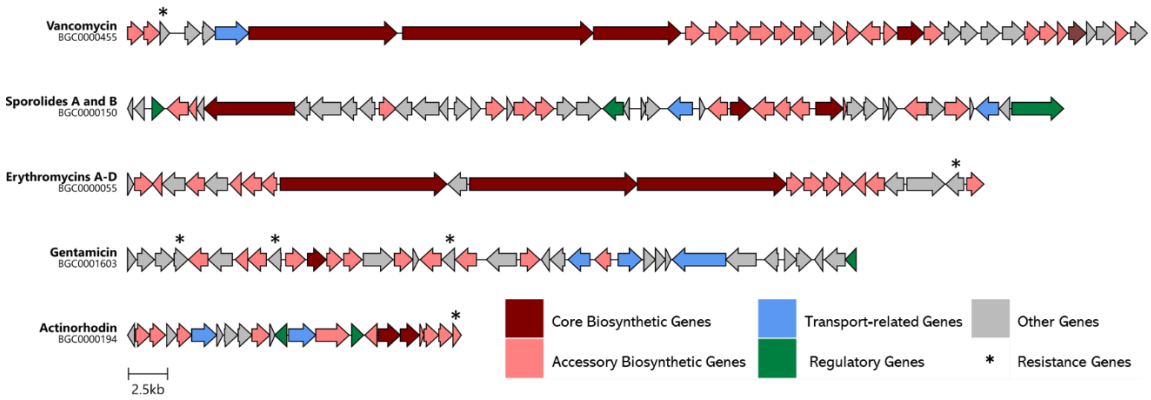


Figure 1-4 Biosynthetic Gene Clusters Contain Conserved Features.

Examples of well-studied biosynthetic gene clusters with genes coloured by their role in the cluster. BGCs were obtained from the MiBiG Database [123] and visualised using clinker [124].

1.4.1. Unlocking the Potential of Cryptic BGCs

The realisation that these cryptic BGCs may in fact hold useful metabolites led to extensive experimentation to try to uncover their hidden natural products through a wide variety of methodologies, termed “elicitation”. One of the easiest of these is to adopt the “one strain, many compounds”, or OSMAC approach. This utilises a variety of culture conditions to induce the production of otherwise hidden metabolites [125]. The simplest way to do this is to vary growth media in order to elicit secondary metabolite production – one of the earliest examples of this was the use of the ISP media to affect soluble pigment production and aid in phenotypic characterisation of actinomycetes [126].

Supporting this hypothesis is the knowledge that carbon-sources can have drastic effects on bacterial phenotype, with the best-understood example of this being diauxic growth in *Escherichia coli* [127] where glucose is preferentially metabolised over other carbon sources through carbon-catabolite repression. This phenomenon also exists in *Streptomyces* under the control of glucose kinase (Glk) and acts to repress secondary metabolism and development [128, 129]. Outwith experimenting with media nutrient composition, other growth conditions which have been shown to alter secondary metabolism include fermentation under hyperbaric oxygen [130] and on inert microparticles [131].

A more finely tuned mechanism for eliciting specialised metabolite production is to exploit known small-molecule regulators of actinomycete development. For example, *N*-acetylglucosamine (GlcNAc), a degradation product of chitin and bacterial cell walls. *Streptomyces* have a genetically encoded sensor for GlcNAc, DasR represses the expression of specialised metabolism in *S. coelicolor* in the absence of GlcNAc, however this repression is alleviated when bound by GlcN-6P. This derepression of specialised metabolism has been shown to occur in other *Streptomyces* species as well [132]. As a product of chitin degradation, GlcNAc may contribute to the induction of specialised metabolism when *Streptomyces* are co-cultured with fungi [133]. Another example of small molecule regulation of specialised metabolism is through 2-isocapryloyl-*3R*-hydroxymethyl-γ-butyrolactone – more simply referred to as “A Factor”. A Factor, as well as other γ-butyrolactones, are themselves specialised metabolites biosynthesised in a pathway that begins with the AfsA mediated condensation between dihydroxyacetone phosphate and a β-ketoacid [134]. The γ-butyrolactones serve as quorum sensing molecules, allowing members of a growing

colony of *Streptomyces* to sense the density of the colony. They trigger the synthesis of specialised metabolites by binding to ArpA to trigger derepression of *adpA* transcription. AdpA then activates multiple pathway-specific transcriptional regulators of specialised metabolism [135].

Knowledge of how specialised metabolism is regulated has also enabled molecular interventions to elicit specialised metabolism. For example, complementation of *adpA in trans* triggered the expression of the lucensomycin BGC in *S. cyanogenus* [136], and it has been shown that variation in AdpA binding sites and promotor sequences strongly effects BGC expression [137]. Another discovery which exploited basic knowledge of actinomycete biology was that deletion of *lsr2* in *S. venezuelae*, which encodes a nucleoid occlusion protein, results in large-scale derepression of specialised metabolite biosynthesis [138].

If all else fails in inducing expression of a BGC in its native host, there is also the option of removing it and relocating it to a more amenable host organism. Thus is the process of heterologous expression [139]. As some actinomycetes are notorious for being difficult to work with, the option of excising a BGC and introducing it to a faster growing, better characterised host is a tempting proposition. Considerations for heterologous expression include the size of the BGC to be transferred, with the large size and high gene count present in many BGCs presenting a challenge and selecting an appropriate host to express the BGC. One of the great success stories from this method is the complete biosynthesis of Erythromycin A – normally produced by *Saccharopolyspora erythrea* [140] – in an *E. coli* host, however this required extensive modification of the native BGC. Whilst the rapid growth and genetic tractability of *E. coli* strains offer advantages to experimentation, *S. coelicolor* as a well-established model actinomycete has also been adopted as a heterologous host. More specifically, *S. coelicolor* strains M1152 and M1154 which lack the BGCs encoding actinorhodin, the calcium dependent antibiotic, and undecylprodigiosin, the cryptic polyketide – thus increasing the available pool of biosynthetic material – as well as mutations to *rpoB* and *rpsL* which improve BGC transcription efficiency [141].

1.4.2. The Evolution of Specialised Metabolism

Much remains to be understood about the evolution of specialised metabolism. The fitness benefit of a given BGC may not be immediately obvious – clinically useful natural products and their derivatives are used at concentrations that far exceed whatever would occur in nature [142]. This raises the question of whether observed

antibiotic activity has been selected for or exists coincident to another ecological role. For example, anthracycline molecules – clinically co-opted for their anti-cancer activity – are able to protect against phage infection [107]. It is likely that the ecological role for other molecules remains to be elucidated.

Alongside asking what role microbial natural products play in a microbe's survival is the question of how that natural product's biosynthetic machinery has evolved. Specifically, how did their BGCs originate, and how do they develop and lead to the diverse molecules that we observe? We know that specialised metabolism is extremely diverse, even at genus and species level [91, 92]. We also know that some BGCs show extensive phylogenetic conservation, developing around their core biosynthetic genes [143, 144]. These two observations indicate that both vertical inheritance of conserved BGCs as well as the horizontal transfer of genetic elements are what drives the evolution of novel BGCs.

Horizontal gene transfer and vertical inheritance as ways of obtaining and maintaining BGCs likely offer different fitness trade-offs for both the BGC and their host. BGCs have been found in both conjugative plasmids [145], integrative and conjugative elements [146], and even in phages (although rarely) [147]. This suggests that for some BGCs, the ability to mobilise offers a fitness advantage to the cluster itself. Conversely, the core of the salinisporamide BGC is estimated to have been fixed in the genus for millions of years, with salinisporamide diversity being achieved by encoding tailoring modules adjacent to the BGC [143]. Although plasmids and phage have both been identified in *Micromonospora* [56, 57], the extent to which they drive the evolution of the genus remains to be seen.

1.5. Sequence, Assemble, Delve: Genome Guided Drug Discovery

A major advance in our ability to investigate actinomycete secondary metabolism has come from advances in genome sequencing. In the two and a half decades since the sequencing and assembly of the *Haemophilus influenzae* chromosome [148], technological advances have led to a rapid fall in both the time and financial cost of bacterial whole genome sequencing. This has revolutionised the process of "genome mining", where genetic information is used to predict the secondary-metabolic potential of organisms of interest [149]. As previously mentioned, the importance of sequencing actinobacteria to fully understand this potential was revealed with the completion of the *S. coelicolor* A3(2) genome sequence, and the realisation that it possessed 18 biosynthetic gene clusters of unknown function – aforementioned

“cryptic” biosynthetic gene clusters [88]. This discovery gave birth to a genomic gold rush, as researchers began to investigate what secrets their strains held within their genetic information and has led to the discovery of several interesting molecules [150-152].

There are three steps that underpin genome mining from a novel organism: first, the process of sequencing its genome. Secondly, the assembly of that sequence information into an accurate model of that genome. Lastly, the prediction of biosynthetic gene clusters and their content in that model.

1.5.1. Sequencing Methodologies

In essence, gene sequencing is the process of translating the chemical information of nucleic acid sequence order into information that can be interpreted by researchers and their tools (Fig. 1.5). This translation can be achieved by several means, each of which involves different ways of manipulating DNA to infer its sequence. This thesis presents genome mining data acquired from de-novo assemblies acquired using both Oxford Nanopore Technologies (ONT) and Illumina sequencing. Discussed below is how these technologies differ from each other, as well as their advantages and limitations in the context of bacterial natural product research.

1.5.1.1. Illumina Sequencing

Contemporary Sanger-sequencing uses fluorescently tagged dideoxynucleotides and electrophoresis to infer the base sequence of a sample of DNA [153]. Illumina can be interpreted as a massive parallelisation of this process, eschewing electrophoresis for imaging as DNA molecules are synthesised [154, 155]. During an Illumina run, adapter sequence added to the DNA as part of library preparation binds to complementary sequence on the surface of an Illumina flow cell. The library is then amplified into clonal spots on the flow-cell utilising solid-state bridge PCR. With this library amplified the sequence of its members is then determined by successive rounds of incorporating fluorescent nucleotides complementary to the spots, imaging them, and then cleaving their fluorescent tags [156].

Through the direct imaging of nucleotides as they become incorporated, Illumina sequencers generate reads with >99% base calling accuracy. Its sequencing covers both ends of a DNA fragment, compensating for drops in quality as strands extend and shifting the most difficult-to-predict region from a fragment’s end to its midpoint [157]. The precision offered by Illumina sequencing is useful – allowing for confidence

in the quality of the assemblies they generate. That said, despite the accuracy of their reads, Illumina is limited by the length of the reads it can generate. Reaching a maximum of 300 bp, these short reads struggle to capture the repetitive sequences present in some BGCs [93, 158].

1.5.1.2. Oxford NanoPore sequencing

Where the read lengths generated by Illumina sequencing constrain their ability to resolve repetitive sequence, Oxford Nanopore Technology's nanopore (ONT) sequencing generates reads that span kilobases of DNA. ONT technology predicts bases by measuring current change as a strand of DNA passes through a protein pore, followed by the complementary strand to that strand of DNA [159]. Although initially held back by low basecalling accuracy, advances in sequencing chemistry and computation have pushed this to >75% [160, 161]. Whilst still lower than Illumina, these advances in accuracy have enabled complete *de novo* whole genome assemblies to be obtained using nanopore exclusively [162]. In addition to relatively low base calling accuracy, ONT sequencing struggles to resolve homopolymeric DNA sequence. As a homopolymer runs through the pore it generates an extended uniform signal, which is computationally challenging to convert back to polymer length.

Where ONT sequencing shines is both in its read length (with a capacity to generate reads in excess of two Mb in length [163]) and, in the case of the MinION sequencer, accessibility and ease of use [159]. The length of the reads generated by ONT allows for it to overcome regions of repetitive sequence and capture genome architecture in a way that short-read sequencing technologies cannot.

1.5.2. Genome assembly for Natural Product Discovery

Genome assembly is the process of using reads generated by sequencers to generate a model of the DNA sequence from which the reads were derived. There are two ways of accomplishing this: Reference-guided alignment or *de novo* assemblies. Reference-guided alignment maps reads from a sequencer onto a provided genome sequence [164, 165]. This method is advantageous in its ease of comparing how sequenced DNA differs from a known reference sequence, and ranges in use from identifying single-nucleotide polymorphisms (SNPs) to large scale deletions [166]. However, references are not always available – for example when sequencing the genome of a novel organism for the first time. For this, *de novo* assembly is needed. This is the process of computationally joining reads algorithmically with methods such as de Bruijn graphs [167-170], with different assemblers built to solve different problems such as error-prone or repetition-heavy reads. These assemblies produce strings of contiguous sequence (contigs) which represent the base order of the DNA sequenced.

Genome assembly software is rarely able to recapitulate a single contig for each replicon sequenced, with assemblies of >100 contigs being common. It is because of this that assemblies are classed into different levels of completion (Table 1) to describe how well assembled they are [171]. There are also bioinformatic tools which provide quality information about query assemblies. Whole-genome aligners such as MUMmer [172] and Mauve [173] can be used to identify large scale rearrangements in assemblies which may be artifacts, whereas dedicated quality control tools such as Quast [174] can compute common quality metrics for a submitted assembly.

1.5.3. *De novo* Genome Assembly

As actinomycetes possess highly diverse genomes [51], *de novo* assembly is often necessary to accurately recapitulate BGC structure. This is complicated by the high GC content of actinomycete genomes [49], linearity of chromosomes such as in the *Streptomycetaceae* [90, 175] as well as the conserved and repetitive sequence present in some BGC classes – for example polyketide synthase and non-ribosomal peptide synthase classes [93].

To generate *de novo* assemblies, software tools such as Unicycler [167], Canu [169], and Flye [170] attempt to reconstruct genomes in the absence of a reference to map reads to, ideally resulting in a single contig per replicon. Unicycler can be used to assemble genomes with either short reads alone, or incorporating both short and long sequencing reads. To do this, it first uses the SPAdes assembler [168] to generate multiple assemblies from short reads using different k-mer sizes. It then attempts to solve the problem of repetitive sequence (inferred from multi-copy contigs that share edges with single copy contigs) with bridges, based on the contig path generated by SPAdes. Without long reads, the assembly ends here – however, the aforementioned repetitive sequence problem may not be solved, and pose problems during genome mining owing to PKS and NRPS genes being rich in repetitive sequence [93].

Long reads, if available, allow Unicycler to complement the SPAdes assembly – by aligning these reads to the assembly and using regions that span contigs with multiple edges to resolve which edge is the “true” edge, to improve contiguity of the assembly.

Long read assemblers, on the other hand, struggle to utilise De Bruijn graphs in assemblies, as k-mers throughout a genome are less uniformly distributed in a set of long reads. [170]. As long-read sequencing reads can be theoretically long enough to cover entire BGCs [176], high quality long read assemblies offer great advantages to

genome mining. Recently, the Flye assembler has sought to solve the problem of long read assembly by employing “disjointigs” in its algorithm – which it then concatenate, map reads to, and attempts to solve repeats through. Long read sequencing and assembly also has the advantage of being able to span the entirety of smaller plasmids with their reads, which enables easier identification of these genetic elements.

Table 1.1 NCBI Assembly Quality Identifiers

Assembly Completeness Level	Description
Complete	All chromosomes covered. No gaps in submitted assembly. No runs of more than ten ambiguous bases. No unplaced scaffolds. Plasmids are included in the assembly.
Chromosome	One or more chromosomes covered. Gaps and scaffolds are permitted
Scaffold	Contigs are connected into scaffolds but not further assembled
Contig	The assembly consists only of contigs

1.5.3. Overcoming Challenges in Assembling Genomes for Mining

BGCs are difficult or even impossible to solve using exclusively short reads which don't span the repetitive regions present in the cluster [176]. In addition, assemblies constructed of only short reads carry the risk of fragmenting larger BGCs across multiple contigs [177], leading to both overestimation and inadequate prediction of an organism of interest's secondary metabolome. On the other hand, the high rate of error present in ONT derived reads reduces the quality of assemblies derived exclusively from them and thus might impact on BGC annotation owing to incorrect gene prediction. It can also impact phylogeny predictions, which depend on comparisons between conserved genes and thus are sensitive to incorrect base predictions [178].

The accuracy of Illumina-derived reads and the length of ONT reads complement each other, and so they can both be incorporated into the assembly process to improve the resultant genome assembly. So-called "hybrid" assemblies incorporate information from both short and long reads – the long ONT reads are used to branch long repetitive sequence regions with the accurate short Illumina reads compensating for ONT's error rate. This methodology has been shown to reduce fragmented BGC prediction [177]. The ideal actinomycete genome assembly for genome mining consists of a single contig for each chromosome with all the BGCs represented, annotatable, and able to be located within their chromosome.

1.5.4. Genome Mining and BGC prediction: Here Be Drugs

Armed with high quality genome sequence, it is possible to predict the locus and nature of BGCs contained within that sequence – the process of which is a developing field. Released in 2011, the antibiotics and secondary metabolite shell (antiSMASH) [179] has gone on to become one of the most popular tools in the field for BGC prediction. At inception, antiSMASH functioned by scouring submitted genome data for proteins and protein domains which were characteristic of a given BGC class, based on profile Hidden Markov Models (HMMs) [180], as well as excluding similar genes that played a role in primary metabolism. These models were derived from published information on what genes play a role in the biosynthesis of secondary metabolites. Over a decade later antiSMASH has processed more than a million queries and has been updated to version 6.0. [181]. Part of the appeal of antiSMASH is its ease of use and its incorporation into other tools and databases [123, 182, 183].

Aside from antiSMASH, there are other genome mining tools that vary in what type of specialised metabolite they identify BGCs of and how they identify clusters. The prediction informatics for secondary metabolomes (PRISM) algorithm, for example, also identifies BGCs using HMMs, but its advantage is its ability to effectively predict specialised metabolite structure [184], through features such as predicting tailoring reactions that occur during natural product biosynthesis. There are also tools which specialise in identifying BGCs of a specific class, such as the suite of tools available for interrogating ribosomally synthesised and post-translationally modified peptides (RiPPs) [185].

Many BGCs code for the production of products that do not act as antibiotics, posing a problem for prioritising which BGC should be interrogated – a genetic triage is needed to optimise drug discovery. As many natural products are inherently toxic, their BGCs often encode self-resistance mechanisms such as the *vanHAX* operon in the *Amycolatopsis orientalis vcm* cluster which encodes vancomycin resistance [186]. These resistance mechanisms can be exploited to help prioritise BGCs that may encode for antibiotic production. The Antibiotic Resistant Target Seeker is one such tool that exploits this, by first identifying BGCs in a query genome with antiSMASH and then identifying clusters containing known models of resistance [187].

1.6. Aims and Objectives

The discovery of novel antibiotics to combat AMR requires an intimate understanding of the biology of antibiotic-producing organisms. To that end, the aim of this project is to characterise three *Micromonospora* species isolated from Atacama Desert soil. We aimed to confirm whether or not these *Micromonospora* strains were capable of producing antibiotics able to inhibit *P. aeruginosa*. We also sought to sequence the genomes of these organisms and generate high-quality assemblies to facilitate mining of their genomes, understanding of any adaptations they may have that contribute to their lifestyle in the Atacama, and how they are phylogenetically related to other *Micromonospora* strains. Finally, in order to gain an insight into the evolution of *Micromonospora* secondary metabolism and how BGCs are distributed throughout the genus, we map the BGC content present in high-quality *Micromonospora* genome assemblies. This work has the fundamental objective of further understanding the specialised metabolism of this important organism.

Chapter 2 Materials and Methods

2.1. Media and Reagent Components

To prepare growth media, medium components were dissolved in dH_2O and pH corrected using solutions of 1 M HCl and 1 M NaOH (also dissolved in dH_2O). After pH adjustment, agar powder was added and mixed, and the media was sterilised by autoclaving at 121°C/1.2 bar for 11 minutes. Trace Salt Solution and Trace Element Solution were prepared and passed through a 0.2 µm filter and stored at 4°C to be added after autoclaving.

Table 2.1 Media Components

Media	Recipe
Reduced-Arginine Starch (RAS) agar	Arginine: 0.1 g Soluble Starch: 12.5 g K_2HPO_4: 1 g NaCl: 1g $MgSO_4$ $7H_2O$: 5 g dH_2O: 1 L pH 7.2
Starch-Casein (SC) agar	Soluble starch: 10 g K_2HPO_4: 2 g KNO_3: 2 g Casein: 0.3 g $MgSO_4$. $7H_2O$: 0.01 g Agar: 15 g dH_2O: 1 L pH 7.2
Oligotrophic agar	Yeast extract: 0.05 g Tryptone: 0.5 g Sodium Glycerophosphate: 0.1 g dH_2O: 1 L pH 7.2
Glucose Yeast Malt extract (GYM) agar	Glucose: 4 g Yeast extract: 4 g

	Malt extract: 10 g $CaCO_3$: 2 g Agar: 12 g dH_2O: 1 L pH 7.2
Lysogeny Broth (LB) agar	Peptone: 10 g Yeast Extract: 5 g NaCl: 5 g Agar: 15 g dH_2O: 1 L pH 7.2
Nutrient Broth	Lab Lemco powder: 1 g Yeast extract: 2 g Peptone: 5 g NaCl: 5 g dH_2O: 1 L pH 7.2
Soft Nutrient agar	Lab lemco powder: 1 g Yeast extract: 2 g Peptone: 5 g NaCl: 5 g Agar: 7.5 g dH_2O: 1 L pH 7.2
Nutrient agar	Lab lemco powder: 1 g Yeast extract: 2 g Peptone: 5 g NaCl: 5 g Agar: 15 g dH_2O: 1 L

	pH 7.2
HiMedia International Streptomyces Project media 2	Yeast extract: 4 g Malt extract: 10 g Dextrose: 4 g Agar: 20 g dH_2O: 1 L pH 7.2
HiMedia International Streptomyces Project media 3	Oats: 20 g Agar: 18 g Trace salt solution*: 1 ml dH_2O: 1 L pH 7.2
*Trace salt solution	$FeSO_4 .7H_2O$: 0.1 g $MnCl_2 .4H_2O$: 0.1 g $ZnSO_4$ $7H_2O$: 0.1 g dH_2O: 100 mL
HiMedia International Streptomyces Project media 4	Soluble Starch: 10 g K_2HPO_4: 1 g $MgSO_4 .7H_2O$: 1 g NaCl: 1 g $(NH_4)_2SO_4$: 2 g Trace salt solution: 1 mL Agar: 20 g dH_2O: 1 L pH 7.2
HiMedia International Streptomyces Project media 5	L-asparagine: 1 g Glycerol: 10 g K_2HPO_4 Trace salt solution: 1 ml dH_2O: 1 L

	pH 7.2
HiMedia International Streptomyces Project media 6	Peptone: 15 g Proteose peptone: 5 g Ferric ammonium citrate: 0.5 g K_2HPO_4: 1 g Sodium Thiosulphate: 0.08 g Yeast extract: 1 g Agar: 20 g dH_2O: 1 L pH 7.2
HiMedia International Streptomyces Project media 7	Glycerol: 15 g L-tyrosine: 0.5 g L-asparagine: 1 g K_2HPO_4: 0.5 g NaCl: 0.5 $MgSO_4 .7H_2O$ Trace element solution**: 1 mL Agar: 20 g dH_2O: 1 L pH 7.3
**Trace element solution	$FeSO_4 .7H_2O$: 1.36 mg $CuCl_2 .7H_2O$: 0.027 mg $CoCl_2 .6H_2O$: 0.04 mg $NaMoO_4 .2H_2O$: 0.025 mg $ZnCl_2$: 0.02 mg H_3BO_3: 2.85 mg $MnCl_2 . 4H_2O$: 1.8 mg Sodium tartarate: 1.77 mg dH_2O: 1 L
Soya Flour Mannitol agar	Mannitol: 20 g Soya flour: 20 g $CaCl_2$: 1.2 g

	Agar: 20 g dH_2O: 1 L pH 7.2
Tryptic soy broth	Pancreatic digest of casein: 17 g Enzymatic digest of soya beans 3 g NaCl: 5 g K_2HPO_4: 2.5 g Glucose: 2.5 g dH_2O: 1 L pH 7.2

2.2. Growth, Culture, and Maintenance of Strains

For maintenance, *Micromonospora* strains were grown on 25 mL of GYM agar in 90 mm round petri dishes. To prepare spore stocks, lawns of *Micromonospora* were incubated at 30ºC until they showed black pigment production indicative of sporulation. To harvest spores, 5 mL of sterile dH_2O was laid over the lawns and the spores were dislodged by agitation with a sterile cotton bud and cleared of hyphal fragments by passing through a cotton plugged syringe. The spore suspension was then centrifuged at 4000 RPM for 10 minutes, and the supernatant decanted. Finally, the spore pellet was resuspended in glycerol (20% v/v in dH_2O) and stored at -80ºC. Maintenance of *Streptomyces* strains was carried out using the same procedure, except on SFM agar with sporulation interpreted by a change of colour after the erection of aerial hyphae.

For the maintenance of pathogens used in bioactivity assays, the organisms were colony purified from previously made stocks and grown on 25 mL of nutrient agar overnight at 37ºC. Single colonies were then inoculated in 5 mL of nutrient agar and grown overnight at 37ºC, shaking at 250 RPM. Overnight cultures were then centrifuged at 4000 RPM for 10 minutes, and the supernatant decanted. Finally, the cell pellets were resuspended in glycerol (40% v/v in dH_2O) and stored at -80ºC.

2.3. Data processing and Statistical Analysis

Where appropriate, statistical analysis was carried out in RStudio V. 1.4.1106. (R V. 4.2.0). Data manipulation and visualisation was carried out using the tidyverse package. Scripts used for data analysis are available at https://github.com/DavidRcoMark/Thesis. The European and Australian Galaxy servers [188] were employed for bioinformatic analysis.

2.4. Isolation of Soil Actinobacteria

In addition to testing *Micromonospora* strains from the Atacama Desert, novel isolates of *Actinobacteria* were also generated to test for anti-pseudomonad activity. Soil for these isolations was obtained from a Glasgow Green footpath, potted basil (*Ocimum basilicum*) and coriander (*Coriandrum sativum*) plants, and from the base of a variety of plants from the Glasgow Botanic Gardens. For isolation of novel actinobacteria, 1 g of sampled soil was incubated at 60ºC to kill vegetative organisms. The soil was then transferred to 9 mL of quarter-strength Ringer's solution and homogenised by vortex mixing for 10 minutes. The suspension was diluted to 10^{-1} and 10^{-2}, and 200

µL of the suspensions were spread on plates of 25 mL RAS, SC, and Oligotrophic agar supplemented with cycloheximide (25 µg mL^{-1}), nystatin (25 µg mL^{-1}) and nalidixic acid (10 µg mL^{-1}) and incubated at 30°C. Plates were checked for growth frequently, with colonies of actinobacteria being selected by colony morphology (matte appearance, pigment production, and/or aerial hyphae production). Colonies were also selected if hyphae were visible at the edge of colonies when observed under a microscope. Pure single colonies were spread to grow as lawns and incubated at 30°C until spores were harvestable. The spores of novel isolates were harvested as described in section 2.2.

2.5. Agar-Plug Bioactivity Assay

In order to be screened for antibacterial activity, spores of isolated actinobacteria were streaked using a sterile cotton bud onto International Streptomyces Project (ISP) media 2,3, 4, and 5. 14 days after inoculation, spores were streaked onto ISP media 6 and 7. Plugs were taken from the media after seven days of growth on ISP 6 and 7 and 21 days of growth on ISP 2, 3, 4, and 5. These plugs were placed onto soft Nutrient agar seeded with either *Acinetobacter baumanii* ATCC 19606, *Pseudomonas aeruginosa* ATCC 27853, or *Escherichia coli* ATCC 25922*,* that had been overlaid onto a plate of Nutrient agar (Fig. 4). Plates were incubated overnight at 37°C and then observed to identify zones of inhibition.

To seed agar before overlaying; single colonies of *A. baumanii, P. aeruginosa,* or *E. coli* were grown overnight (37°C, 250 RPM) and diluted to an optical density at 600 nm of 0.01 in 7.5 mL of soft nutrient agar.

2.6. Characterisation of Hyper-arid *Micromonospora*

2.6.1. Microscopy of Isolates

Phase contrast microscopy was used to observe the hyphal and spore morphology of *Micromonospora* sp. O3, O5, and PH63. Flame-sterilisefd square coverslips were inserted into GYM agar, and 10 µL of spores of each strain were inoculated on the underside of the coverslips, in triplicate. For each strain, seven plates with three coverslips per plate were inoculated and incubated at 37°C.

To stain, cultures were fixed in 100% methanol for 30 seconds, submerged in 1% aqueous methylene blue solution foroneminute, and then washed using methanol and blotted dry. Coverslips were mounted in 10 µL of 40% glycerol, sealed with nail

varnish, and stored at -20°C. Phase contrast images were acquired on a Nikon Eclipse TE2000-S microscope, using NIS Elements Imaging Software (v3.22.11 Build 728). Image processing and quantification was performed using Fiji [189]. Spore size, interbranch distance, and width of hyphal bundles in *Micromonospora* sp. O3, were measured.

For *Micromonospora* sp. PH63, scanning electron microscopy (SEM) and epifluorescence microscopy were also employed. SEM of seven day old culture grown on GYM agar was performed by Dr. Jana Schniete on a JEOL InTouchScope, with a 20 kV acceleration voltage. Nascent peptidoglycan/Nucleoid staining of three day old culture grown at 30°C was performed by Jan DeWald, as previously described [190]. Briefly, coverslips were fixed in 500 µL of glutaraldehyde/paraformaldehyde (0.0045%/2.8% in PBS) for 15 minutes, washed twice with PBS, and then dried. The mycelium was then rehydrated in PBS for five minutes, washed once more, and incubated with Bovine Serum Albumin (BSA) (2% in PBS) for five minutes. To stain coverslips were incubated in the dark with fluorescein conjugated wheat-germ agglutinin and propidium iodide solution (2 µg mL^{-1} FITC-WGA, 10 µg mL^{-1} PI, in 2% BSA in PBS) for two hours at room temperature. Following this incubation, slides were washes eight times with PI (10 µg mL^{-1} PI, in 2% BSA in PBS). The stained coverslips were mounted and stored as above. Fluorescent images were acquired on the same microscope and software as phase contrast images; with FITC-WGA visualised using a FITC filter (excitation 492/18; emission 520/20) and PI with a TRITC filter (excitation 572/23; emission 600/20) [191].

2.6.2. Starvation Tolerance Assay

To examine the phenotype of *Micromonospora* grown on nutritionally poor media, 10 µl of a serial dilution of freshly harvested spores (10^0 through to 10^{-5}) was spotted onto 1 mL of increasingly dilute GYM agar (1x GYM through to 0.125x GYM) in a 24 well plate. Spore dilutions were plated along the columns of the plate, with media dilutions down the rows of the plate.

To prepare different concentrations of GYM nutrient base, a 2x concentrated GYM broth (8 g Glucose, 8g Yeast Extract, 20 g Malt extract L^{-1}) was prepared separately from a 2x agar base (24 g agar agar powder, 4 g $CaCO_3^-$ L^{-1}). The 2x GYM broth was mixed 1:1 with the molten agar base either neat, or at 1x, 0.5x, or 0.25x dilutions in dH_2O to prepare final concentrations of 1x, 0.5x, 0.25x, and 0.125x.

After seven days of incubation at 37ºC, plates were overlaid with soft nutrient agar spiked with *P. aeruginosa* prepared as above and incubated for a further 24 hours before being photographed.

2.7. Genomic DNA Isolation from *Micromonospora* sp. O3, O5, and PH63 and isolate GBG 9AT

50 mL GYM broth was inoculated with single colonies of each respective strain and incubated for seven days (30ºC 250 RPM), after which cultures were pelleted by centrifugation at 4000xG and transferred to sterile 1.5 mL Eppendorf tubes. For lysis cells were suspended in TE buffer (50 mM Tris, 20 mM EDTA) before incubation with lysozyme (20 mg ml^{-1}) and RNase A (10 mg ml^{-1}) for one hour (37ºC 300 RPM), after which 50 µL SDS (10% w/v) was added and thoroughly mixed. 85 µL of NaCl (5 M) was added before three successive washes with phenol:chloroform:isoamyl alcohol (25:24:1) to separate lipids and cellular debris, after which the aqueous phase was transferred to a fresh Eppendorf tube. Genomic DNA was precipitated with the addition of 500 µL isopropanol (100%), mixing by inversion before centrifugation at maximum speed. The isopropanol was decanted before the pellet was washed twice in cold 100% ethanol and left to dry completely at room temperature before resuspending in 200 µL dH_2O. The 260/230 and 260/280 ratios of purified gDNA and quantity were assessed using a Nanodrop spectrophotometer and Qubit fluorometer.

2.8. Whole Genome Sequencing

Sequencing of strains was performed by MicrobesNG on an Illumina MiniSeq system with a library prepared on-site from genomic DNA isolated as described in section 2.7 to generate 250 bp paired end reads. In addition, for *Micromonospora* sp. O5 Nanopore long reads were generated using the GridION system, with genomic DNA isolation and library preparation being performed on-site. Sequencing of the three strains was performed by MicrobesNG on an Illumina MiniSeq system with a library prepared on-site from genomic DNA isolated as described in section 2.7. to generate 250 bp paired end reads. In addition, for *Micromonospora* sp. O5 and *Micromonospora* sp. PH63 nanopore long reads were generated using the GridION system, with genomic DNA isolation and library preparation being performed on-site.

2.9. Genome Assembly

An iterative approach was taken to maximise quality when assembling the genomes of the three *Micromonospora* strains. For *Micromonospora* sp. O5 and *Micromonospora* sp. PH63, long read assemblies were generated using the Flye [170] (Galaxy Version 2.9.1+galaxy0) and Canu [169] (Galaxy Version 2.2+galaxy0) algorithms. Hybrid assemblies incorporating both Illumina and Nanopore reads were generated using Unicycler [167]. The contigs generated by Unicycler were then scaffolded with MeDuSa [192] using the Flye and Canu assemblies as references. Closely related strains were identified by autoMLST [178] (Fig. 4.9.; Clade 2), and nucmer [172]was used to visualise any misassemblies compared to those related strains.

For *Micromonospora* sp. O3, Illumina reads were assembled with the Unicycler algorithm, and the assembly submitted to autoMLST to identify closely related organisms. These closely related organisms (Fig. 4.9.; Clade 1) were the used as references for MeDuSa scaffolding.

For strain GBG 9AT, Illumina reads were assembled as above for *Micromonospora* sp. O3.

2.10. Genome Mining

Initial BGC content prediction of isolates was performed using antiSMASH 6.0 [181]. To minimise the risk of false positives, the clusters counted and reported on were identified using the “strict” setting on antiSMASH, which omits BGCs unless they bear all the characteristics of their BGC type (e.g. a Type I PKS will be omitted if it lacks all the associated Type I PKS domains) unless otherwise stated. The assemblies were scored by number and type of BGC identified by antiSMASH.

We followed several routes to examine the novelty of the BGCs predicted by antiSMASH. The tool runs its BGC predictions through the ClusterBLAST alignment algorithm against two database – the MIBiG database of characterised BGCs, and the antiSMASH database of BGCs predicted from publicly available genomes. If a BGC was predicted to have >50% similarity to a BGC in one of these databases, it indicated that those clusters shared more than half of their genes and so we compared them using clinker.

To test how many of our organisms' BGCs were unique, we predicted the BGC content of each of the neighbours in their clade and used BiG-SCAPE to network these BGCs. Singletons appearing in these networks that belonged to our organisms were considered unique to them.

2.11. Identification of strains related to isolates

Preliminary multi-locus-sequence-analysis of the strains was carried out by uploading the assembled contigs in FASTA format to the autoMLST server [178] (DeNovo mode, default nearest organisms and MLST genes, IQ-Tree bootstrapping and modelfinder enabled, inconsistent MLST genes filtered), with the resulting concatenated tree permitting their placement into taxonomic clades and identification of the biosynthetic gene clusters possessed by neighbouring organisms, with *Micromonospora phaseoli* automatically selected as an outgroup.

To test if our strains represented new species, the FastANI algorithm [193] (Galaxy ver. 1.3), with a unique species in the tree being defined as having no neighbours with an ANI >95%. The phylogenetic tree was plotted and annotated using the treedataverse package [194] (ver. 0.0.1) implemented in R.

2.12. Pangenome prediction

To estimate the size of the *Micromonospora* pangenome, FASTA assemblies of the strains included in the autoMLST tree were annotated using Prokka (Galaxy Version 1.14.6+galaxy1) [195] to produce annotations in GFF3 format. Roary (Galaxy Version 3.13.0+galaxy2) [196] was then used to calculate the pangenome of these organisms with an 80% blastp identity cutoff used to group proteins [86].

To confirm all strains selected by AutoMLST as *Micromonospora*, we plotted the ANI values against proportion of query genome fragments that mapped to reference fragments for that calculation. Where *M. pataloongensis* was the reference or query sequence for FastANI stood out as separate from the rest of the members of the tree in this plot, with fewer fragments mapping. We therefore reran Roary without this strain to optimise the pangenome, with the hypothesise that *M. pataloongensis* inclusion warps the core genome calculation. We also generated pangenomes with and without split paralogs.

2.13. Codon Prediction

Codon usage for *Micromonospora* sp. O5, O3, and PH63 was calculated using the 41ordon package (ver. 1.12) [197] in R, using coding sequences predicted by Prokka as inputs. Overall codon distribution was calculated, and developmentally regulated genes were predicted as those containing TTA codons that did not encode for transfer or ribosomal RNAs. The presence of *bldA* homologs was confirmed by searching for predicted leucyl-tRNAs carrying a UAA anticodon. The secondary structures of these tRNAs as well as that of the *bldA* of *S. coelicolor* A3(2) were predicted using the RNA Fold web server [198] using the minimum free energy and partition function algorithm and compared. Alignments of the *bldA* genes from the four organsisms were generated using the ClustalW algorithm, implemented in the msa package [199] (ver. 1.28) and visualised using the ggmsa package [200] (ver. 1.3.4) in R.

2.14. Prediction of conserved developmental regulators

To identify the presence of confirmed regulators, we centred our focus around the presence or absence of members of the BldD regulon [72] in our *Micromonospora* strains. To do this, we obtained amino acid sequences of conserved regulators in *S. coelicolor* A3 (2) from StrepDB (https://strepdb.streptomyces.org.uk, Table 2.2.) and ran them against our organisms using BLASTp with default search parameters (Galaxy Version 2.10.1+galaxy2) as an initial screen, followed by BLAST Reciprocal Best Hits (RBH), again with default search parameters [201] (Galaxy Version 0.1.11) to confirm orthologs. Alignments of the best hits from BLASTp (by lowest e-value) were generated as above.

Table 2.2. *S. coelicolor* regulators selected for analysis

Locus Tag	Gene Name	Protein	Function	Reference
SCO2792	adpA	AdpA	• Cessation of DNA replication • Activation of *bldA* transcription	[135]
SCO5133	NA	Ortholog of AmfR of S. griseus	• Activates expression of spore coat peptide SapB	[202]
SCO1489	bldD	BldD	• Master repressor of developmental transition	[78]
SCO4768	bldM	BldM	• Response regulator • Initiation of sporulation	[203]
SCO3323	bldN	BldN	• Sigma factor • Required for aerial hyphae formation	[204]
SCO3324	rsbN	RsbN	• Anti sigma factor which antagonizes BldN	[205]
SCO1950	whiA	WhiA	• Initiates transition from aerial hyphae to sporulation	[206]
SCO3034	whiB	WhiB	• Small transcription factor • Regulates onset of sporulation in tandem with WhiA	[207]
SCO5621	whiG	WhiG	• Sigma factor for transcription of late sporulation genes	[208]

SCO6029	whiI	WhiI	• Transcription factor • Forms a heterodimer with WhiI to transcribe late sporulation genes	[209]

2.15. Curation of Single Contig *Micromonospora* Assemblies

To understand the evolution of specialised metabolism in *Micromonospora*, we chose to investigate the relative locations of biosynthetic gene clusters in *Micromonospora* chromosomes. To enable this, we first sought out to collate a set of single-contig assemblies of *Micromonospora*, by downloading NCBI Refseq assemblies with completion levels of either “Complete” or “Chromosome”. We oriented these assemblies using SnapGene® [210] so that they started at the first base of *dnaA*, allowing us to use *dnaA* as a reference point to compare the locations of BGCs. To identify large misassemblies in our dataset, we performed an all vs all comparison using Nucmer (Galaxy Version 4.0.0rc1+galaxy2) [172] in the Galaxy European server [188] discarding incongruent assemblies. The raw assembly data used for analysis were obtained on May 24th 2021.

Table 2.3 Assemblies Selected for and Included in Analysis

Organism	Accession	Assembly Level	Included
M. aurantiaca DSM 27029	GCA_000145235.1	Complete	Yes
M. carbonaceae aurantiaca	GCA_013389765.1	Complete	Yes
M. maris AB-18-032	GCA_000204155.1	Complete	Yes
M. cranelliae LHW63014	GCA_014764405.1	Complete	Yes
Micromonospora terminaliae DSM 101760	GCA_009671205.1	Complete	Yes
M. sagamiensis JCM 3310	GCA_014680085.1	Complete	Yes
M. aurantiaca 110B	GCA_003351365.1	Complete	Yes
M. endophytica NBRC 109090	GCA_018326245.1	Complete	Yes
Micromonospora sp. HM134	GCA_007833915.1	Complete	Yes
Micromonospora sp. B006	GCA_003408515.1	Complete	No
Micromonospora sp. L5	GCA_000177655.2	Complete	No
M. echinospora DSM 43816	GCA_900091495.1	Chromosome	Yes
M. rifamycinica DSM 44983	GCA_900090265.1	Chromosome	Yes
M. echinofusca DSM 43913	GCA_900091445.1	Chromosome	Yes
M. aurantinigra DSM 44815	GCA_900089595.1	Chromosome	Yes
M. purpureochromogenes DSM 43812	GCA_900091515.1	Chromosome	Yes
M. echinaurantiaca DSM 43904	GCA_900090235.1	Chromosome	Yes
M. zamorensis DSM 45600	GCA_900090275.1	Chromosome	Yes
M. krabiensis DSM 45344	GCA_900091425.1	Chromosome	Yes
M. viridifaciens DSM 43909	GCA_900091545.1	Chromosome	Yes
M. coriariae DSM 44875	GCA_900091455.1	Chromosome	Yes
M. chokoriensis DSM 45160	GCA_900091505.1	Chromosome	Yes
M. coxensis DSM 45161	GCA_900090295.1	Chromosome	Yes
M. inositola DSM 43819	GCA_900090285.1	Chromosome	Yes
Micromonospora sp. 28ISP2-46	GCA_013694245.1	Chromosome	Yes
M. narathiwatensis DSM 45248	GCA_900089605.1	Chromosome	Yes

M. siamensis DSM 45907	GCA_900090305.1	Chromosome	Yes
Micromonospora sp. WMMA2032	GCA_002688545.1	Chromosome	Yes
Micromonospora sp. WMMC415	GCA_009707425.1	Chromosome	Yes
M. tulbaghiae CNY-010	GCA_003612775.1	Complete	Yes

2.16. ANI Calculation of Complete *Micromonospora* assemblies

To understand the relative-relatedness of the complete *Micromonospora* assemblies in our analysis in chapter five, and *Micromonospora* closely related to the Chilean strains as identified by AutoMLST, we generated an all-vs-all average nucleotide identity (ANI) comparison using FastANI (Galaxy Version 1.3). The values generated by this were used to generate a heatmap using the gplots package (ver. 3.1.3.) [211] in R.

2.17. Biosynthetic Gene Cluster Locus Mapping

To map the loci of BGCs present in the dataset of single contig assemblies, we utilised antiSMASH 6.0 [181]. Sequence data from our dataset was exported in FASTA format and uploaded to antiSMASH, with BGCs being predicted under the antiSMASH "strict" parameters to reduce the risk of false-positive BGCs being identified.

Once the BGCs present in our assemblies were identified, we then calculated the midpoint of the BGCs with the formula: $B_M = \frac{B_1+B_2}{2}$; where B_M is the midpoint of a given BGC; B_1 is the the position of the first nucleotide in the BGC, and B_2 is the the position of the last nucleotide in the BGC. To control for variation in chromosome size, we then normalised B_M on a scale of 0-100 using the formula $NB_M = 100(\frac{B_M}{S})$, where NB_M is the normalised midpoint of a given cluster and S is the size in base pairs of the chromosome the BGC occupies.

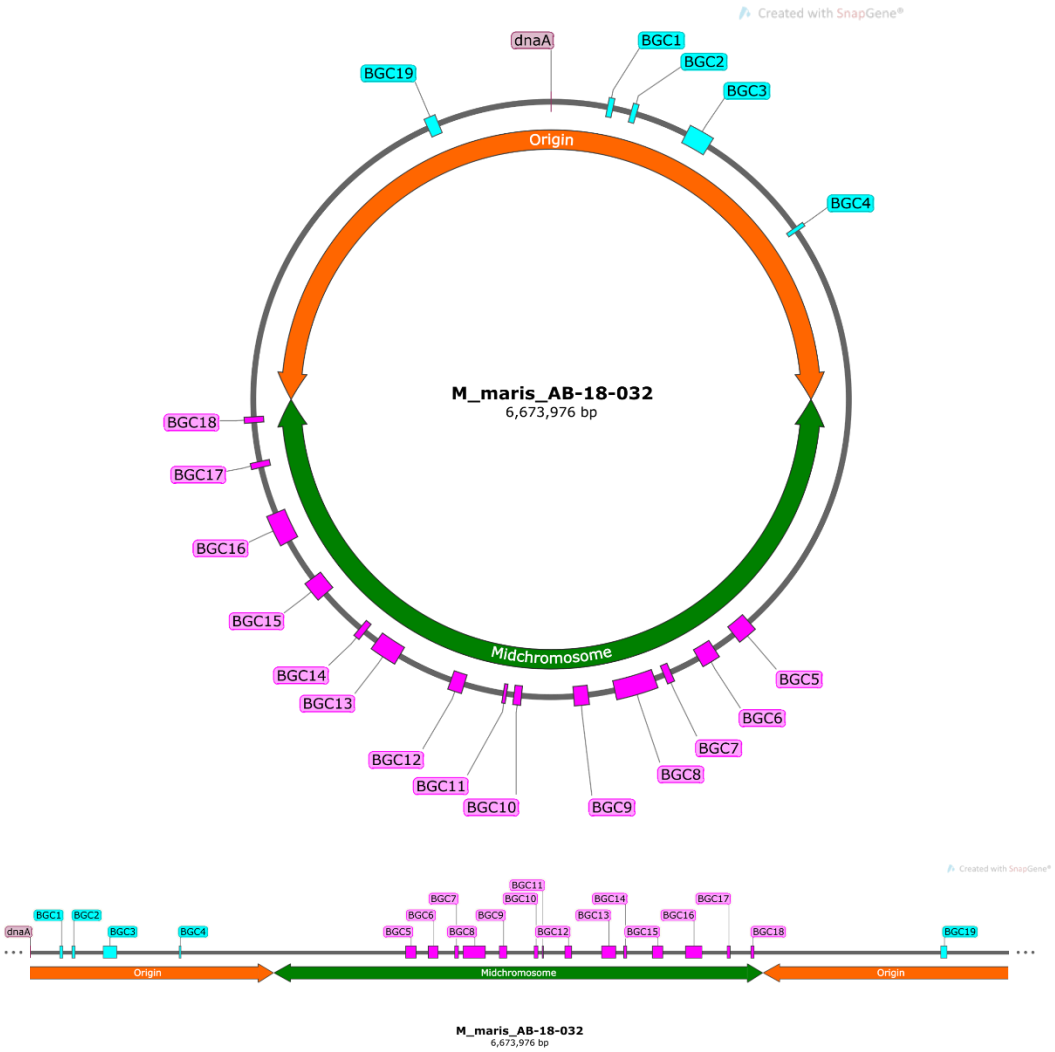


Figure 2-1 Linear transposition of the *M. maris* AB-18-032 chromosome.

BGCs are designated as ori-proximal (cyan) or ori-distal (magenta).

2.18. Between-Cluster Comparisons

To test if BGCs of the same type at similar loci between chromosomes were homologous, we utilised BiG-SCAPE [182] and Clinker [124]. BiG-SCAPE enabled us to generate networks of related BGCs, and Clinker used to generate alignments of syntenous BGCs to investigate their gene content. These networks were then further investigated and using Cytoscape software. Clinker was used to generate pairwise alignments between BGCs present in a network, which were coloured according to their functions predicted by antiSMASH.

2.19. Ecological Modelling of *Micromonospora* BGCs

To test if there was a difference in the diversity of BGCs present in either half of the *Micromonospora* chromosome, the BGCs in our dataset were split into two sets: “origin-proximal”, i.e. a quarter chromosome upstream and downstream of *dnaA*, or “midchromosome”, i.e. a quarter chromosome upstream and downstream of the midchromosome. These were analysed by BiG-SCAPE and the resulting presence-absence tables were then used to calculate the Shannon, Simpson, and Inverse-Simpson diversity indices using the vegan package [212] (ver. 2.6-2) in R.

2.20. Strand Bias Testing

To compare gene strand bias between the organisms in our dataset we annotated the reoriented genomes using Prokka (Galaxy Ver. 1.14.6) to obtain GFF3 format files. The files were imported using the read.gff function in ape (ver. 5.6-2) [213], filtered for genes, with their position normalised using the same calculation applied for BGCs. GFF3 format files assign “+” or “-“ to indicate if genes are on the top or bottom strand of a DNA molecule – to convert this to leading or lagging, we considered bottom strand genes on the left replichore to be leading, and top strand genes to be lagging. The opposite was applied to genes in the right replichore. We also calculated the GC skew of the organisms using SkewIT [214].

Table 2.4 Decision matrix for designation of strands as leading or lagging

	Left Replichore	**Right Replichore**
Top Strand	Lagging	Leading
Bottom Strand	Leading	Lagging

Chapter 3 : Establishing and Screening a Strain Collection of Actinobacteria for Anti-pseudomonad Bioactivity, and Characterising Three Hyper-Arid *Micromonospora* Isolates

3.1. Introduction

As previously discussed, AMR is a critical threat to humanity and the discovery of novel drugs is a key facet to the fight against it. Previously, a number of *Micromonospora* strains were isolated from Chilean desert soil to be screened for novel bioactive metabolites. Here, we sought to carry out this screen and characterise the biology of any bioactive *Micromonospora.* To that end, we set out with three aims:

1. To use the OSMAC approach to screen these *Micromonospora* for the ability to inhibit the growth of *P. aeruginosa*

2. To generate a bank of novel isolates from Scottish soil to screen in parallel to this

3. To microscopically analyse bioactive Chilean isolates to gain an insight into their morphology

3.2. Results

3.2.1. Antibacterial activity of isolates

Following growth on solid media, plugs were taken of *Micromonospora* sp. O3, *Micromonospora* sp. O5, and *Micromonospora* sp. PH63 and placed over soft Nutrient Agar seeded with an indicator organism. Reproducible zones of inhibition were seen around *Micromonospora* sp. O3 grown on ISP 2 and *Micromonospora* sp. O5 grown on ISP 3 and ISP 5 (Fig. 3.1.). In addition, one isolated from the rhizosphere of a *Leucadendron argeneum* was bioactive: GBG 9AT. This strain inhibited the growth of *A. baumanii*, and *P. aeruginosa* when grown on ISP 3 or ISP 4, and inhibited *E. coli* on ISPs 3 and 5 (Fig. 3.2.). As these strains were bioactive, we elected to sequence their genomes (The genomes of *Micromonospora* sp. O3 and *Micromonospora* sp. O5 are discussed further in Chapter 3, the genome of strain GBG 9AT is discussed in this chapter).

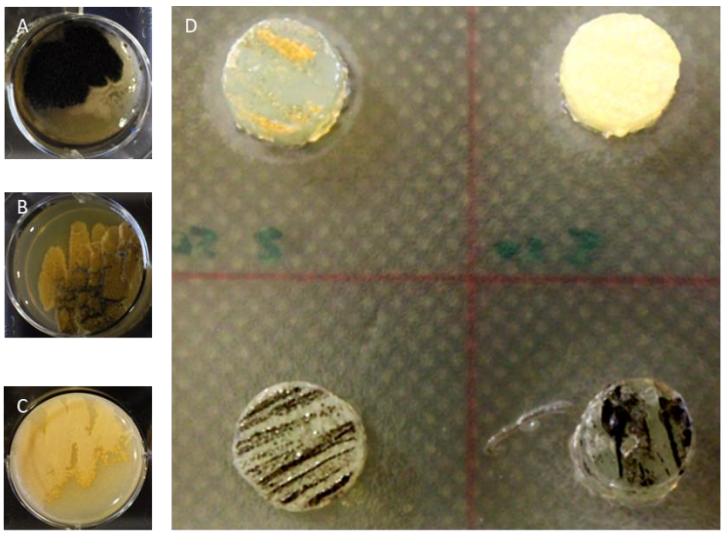


Figure 3-1 Morphology and bioactivity of *Micromonospora* strains.

Morphology of (**A**) Oligo 3 on ISP2 (**B**) Oligo 5 on ISP2 (**C**) Oligo 5 on ISP3 (**D**) Zones of *P. aeruginosa* inhibition around plugs of *Micromonospora* sp. O3 and *Micromonospora* sp. O5.

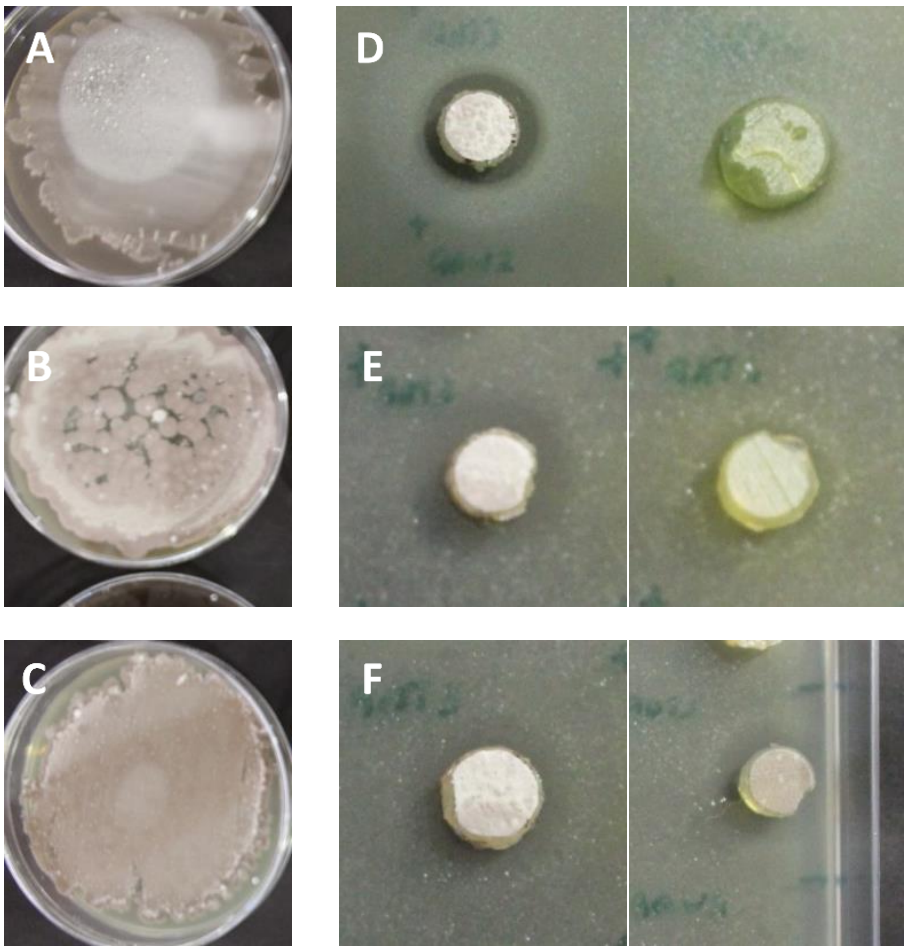


Figure 3-2 Antibacterial activity of isolate GBG 9AT.

Morphology of GBG 9AT on (**A**) ISP 3 (**B**) ISP 4 (**C**) ISP5. Bioactivity of 9AT against (**D**) *P. aeruginosa* [Left: ISP 3; Right: ISP 4] (**E**) *A. baumanii* [Left: ISP 3; Right: ISP 4] (**F**) *E. coli* [Left: ISP 3; Right: ISP 5].

3.2.2. GBG 9AT is a strain of *Streptomyces xanthophaeus*

As well as observing antipseudomonad bioactivity from *Micromonospora* strains O3 and O5, we identified a putative streptomycete which inhibited the growth of *P.* aeruginosa. The potent bioactivity of strain GBG 9AT raised it as a candidate for whole genome sequencing. Genomic DNA was extracted for the strain, sequenced by MicrobesNG through the Illumina platform, and assembled in Unicycler to generate a 783 contig, 8.2 Mbp assembly. This assembly was then submitted to the AutoMLST server, generating a phylogenetic tree that identified GBG 9AT as a strain of *Streptomyces xanthophaeus* (Fig. 3.3). with an estimated ANI of 99.4% to *S. xanthophaeus* NRRL B-3004 and 99.3% to *S. xanthophaeus* NRRL B-5414 The assemblies of both *S. xanthophaeus* strains were then used as references in MeDuSa to scaffold the assembly of GBG 9AT before submitting the assembly to antiSMASH for BGC prediction. Scaffolding reduced the number of predicted BGCs to 33 from 36, with a second round of scaffolding reducing this to 32 BGCs. Notably, scaffolding helped recapitulate eight fragmented BGCs into larger hybrid BGCs and helped uncover missing adjacent genes (Fig. 3.4.). Three BGCs were predicted from the scaffolded assembly that were absent from the unscaffolded assembly, and two *vice versa*.

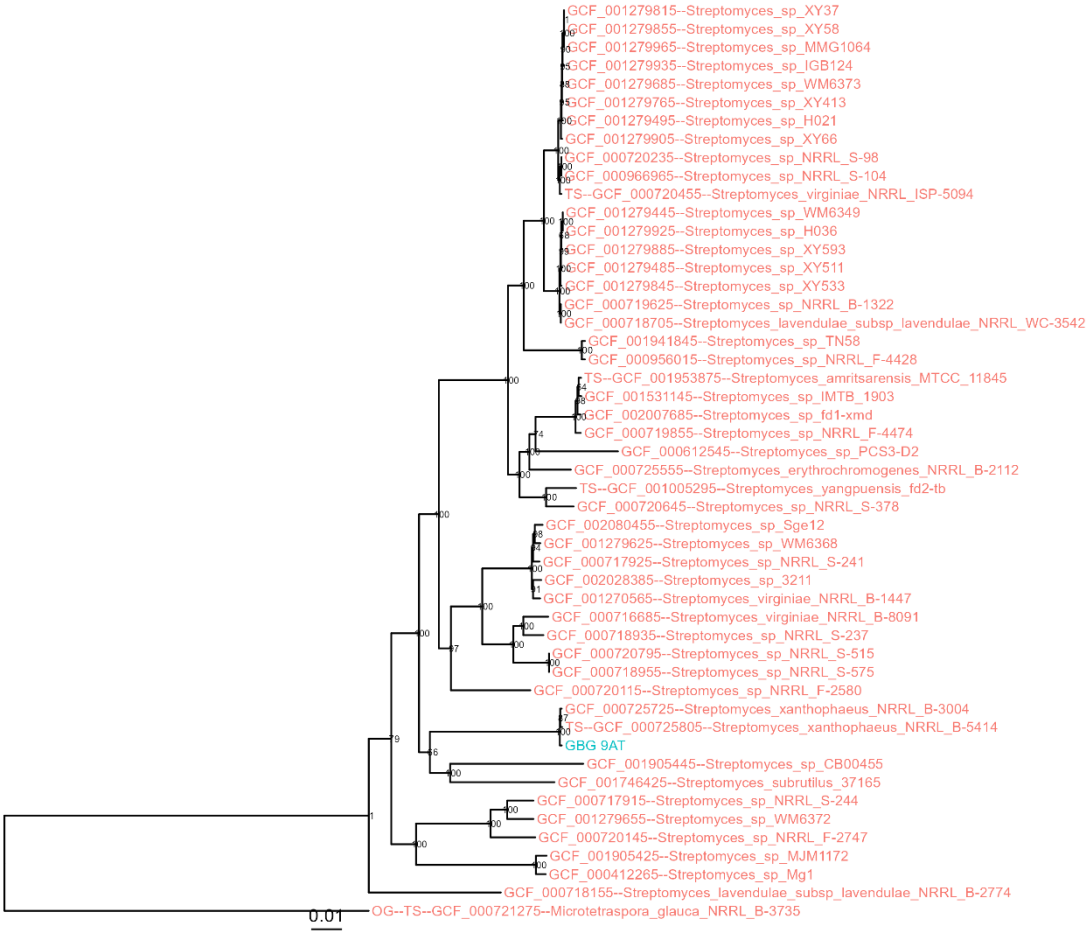


Figure 3-3 AutoMLST generated phylogeny of strain GBG 9AT.

GBG 9AT (Blue) clustered most closely with *Streptomyces xanthophaeus* strains NRRL B-3004 and NRRL B-5414. Scale bar: nucleotide substitutions per site.

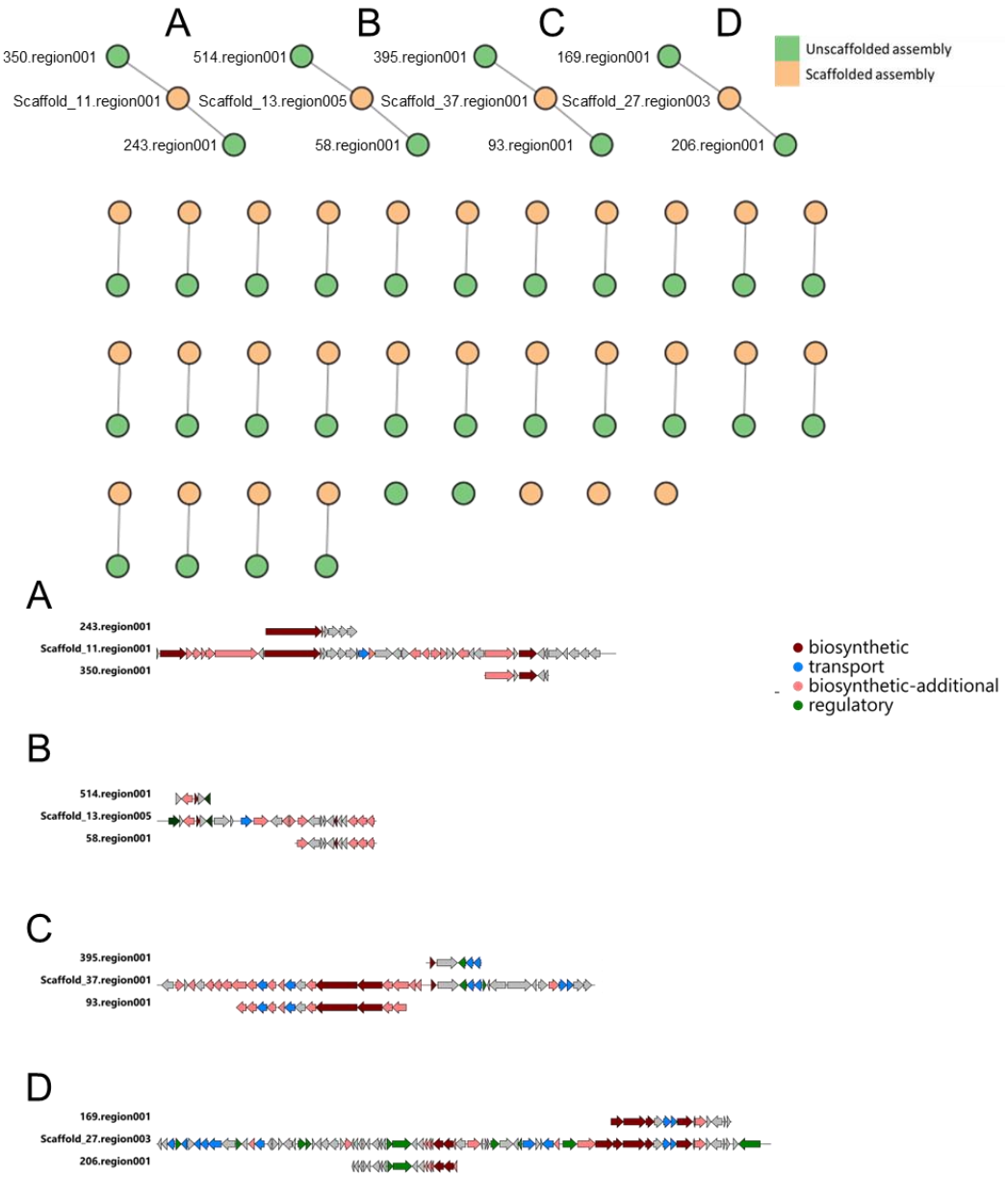


Figure 3-4 Scaffolding recapitulates BGCs fragmented by Unicycler.

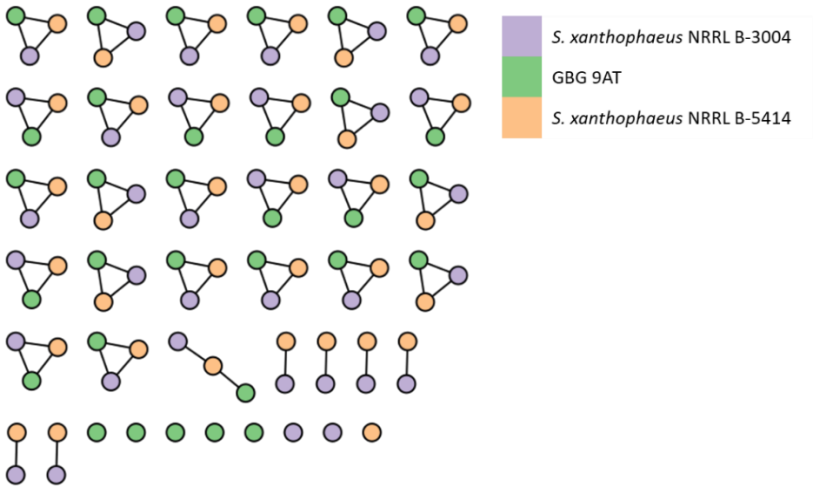
Top panel: Network of GBG 9AT's BGCs before and after one round of scaffolding against *Streptomyces xanthophaeus* assemblies, as predicted by antiSMASH. Scaffolding identified genomically close BGCs and extended the predicted boundaries of the BGC.

Bottom panel: alignments of networks A, B, C, and D generated by clinker, which illustrate that scaffolding improves BGC prediction

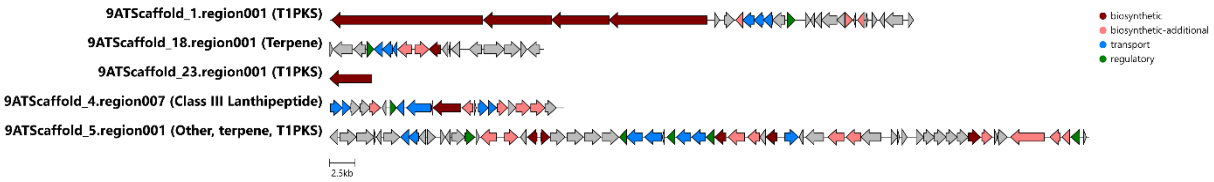
3.2.3. Specialised metabolism is highly conserved between *S. xanthophaeus* and its related strains

After demonstrating the utility of scaffolding to improve BGC prediction from fragmented assemblies we also compared the BGC content predicted from the scaffolded GBG 9AT assembly to the BGCs of *S. xanthophaeus* NRRL B-3004 and *S. xanthophaeus* NRRL B-5414. BiG-SCAPE networking of the antiSMASH predicted BGCs from the three isolates revealed highly conserved secondary metabolism within the species – 27 out of the 41 gene cluster families predicted contained a representative BGC from each isolate queried (Fig. 3.5. A). GBG 9AT contains five BGCs that were unique to it (3.5.B). Of these, one was a T1PKS gene present by itself on a contig, one was a terpene cluster, one was a large T1PKS gene cluster at a contig boundary. Lastly, a BGC encoding machinery for biosynthesis of a SapB type lanthipeptide fell into a node by itself, despite the core lanthionine synthase and prepeptide being 100% conserved. This separation from the SapB BGCs of *S. xanthophaeus* NRRL B-3004 and *S. xanthophaeus* NRRL B-5414 is attributable to differences in the gene content of the right side of the BGC (Fig. 3.5.C)

A



B



C

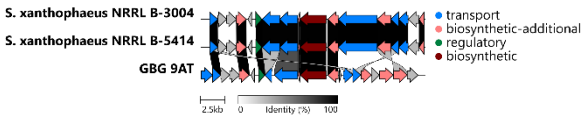


Figure 3-5 Unique BGCs of GBG 9AT.

(**A**) Network of BGCs predicted from GBG 9AT, *S. xanthophaeus* NRRL B-3004, and *S. xanthophaeus* NRRL B-5414. GBG 9AT was predicted to encode five unique BGCs. (**B**) Clinker illustration of the singleton BGCs of GBG 9AT. The three smallest clusters represented a single T1PKS gene, an ectoine cluster, and a type III lanthipeptide cluster encoding for SapB biosynthesis. The other two BGCs were a large modular T1PKS and a hybrid other, terpene, T1PKS class cluster (**C**) Clinker alignment of the SapB clusters from GBG 9AT and the two related *S. xanthophaeus* neighbours illustrating the difference in gene content between GBG 9AT's SapB BGC and the SapB of its neighbours.

3.2.4. Development of *Micromonospora* sp. PH63 is under carbon catabolite repression

As strain GBG 9AT was a previously characterised organism, we chose instead to focus on strains of *Micromonospora* for further characterisation. The effects of carbon catabolite repression on the gene expression of *Streptomyces* is well studied [128, 129], and so we sought to examine if nutrient starvation may affect the observed phenotypes of *Micromonospora* sp. PH63, *Micromonospora* sp. O3, and *Micromonospora* sp. O5.

To characterize the growth of our organisms on nutrient poor media, we grew them on GYM media with progressively poorer nutrient content. The development of *Micromonospora* sp. PH63 was accelerated under poor nutrient conditions, with the onset of sporulation (judged by production of black spore pigment) accelerated on 0.5x GYM agar compared to 1x for undiluted spores after seven days of growth (Fig. 3.6.). The relationship between media richness and earlier germination broke down at lower inoculums, possibly reflecting less competition for nutrients for those colonies. Development of *Micromonospora* sp. O3 was unaffected by media richness at high inoculum density, however at low CFU development was repressed in poor media. *Micromonospora* sp. O5 was unaffected by media richness. After seven days of growth, the wells were overlaid with *P. aeruginosa* spiked soft nutrient agar to assay for antibiotic induction – however there was no difference in observed *P. aeruginosa* inhibition between nutrient poor agar with *Micromonospora* and *Micromonospora* free controls.

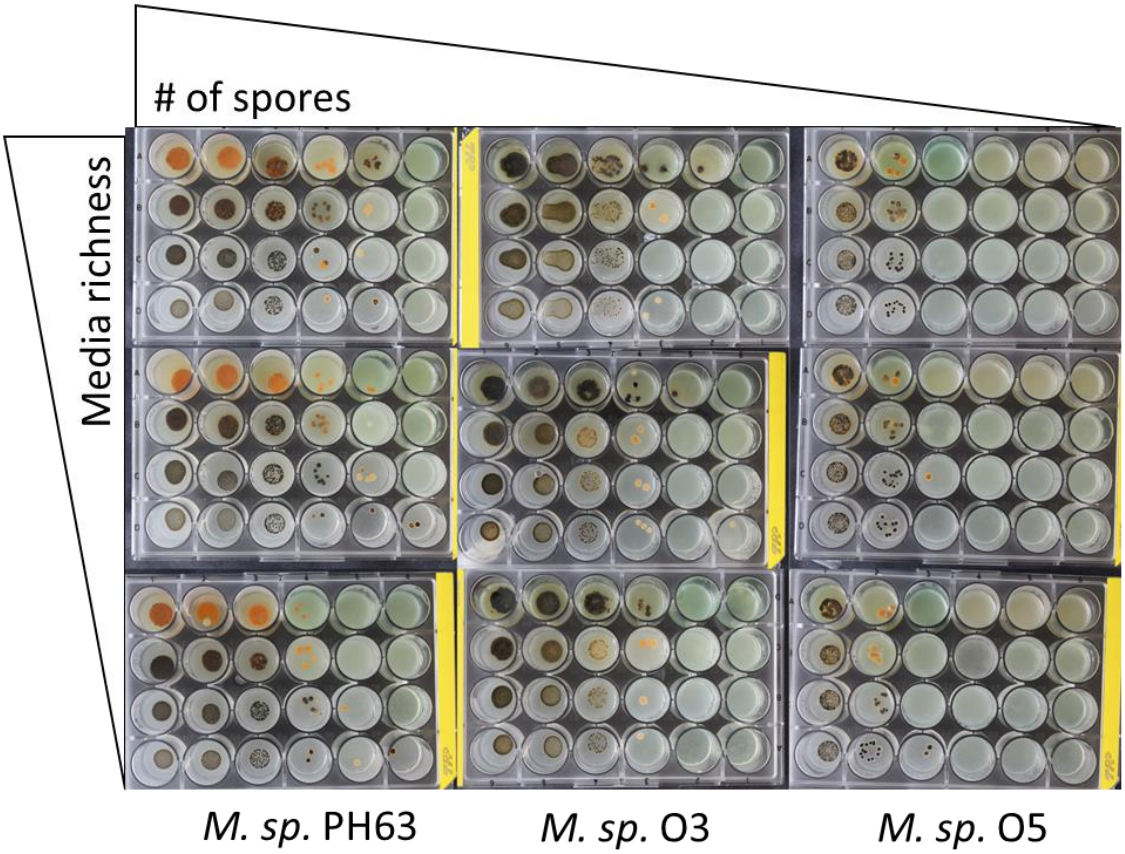


Figure 3-6 Carbon Catabolite mediated repression of *Micromonospora* sp. *PH63* development.

Morphology of Chilean *Micromonospora* isolates grown on 2-fold dilutions of GYM agar after seven days. *Micromonospora* sp. PH63 showed repressed sporulation on 1x GYM compared to 0.5x whereas *Micromonospora* sp. O3 and *Micromonospora* sp. O5 did not.

3.2.5. Vegetative *Micromonospora* hyphae fluoresce under ultraviolet light, whereas sporulating hyphae absorb it

We hypothesised that the pigmentation of the three Chilean *Micromonospora* may be protective against ultraviolet radiation, a particular problem in the Atacama Desert. We imaged colonies of *Micromonospora* sp. PH63, *Micromonospora* sp. O5, and *Micromonospora* sp. O3. The orange pigmented vegetative hyphae at the colony edges fluoresced under UV illumination, whereas the darker brown/black spore-bearing hyphae strongly absorbed the UV light (Fig. 3.7.). This suggests that pigmentation of the colonies has a role to play in the colonies' response to UV irradiation.

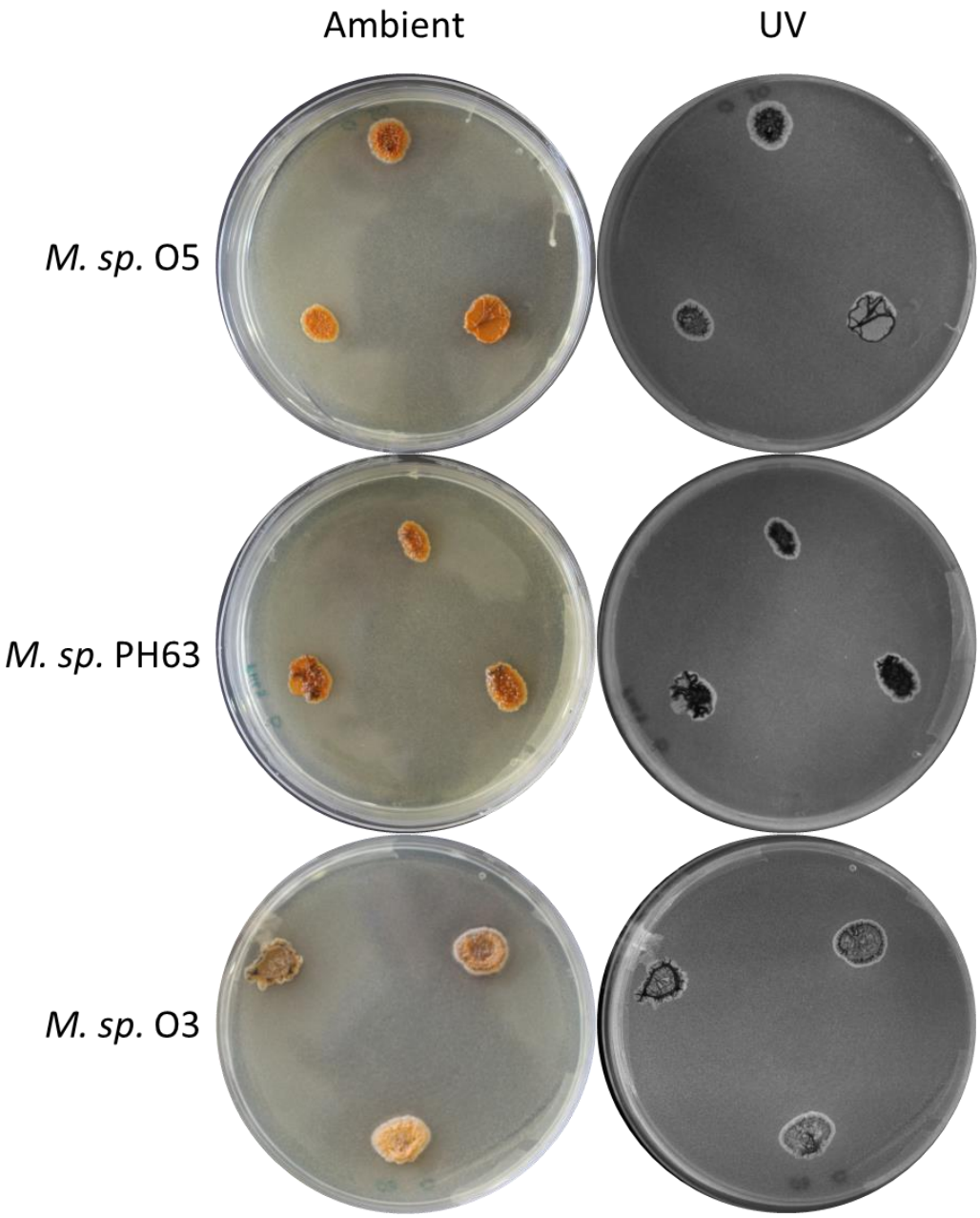


Figure 3-7 Colonies of Chilean *Micromonospora* both absorb and fluoresce under UV light.

Vegetative hyphae at the colony edges fluoresced, whereas spore-bearing hyphae absorbed the UV.

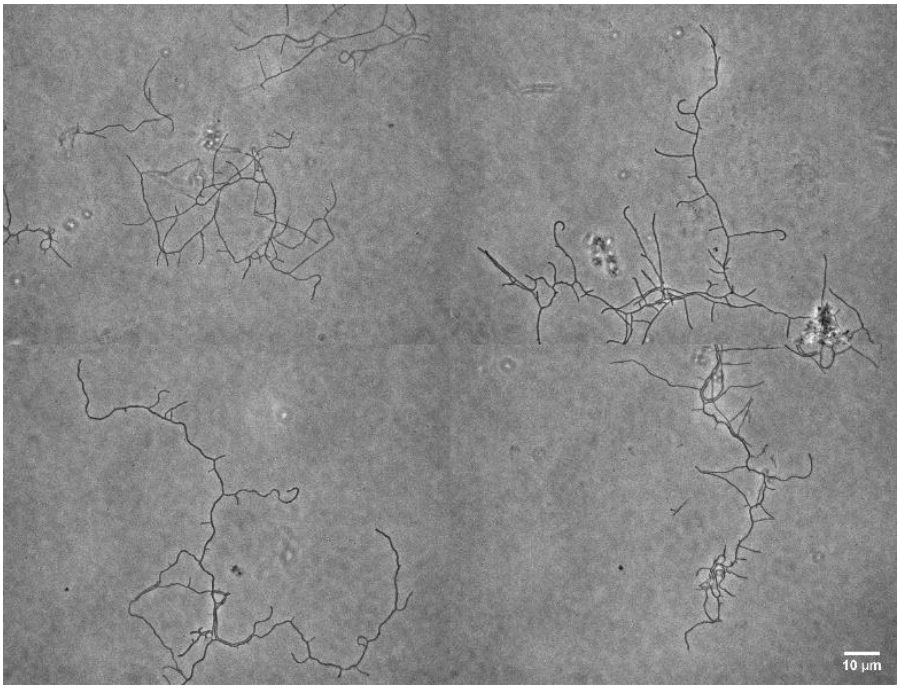
3.2.6. Micromorphology of Isolates

In order to better understand the morphology of our strains we elected to observe them microscopically. Discussed below are observations gathered from observations made of agar-grown mycelium from *Micromonospora* sp. O5, *Micromonospora* sp. O3, and *Micromonospora* sp. PH63.

3.2.6.1. Micromorphology of *Micromonospora* sp. O5

The vegetative hyphae of *Micromonospora* sp. O5 are highly branched and infrequently curved, extending straight from their branch point away from their central colony. Spore bearing hyphae are characterised by frequent branching of short hyphae, which develop engorged monospores at the hyphal tip. Dense sporulation occurs at the centre of colonies. Extensive simultaneous branching suggests coordinated onset of sporulation (Fig. 3.8.). There was a mean distance of 5.43 ± 0.29 µm between branches, and spores were a mean diameter of 0.99 ± 0.012 µm (Fig. 3.9.).

A



B

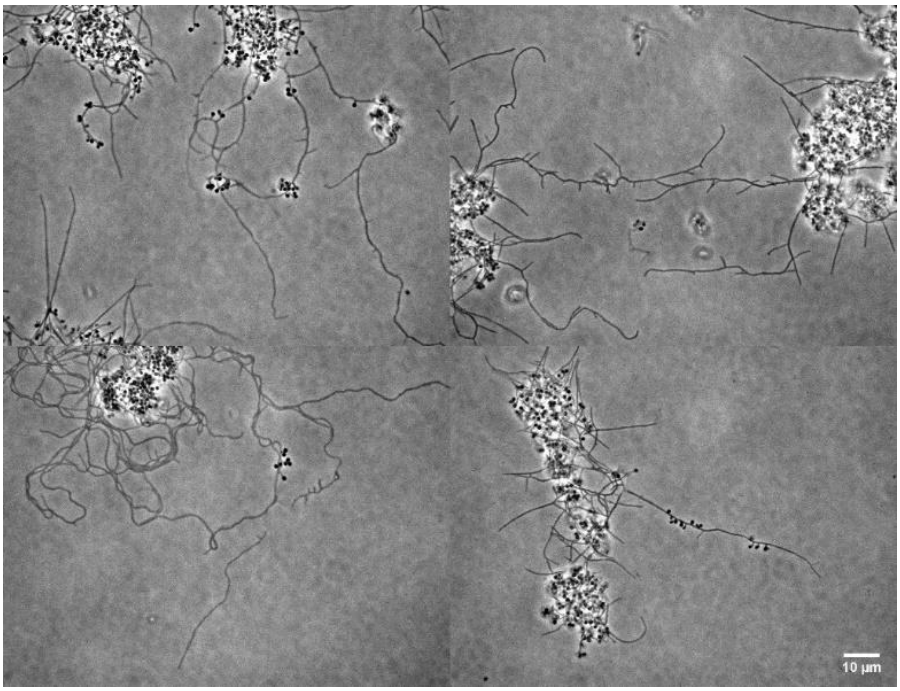


Figure 3-8 Micromorphology of *Micromonospora* sp. *O5*.

(**A**) Vegetative hyphae of five day old culture (**B**) Spores of five day old hyphae.

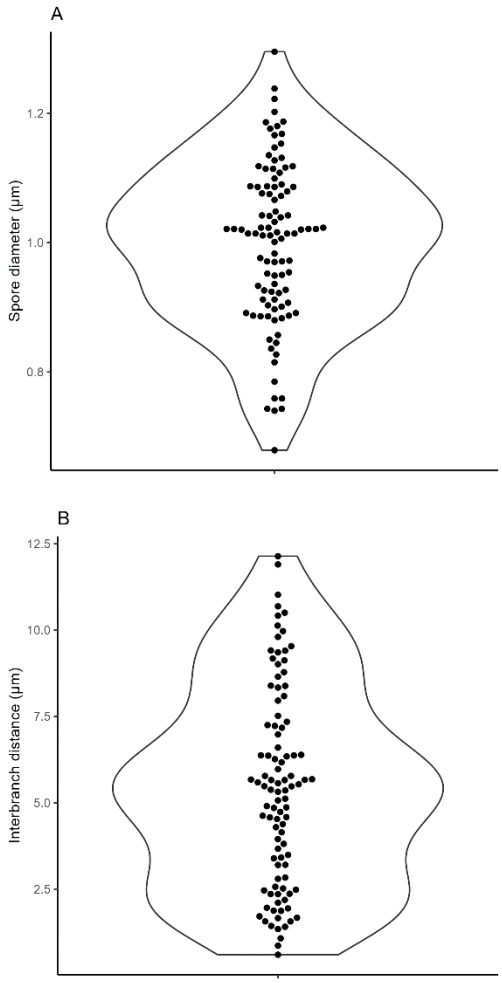


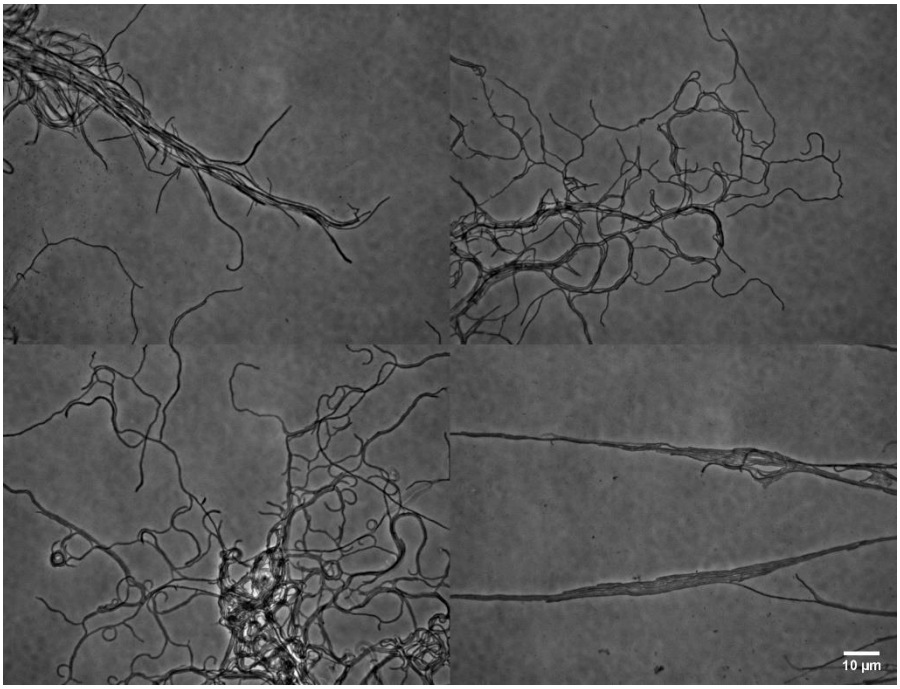
Figure 3-9 *Micromonospora* sp. *O5* Micromorphology measurements.

Measurements of (**A**) Spore diameter of five day old culture and (**B**) Interbranch distance of five day old culture. N=100 for each measurement.

3.2.6.2. Micromorphology of *Micromonospora* sp. O3

The vegetative hyphae of *Micromonospora* sp. O3 branch infrequently, extending outward from the colony and frequently curving (Fig. 3.10). The hyphae grow in parallel bunches as wide as 5.4 µm (mean 2.05 ±0.09 µm) across (Fig. 3.11) – *Micromonospora* sp. O3 was the only one of the three Chilean *Micromonospora* to exhibit this morphology. Monospores are borne on short branches from the vegetative hyphae were visible from day five of growth, are phase-dark, round, and on average 0.95 ±0.013 µm in diameter. Sporulation occurred in clusters, which suggests coordination and simultaneous development. Representative images of hyphae and spores are given in Figure 1.3. with measurements in Fig. 1.4. of spore diameter, bundle width, and interbranch distance.

A



B

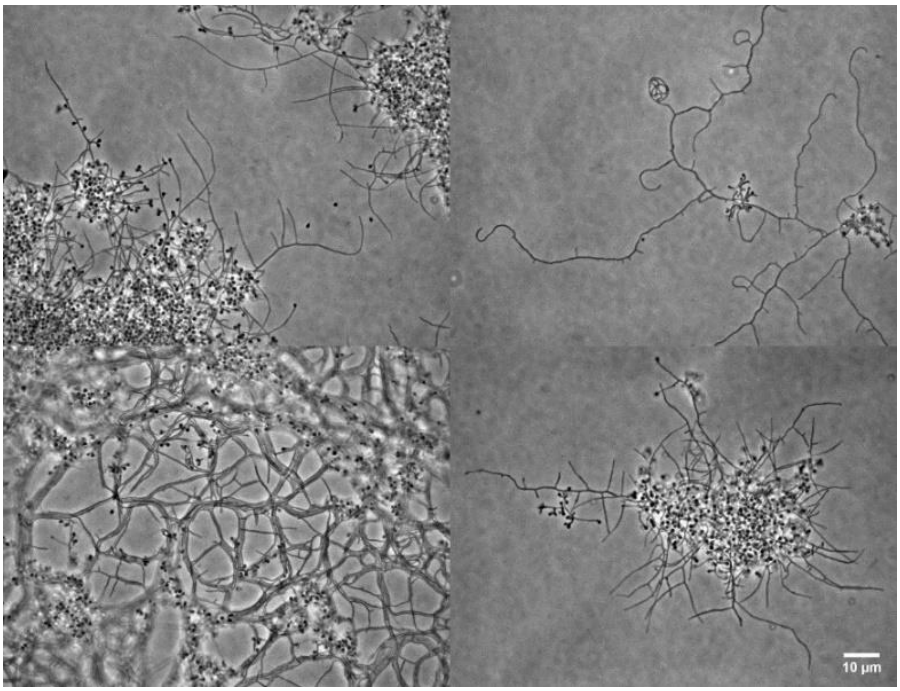


Figure 3-10 Micromorphology of *Micromonospora* sp. *O3*.

(**A**) Representative images of 4-day old hyphae illustrating bundles. (**B**) Representative images of spore morphology in five day old culture.

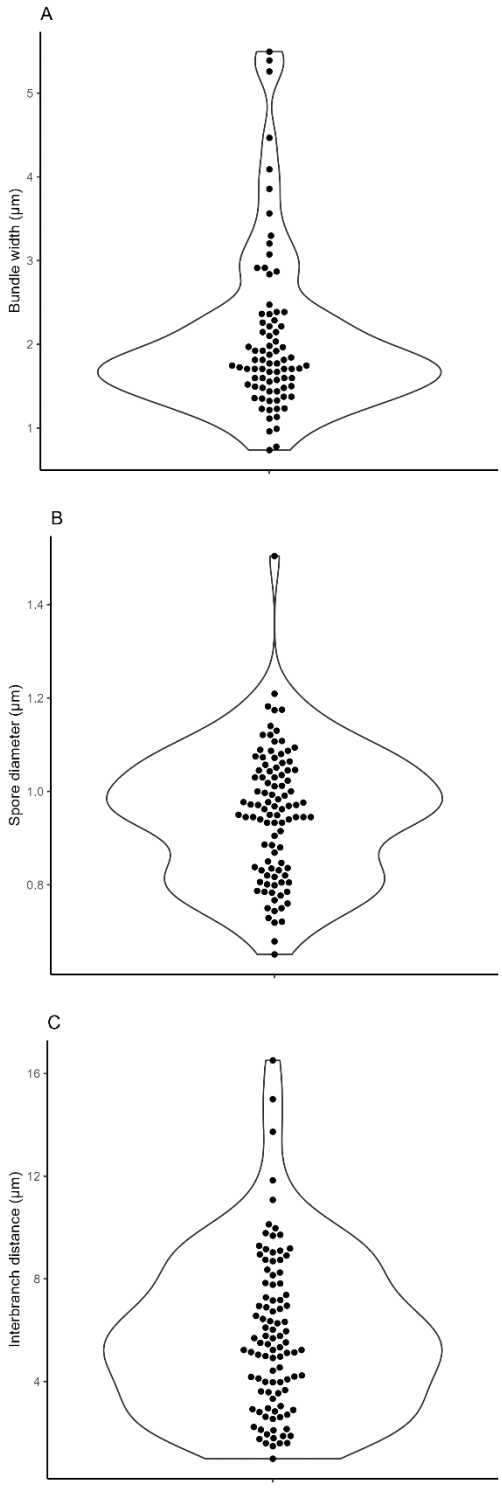


Figure 3-11 Micromonospora sp. O3 Micromorphology measurements.

Measurements of (**A**) Width of parallel hyphal bundles (**B**) spore diameter and (**C**) Interbranch distance. N=100 for each measurement.

3.2.6.3. Micromorphology of *Micromonospora* sp. PH63

The vegetative hyphae of *Micromonospora* sp. PH63 are highly branched and infrequently curved, extending straight from their branch point away from their central colony. Spore bearing hyphae are characterised by frequent branching of short hyphae, which develop engorged monospores either singly or in pairs at the tip. SEM of the spores revealed warty protrusions from their surface (Fig. 3.13.). PI/FITC-WGA staining revealed multinucleate hyphae (Fig. 3.12.), segmented by peptidoglycan crosswalls. There was a mean of 5.31 ± 0.45 μm between branches, and the spores of *Micromonospora* sp. PH63 were a mean diameter of 0.9 ± 0.015 μm (Fig. 3.14).

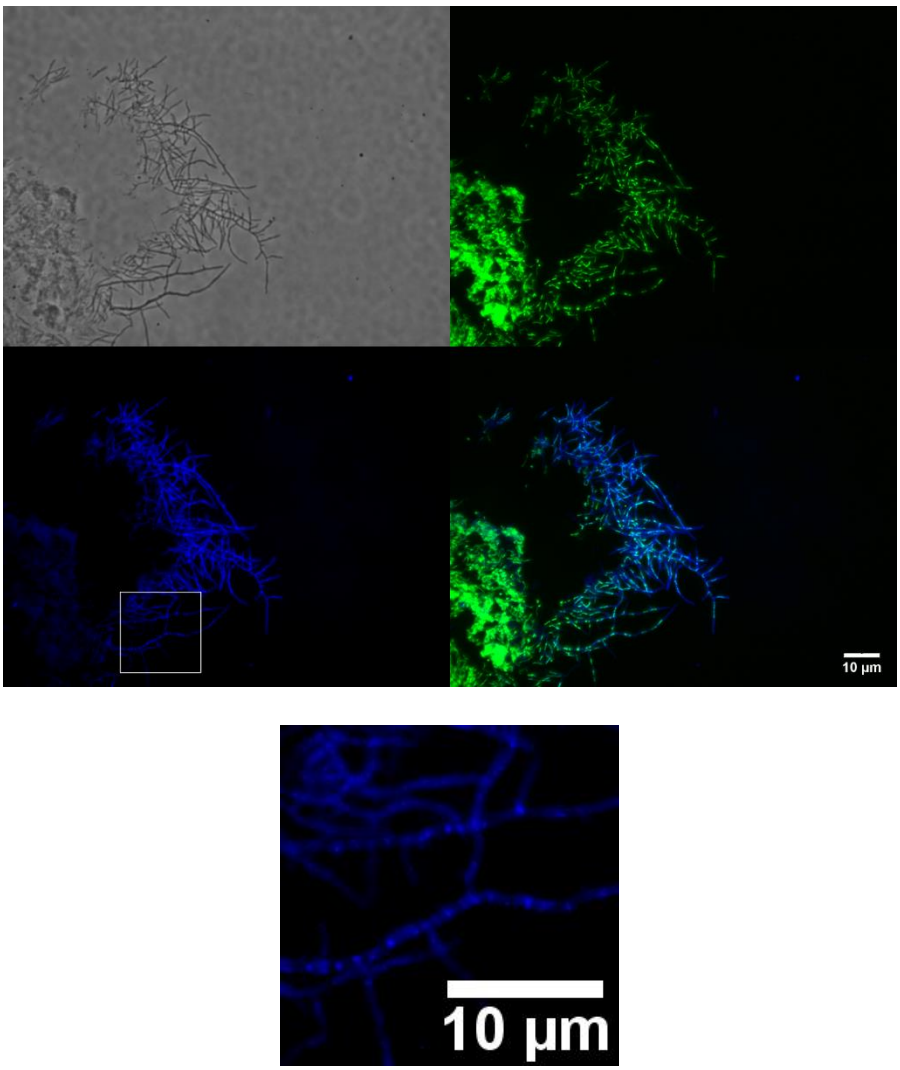
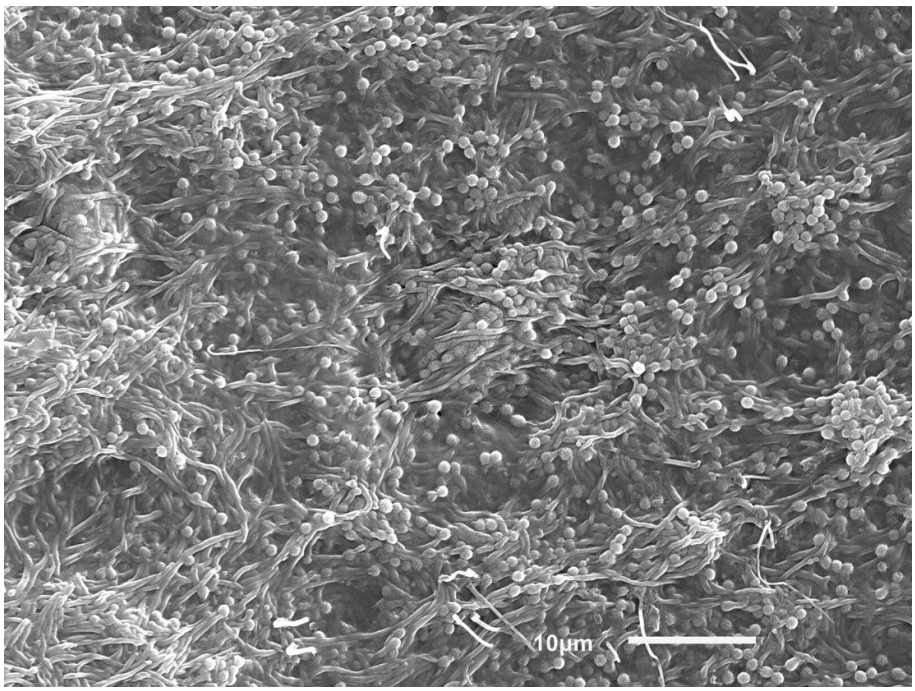


Figure 3-12 Multinucleate hyphae and peptidoglycan cross walls of *Micromonospora* sp. *PH63.*

Widefield phase contrast (top left) and epifluorescence (top right, bottom) of three day old hyphae. Nucleoids stained with PI (Green), peptidoglycan stained with FITC-WGA (Blue). Bright FITC foci indicate peptidoglycan crosswalls.

A



B

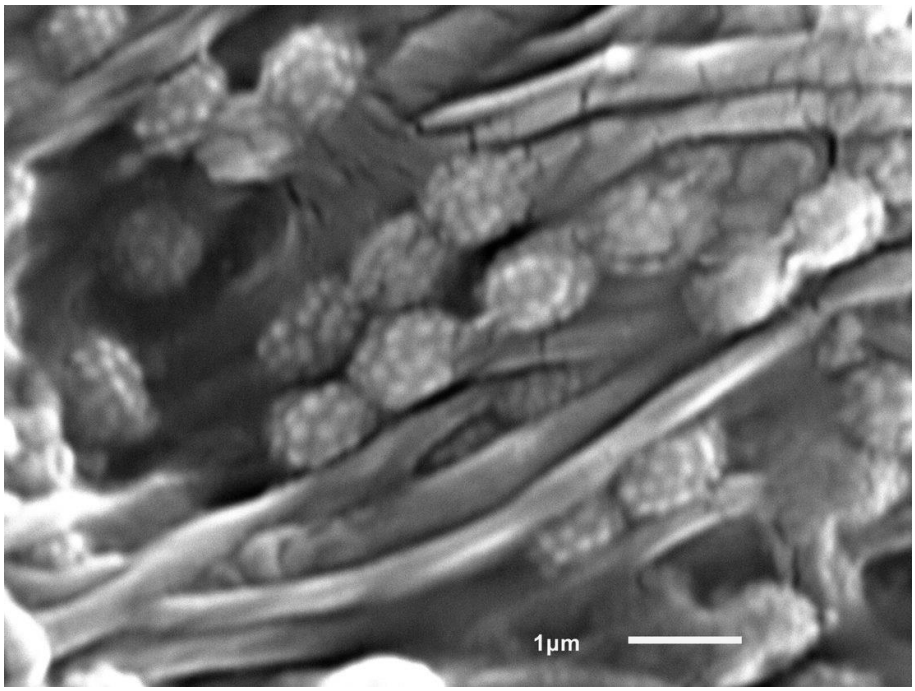
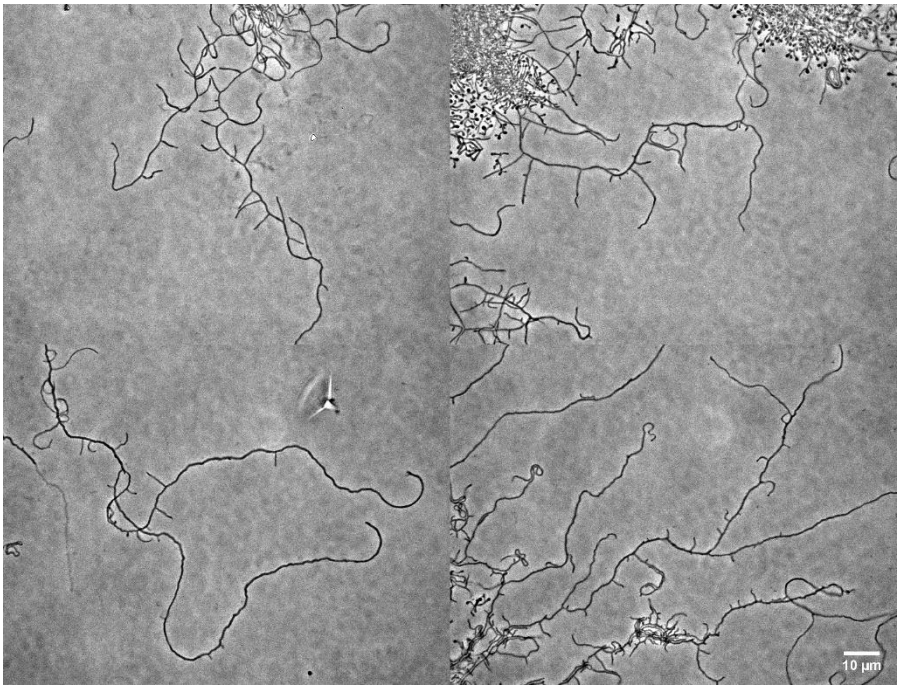


Figure 3-13 Scanning electron microscopy of *Micromonospora* sp. *PH63*.

(**A**) Micrograph of seven day old culture grown on GYM, showing dense sporulation (**B**) Warty spores of *Micromonospora* sp. PH63.

A



B

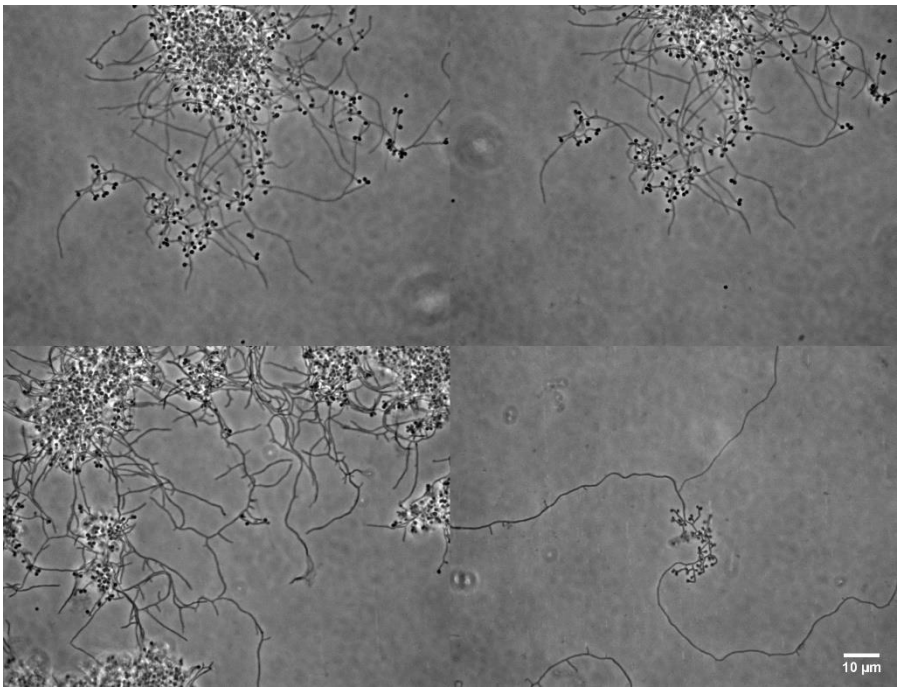


Figure 3-14 Representative widefield phase contrast images of M. sp. PH63.

(**A**) Vegetative hypha of 4-day old culture (**B**) Mature monospores of six-day old culture.

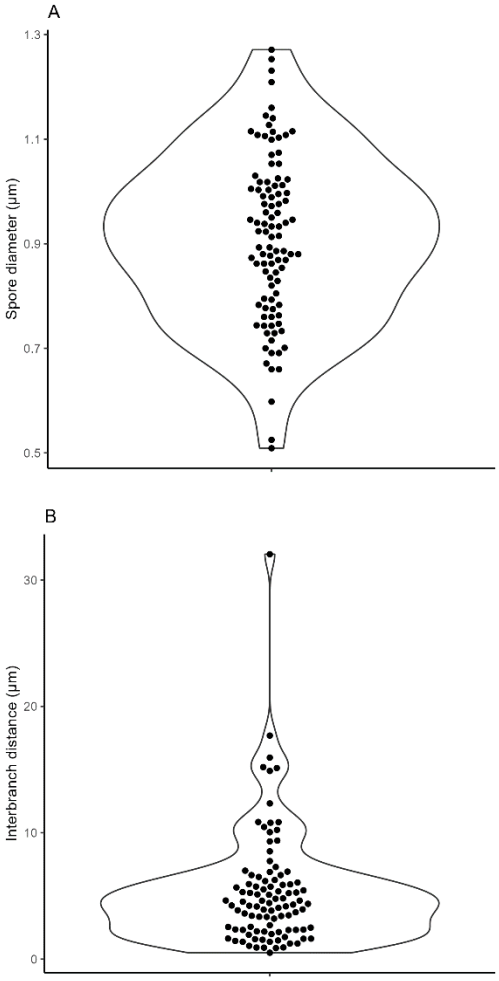


Figure 3-15 *Micromonospora* sp. *PH63* Micromorphology measurements.

Measurements of (**A**) Spore diameter of six day old culture and (**B**) Interbranch distance of four day old culture. N=100 for each measurement.

3.3. Discussion

3.3.1. One strain, many compounds, little bioactivity

Combatting AMR requires the development of novel antibiotics, and the development of those in turn requires the characterisation of novel organisms [215]. Here, we attempted to isolate a diversity of actinobacteria from a diversity of soils in order to maximise the odds of discovering novelty. In particular, the Glasgow Botanic Gardens were selected as a sampling site for the diversity of both the plants it plays host to, but also for the range of ambient temperatures present in its various glasshouses [216, 217]. Unfortunately, only one isolate from Scottish samples displayed anti-pseudomonad activity, a reflection *P. aeruginosa*'s ability to survive antibiotic exposure [218].

We chose to employ the OSMAC approach to elicit cryptic BGCs from our organisms, specifically by culturing them on ISP formulated growth medium. Originally developed to aid in phenotypic taxonomy of *Streptomyces* [126], these nutritionally diverse media elicit diverse phenotypes from actinomycetes cultured on them. As well as the immediately obvious morphological phenotypes, this nutritional diversity can also induce antibiotic biosynthesis [125]. The importance of employing the OSMAC approach is obvious considering that all bioactivity observed here was media dependent. GBG 9AT produced antibiotics on ISPs 3, 4, and 5 – with antibacterial activity against *P. aeruginosa, A. baumanii, and E. coli*. These media vary in both complexity and level of definition, and it remains to be seen what specifically induces the antibacterial activity we observed and if this activity is derived from the products of one or multiple BGCs. Strains of *S. xanthophaeus* have been shown to biosynthesise a number of interesting molecules including phytotoxins and glucosidase inhibitors [219]; guanitrypmycins [220]; rishirilides and lupinicidin A [221]; and the pyroglutamyl peptidase inhibitor benarthin [222-224]. Previous work has identified strains of *S. xanthophaeus* which can inhibit the growth of *P. aeruginosa* [225] however this is not a universal property of the species and suggests diversity of secondary metabolism. Whilst we can conclude that GBG 9AT certainly produces an antibiotic with anti Gram-negative properties, the nature of this antibiotic, and whether it is unique to GBG 9AT or distributed throughout members of *S. xanthophaeus* remains to be seen.

3.3.2. Exploiting phylogeny to improve genome mining

We compared the BGC predictions from GBG 9AT to the two *S. xanthophaeus* strains that autoMLST identified as being closely related to it. There was a high level of conservation of BGCs between the three organisms – suggestive of high levels of vertical inheritance [89, 143]. Interestingly, all three organisms contained BGCs capable of biosynthesising the SapB lanthipeptide responsible for spore hydrophobicity [226].

3.3.3. The morphology of *Micromonospora* sp. PH63, *Micromonospora* sp. O3, and *Micromonospora* sp. O5 bears the classic hallmarks of *Micromonospora*

The genus *Micromonospora* is morphologically characterised by colonies of orange-to-red pigmented vegetative hyphae, which darkens to brown-black with the onset of sporulation [49, 68, 69]. To that end, *Micromonospora* sp. PH63, *Micromonospora* sp. O3, and *Micromonospora* sp. O5 bear the classic phenotypes of *Micromonospora.* All three organisms strongly absorbed ultraviolet light. In *Salinispora*, hyphal pigment has been demonstrated to be the glycosylated carotenoid sioxanthin – however, the nature of spore pigment remains to be elucidated [227].

The parallel hyphal growth of *Micromonospora* sp. O3 is reminiscent of that of *Micromonospora chalcea* [69] and occurred more commonly than was observed in *Micromonospora* sp. O5 and *Micromonospora* sp. PH63, who primarily displayed this growth in older, spore bearing hyphae– suggesting it may be a developmentally regulated trait. The molecular driver of this growth mode remains to be seen. The strain is closely related to multiple endophytic strains [228, 229], and so this coordinated polar growth may aid in penetration of plant tissue. Little is known about the control of growth direction in *Micromonospora,* although they are predicted to encode a number of membrane associated signal transduction proteins that may mediate directionality [230].

All three strains sporulated in clusters which suggest simultaneous onset of sporulation. In liquid culture in defined minimal media, sporulation of *Micromonospora echinospora* occurs after ~10 days and is associated with production of extracellular proteases which may be essential to sporulation [231].

Of the three *Micromonospora* discussed here, we were best able to characterize the morphology of *Micromonospora* sp. PH63. By staining for nucleoids with propidium iodide and peptidoglycan with FITC-conjugated wheat germ agglutinin [190] we

observed multinucleate hyphal compartments demarcated by peptidoglycan cross walls. Compartmentalised hyphae have been previously observed in *Streptomyces* [75, 232] with septa biosynthesised by a distinct divisome from sporulation septa [75]. We also blunt spiny protrusions from the surface of the spores of *Micromonospora* sp. PH63, analogous to those of the closely related *M. echinospora* [233]. Spore morphology varies widely between actinomycetes and such has been used as a taxonomical marker, such as with *Streptomyces* species [234, 235].

3.3.4. Nutrient availability and early sporulation

Despite their phylogenetic relatedness, *Micromonospora* sp. PH63 and *Micromonospora* sp. O5 showed differential responses to nutrient starvation, with *Micromonospora* sp. PH63 sporulation being repressed on nutrient rich media whilst *Micromonospora* sp. O5 sporulation was not. The primary metabolism of *Streptomyces* species is closely intertwined with both secondary metabolism and their growth and development cycle [128, 236]. Despite this, it remains to be seen what role this plays in *Micromonospora*.

4. Summary

As demonstrated by this chapter, antibacterial activity against *P. aeruginosa* is a rare phenotype. Of the strains we isolated, only GBG 9AT was reproducibly able to inhibit the growth of this pathogen. We identified this strain as an isolate of *S. xanthophaeus*. As we sought novelty, we chose to not investigate *S. xanthophaeus* any further owing to its extensive characterisation and the conservation of BGCs in our strains as compared to other *S. xanthophaeus* isolates.

Thus, we chose to proceed with three *Micromonospora* isolates. *Micromonospora* sp. O5 and *Micromonospora* sp. O3 displayed bioactivity against *P. aeruginosa*. As *Micromonospora* sp. PH63 had been previously sequenced but not analysed, we chose to include this organism for genome mining in the work described in the next chapter. All three strains showed morphology characteristic of *Micromonospora*, with septate hyphae observed by widefield epifluorescence microscopy of *Micromonospora* sp. PH63.

Chapter 4 Assembly and Annotation of Hyper-Arid *Micromonospora* Chromosomes, and Investigating the Evolution of Specialised Metabolism in *Micromonospora*

4.1. Introduction

With the bioactivity and morphology of *Micromonospora* sp. O5, *Micromonospora* sp. O3, and *Micromonospora* sp. PH63 characterised – as well as evidence of *Micromonospora* sp. O5 and *Micromonospora* sp. PH63 being able to inhibit the growth of *P. aeruginosa* – we next sought to gain insight into the genetics that underpin their biology. To achieve this, we undertook a whole genome sequencing study of the three Chilean strains, to enable mining of their genomes for BGCs. In this study, we sought to answer three questions:

1. What is the biosynthetic potential of the three Chilean *Micromonospora* strains?
2. How might the BGCs of these organisms be regulated, and what insights can their genomes give us to the organisms' development?
3. Where do our strains fall in the phylogenetic tree of *Micromonospora*?

4.2. Results

4.2.1. Whole genome assembly of *Micromonospora* genomes

4.2.1.1. MeDuSa Improved Fragmented Assemblies

The first step we undertook was sequencing the strains using the MicrobesNG service. Genomic DNA of *Micromonospora* sp. O5 and *Micromonospora* sp. PH63 were sequenced using both long read Oxford Nanopore Technologies sequencing and short read Illumina sequencing. *Micromonospora* sp. O3 was sequenced using only Illumina sequencing. Details of the assembly process are described in Materials and Methods.

The MeDuSa scaffolder [192] uses user-submitted reference assemblies as templates to improve a query assembly, based on a presumption of collinearity between the query assembly and reference assembly or assemblies. We chose to use this to improve the assemblies of our organisms. Oligo 3, which was assembled using only Illumina short reads, benefited the most from this, with 135 contigs being reduced to 12 scaffolds. Scaffolding was also able to reduce the assemblies of *Micromonospora* sp. O5 and *Micromonospora* sp. PH63 to single contiguous scaffolds (Fig. 4.1.). BUSCO analysis of how many single-copy orthologs were present in the assemblies out of what would be expected suggests the assemblies were nearly complete, at >95% complete BUSCOs.

The chromosome of *Micromonospora* sp. O3 was assembled as 12 scaffolds. The sum of these scaffolds is 6583245 bp, of which 99.4% are on a single scaffold. The GC content of the organism is 72.24%. The organism contains 6,117 CDS, three rRNA genes, one tmRNA gene, and 64 tRNAs. The chromosome of *Micromonospora* sp. O5 was recapitulated as a single 7474902 bp scaffold. The GC content of the organism is 72.87%. *Micromonospora* sp. O5 encodes 6528 CDS, 19 rRNA genes, 76 tRNAs, three repeat regions, and one tmRNA. Lastly, the chromosome of *Micromonospora* sp. PH63 was recapitulated as a single 6996828 bp scaffold. The GC content of the organism is 72.87%. *Micromonospora* sp. O5 encodes 6071 CDS, 12 rRNA genes, 73 tRNAs, and one tmRNA.

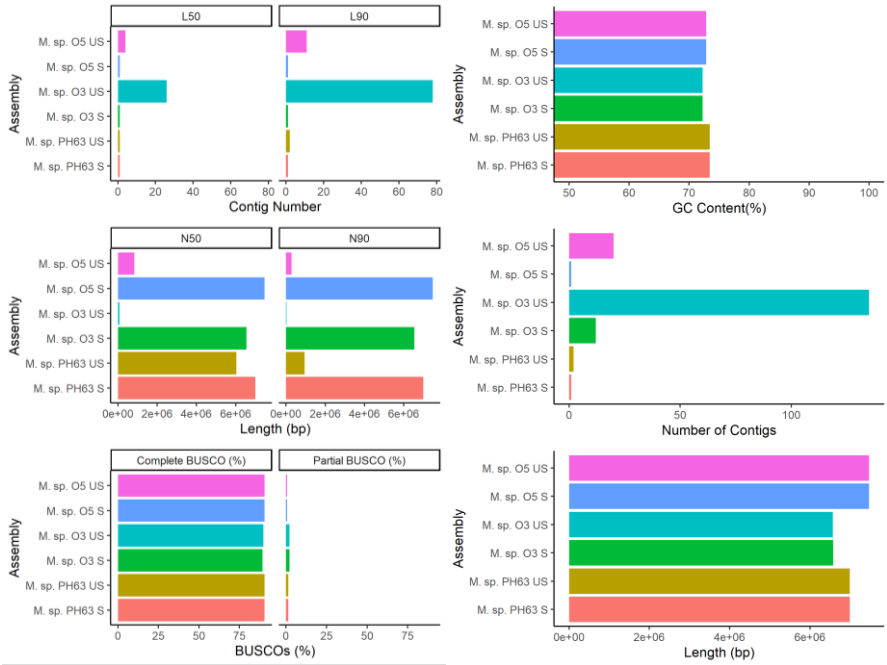


Figure 4-1 Scaffolding improves contiguity of *Micromonospora* assemblies.

Quality checks of our assemblies after scaffolding by MeDuSa. The assemblies are characteristic of *Micromonospora* with large, high GC content genomes. MeDuSa improved the assemblies by ordering them and reducing the number of contigs they were in. US = Unscaffolded assemblies, S = Scaffolded against closely related strains.

4.2.2. Pathway analysis of Chilean *Micromonospora* strains

We employed the RAST server and SEED viewer to predict the metabolic pathways present in the Chilean *Micromonospora*. All three organisms generated similar predictions, and were rich in genes encoding for carbohydrate, protein, and amino acid metabolism (Fig. 4.2.). This suggests that all three *Micromonospora* are generalists, hypothetically capable of utilising a range of nutritional substrates.

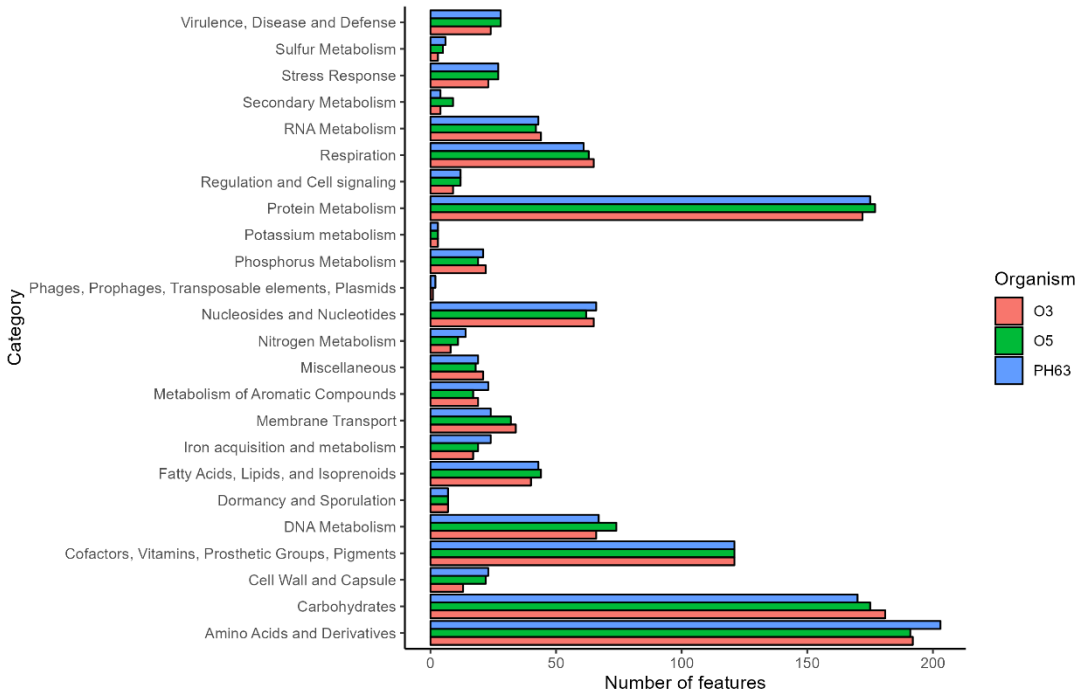


Figure 4-2 SEED subsystem category predictions of the Chilean *Micromonospora*.

All three organisms are enriched for amino acid metabolism, protein metabolism, and carbohydrate metabolism.

4.2.3. Developmental regulation is conserved in *Micromonospora* sp. O3, *Micromonospora* sp. O5, and *Micromonospora* sp. PH63

In *Streptomyces*, the developmental transition from vegetative hyphae to spore-bearing aerial hyphae is tightly regulated by a cascade of developmental proteins starting with the c-di-GMP mediated dimerization of BldD [72], and is often synchronised to antibiotic production. To interrogate the three Chilean isolates for the presence of genes involved this cascade, we carried out BLASTP and BLAST RBH searches of identified regulatory proteins from *S. coelicolor* A3(2) against their predicted proteomes (Fig. 4.3.).

4.2.3.1. *Micromonospora* sp. O3, *Micromonospora* sp. O5, and *Micromonospora* sp. PH63 do not encode an AdpA ortholog

In *S. coelicolor*, AdpA inititates the transition from vegetative growth to the erection of aerial hyphae in response to γ-butyrolactone accumulation . The best hits from our three *Micromonospora* were annotated as CdhR, a separate AraC type transcriptional regulator. These were <50% identical to the AdpA of *S. coelicolor*, which was 78 aa larger than the *Micromonospora* proteins. Based on this, we concluded that these proteins were unlikely to be AdpA orthologs. That said, all 4 proteins were predicted to have conserved domain architecture (Appendix 2; Fig. 7.1.B)

4.2.3.2. *Micromonospora* sp. O3, *Micromonospora* sp. O5, and *Micromonospora* sp. PH63 do not encode an AmfR ortholog

AmfR in *S. griseus* is a participant in AdpA mediated regulation [202], and is essential for aerial mycelium formation. In all three of our organisms, the best hits against an AmfR like ortholog in *S. coelicolor*, were less than 50% identical. In addition, the hit from *M. sp.* O3 contained was annotated as a mycothiol acetyltransferase with an approximately 80 AA region in the middle of the protein which did not align to any of the three other proteins. All four proteins contain c-terminal N-acetyltransferase machinery (Appendix 2; Fig 7.2. B). This evidence was not enough for us to conclude that our three *Micromonospora* encode AmfR.

4.2.3.3. *Micromonospora* sp. O3, *Micromonospora* sp. O5, and *Micromonospora* sp. PH63 encode a BldD ortholog

In *S. coelicolor*, BldD is the master repressor of morphological and physiological development [78]. Our analysis suggests that all three *Micromonospora* genomes encode BldD orthologs – the amino acid sequences of the best BLASTP hits aligned well to BldD of *S. coelicolor* (Fig. 4.4, Appendix 2 Fig. 7.3. A) and all three contained both the DNA binding domains and c-di-GMP binding domains of BldD (Fig. 7.3. B). Based on this, we concluded that all three *Micromonospora* strains possess BldD orthologs

4.2.3.4. *Micromonospora* sp. O3, *Micromonospora* sp. O5, and *Micromonospora* sp. PH63 are unlikely to encode a BldM ortholog

BldM is a response regulator, which heterodimerises with WhiI to control expression of its target genes [203]. In all three *Micromonospora*, the best BLASTP hits against BldM all having amino acid identity less than 41%. These proteins also aligned poorly to BldM (Fig. 4.4, Appendix 2 Fig 7.4. A) and based on domain architecture are likely unrelated response regulators (Fig. 7.4. B). Thus, we conclude it is unlikely that any of the three sequenced *Micromonospora* encode a BldM ortholog.

4.2.3.5. *Micromonospora* sp. O3, *Micromonospora* sp. O5, and *Micromonospora* sp. PH63 do not encode a BldN ortholog

BldN acts as a sigma factor, controlling the transition to aerial hyphae development, with roles such as initiating transcription of *bldM*. The three best hits from our sequenced *Micromonospora* are likely also sigma factors – the C terminal domains of all 4 proteins align well, however the N terminal of the three *Micromonospora* proteins did not align well to BldN (Fig. 4.4., Appendix 2 Fig. 7.5. A). All 4 proteins encode C-terminal domains characteristic of sigma factors (Fig. 7.4. B). Based on this, we do not believe that the three *Micromonospora* sequenced encode BldN orthologs.

4.2.3.6. *Micromonospora* sp. O3, *Micromonospora* sp. O5, and *Micromonospora* sp. PH63 do not encode a RsbN ortholog

RsbN acts as an anti-sigma factor, antagonizing BldN and preventing transcription of BldN-regulated genes. The bests hits from the three *Micromonospora* all aligned poorly to the RsbN of *S. coelicolor* (Fig 4.4., Appendix 2. Fig. 7.6. A), and were likely all sigma factors, based on their domain architecture (Fig. 7.6. B) – based on this, and the presumptive lack of BldN in our organisms, we concluded they lack RsbN.

4.2.3.7. *Micromonospora* sp. O3, *Micromonospora* sp. O5, and *Micromonospora* sp. PH63 encode a WhiA ortholog

WhiA is widely distributed amongst bacteria and required for completion of sporulation in *Streptomyces* [206, 208]. The best hits from our three *Micromonospora* aligned well to the WhiA of *S. coelicolor*, (Fig. 4.4., Appendix 2 Fig. and contained domains characteristic of WhiA – as well as producing a positive BLAST RBH result (Fig. 4.3.). This led us to conclude that all three *Micromonospora* encode WhiA.

4.2.3.8. *Micromonospora* sp. O5, and *Micromonospora* sp. PH63 encode a WhiB ortholog

In *Micromonospora* sp. O5 and *Micromonospora* sp. PH63, a WhiB type transcriptional regulator aligned well to the WhiB of *S. coelicolor* (Fig. 4.4., Appendix 2 Fix 7.16), with 82.456% amino acid identity in both cases (Fig. 4.3). The best hit from *Micromonospora* sp. O3 produced no such alignment, with the best hit only having 39% identity to the WhiB of *S. coelicolor*. Although all 3 hits were WhiB type regulators (Fig. 7.17), and so we could not rule out *Micromonospora* sp. O3's best hit being an ortholog of the *S. coelicolor* WhiB.

4.2.3.9. *Micromonospora* sp. O3, *Micromonospora* sp. O5, and *Micromonospora* sp. PH63 do not encode a WhiG ortholog

The closest hit to *S. coelicolor*'s WhiG from all three *Micromonospora* were proteins annotated as SigF, a distinct sigma factor from WhiG [208], and bore only 31% identity to WhiG (Figure 4.3.). The three *Micromonospora* proteins aligned well to eachother, but poorly to WhiG from *S. coelicolor* (Fig. 4.4, Appendix 2 Fig. 7.18.). All 4 contained domains characteristic of sigma factors, however the three *Micromonospora* proteins were predicted to be SigF class sigma factors whereas *S. coelicolor*'s WhiG was, unsurprisingly, annotated as a WhiG class sigma factor (Fig. 7.19). Based on this, consider it unlikely for the *Micromonospora* to encode WhiG.

4.2.3.10. *Micromonospora* sp. O3, *Micromonospora* sp. O5, and *Micromonospora* sp. PH63 do not encode a WhiI ortholog

WhiI in *Streptomyces* acts by forming heterodimers with BldM, acting as a response regulator. The best hit from *Micromonospora* sp. O3 was annotated as a protein-glutamate methylesterase/protein glutamine glutaminase; the best hit from

Micromonospora sp. O5 was annotated as an oxygen regulatory protein NreC; and the best hit from *Micromonospora* sp. PH63 was annotated as a transcriptional regulator DegU. All three proteins bore less than 35% amino acid identity to WhiI (Fig. 4.3.), aligned poorly to it (Fig. 4.4., Appendix 2. Fig. 7.20) and lacked the predicted C-terminal DNA binding residues and dimerization interface of WhiI (Fig. 7.21.) Thus, we concluded that the three *Micromonospora* lack WhiI.

Despite not generating a BLAST RBH hit, the best BLASTP hits for BldD and BldN were both over 60% amino acid (AA) identity. We generated multiple alignments of the best hits from each organism against each protein to further examine the conservation of these putative regulatory proteins. (Fig. 4.4., Appendix 1.). These alignments confirmed that the Chilean *Micromonospora* did not contain orthologs of AdpA, AmfR, BldM, RsbN, WhiG, or WhiI. It did, however, identify BldD and WhiA within all three *Micromonospora*. WhiB from *Micromonospora* sp. PH63 and *Micromonospora* sp. O5 aligned well to the WhiB of *S. coelicolor*, however the best hit from *Micromonospora* sp. O3 aligned poorly to it – despite being annotated as a WhiB type regulator. These results illustrate that developmental regulation is only partially conserved between the Chilean *Micromonospora* and *S. coelicolor,* and that it is not conserved between the three organisms.

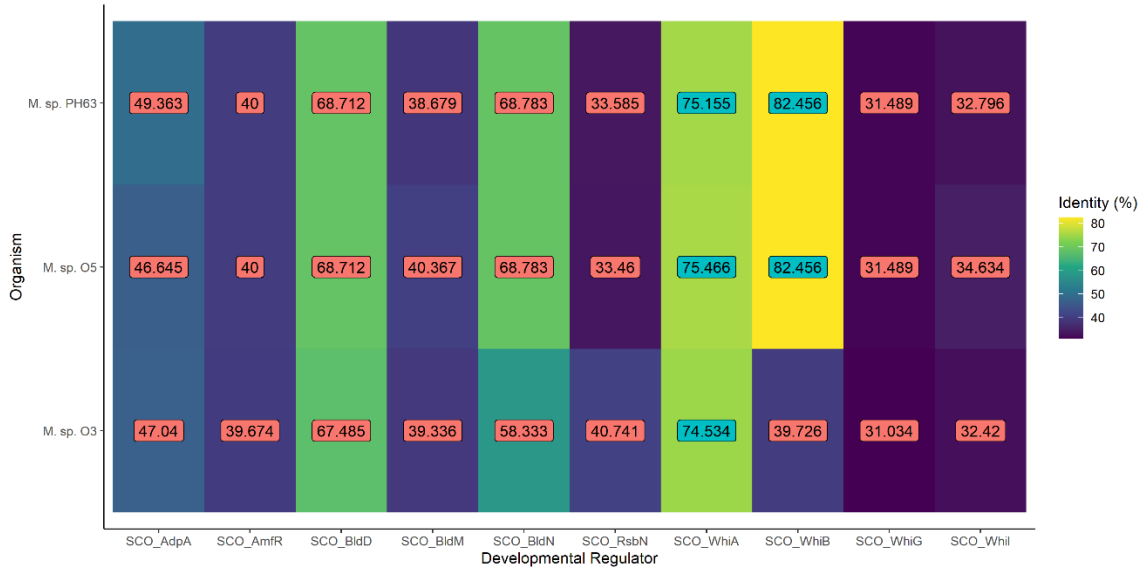


Figure 4-3 Chilean *Micromonospora* isolates lack several developmental regulators present in *S. coelicolor*.

Heatmap of best BLASTp hits (by e-value) of *S. coelicolor* developmental regulators against the proteomes of our organisms, coloured by % amino acid identity. Labels are % amino acid identity values, those with blue backing indicate proteins which had BLAST RBH results. *Micromonospora* sp. O5 and *Micromonospora* sp. PH63 encode WhiA and WhiB regulators, whilst *Micromonospora* sp. O3 encodes only WhiA.

Figure 4-4 ClustalW alignment of Chilean *Micromonospora* putative regulatory proteins to *S. coelicolor* developmental regulators.

Alignment generated using ClustalW and visualised with ggmsa. High resolution alignments are printed in Appendix 1

4.2.4. *Micromonospora* sp. O3, *Micromonospora* sp. O5, and *Micromonospora* sp. PH63 are predicted to utilise BldA mediated translational regulation

After identifying conserved regulatory elements amongst the isolates, we chose to investigate developmental regulation of our organisms by means of codon usage in their genomes. We were particularly interested in the TTA codon, read by developmentally regulated leucyl-tRNA encoded by *bldA*. In all three organisms, TTA was the rarest codon in coding sequences (Fig. 4.5.). *Micromonospora* sp. PH63 contained 145 TTA-containing genes (17 of which were present in BGCs), *Micromonospora* sp. O5 contained 191 (29 in BGCs), and *Micromonospora* sp. O3 contained 224 (nine in BGCs) (Fig. 4.6.).

We followed this by identifying a leucyl-tRNA(UUA) gene in each of our organisms, followed by predicting the folding of the tRNA encoded by that gene and aligning the three genes. We compared these alignments and predicted secondary structures to the BldA of *S. coelicolor* A3(2). The secondary structure prediction suggests that the BldA homologs in the *Micromonospora* isolates does not fold correctly (Fig. 4.7.). Alignment of the three *bldA* genes from the Chilean *Micromonospora* to *S. coelicolor* indicated a four bp deletion in the *Micromonospora* genes which may be responsible for this (Fig. 4.7.) Despite this, the UUA anticodon is still able to form correctly, so the BldA of these genes may retain some function.

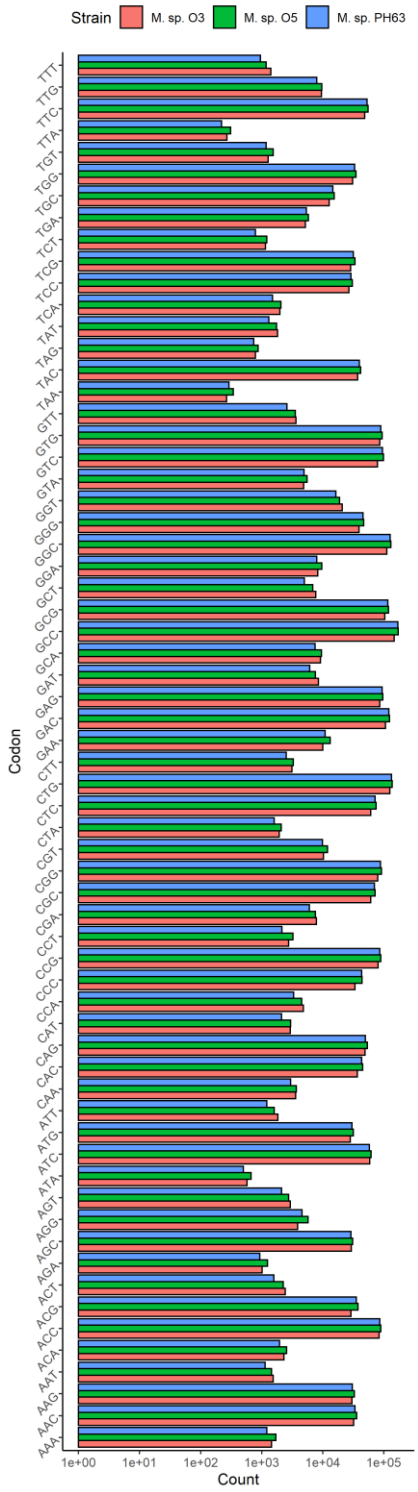


Figure 4-5 Codon Usage in Micromonospora isolates.

Total codon usage in predicted coding sequences, as predicted by Cordon. The leucyl codon, TTA, which is indicative of developmental regulation, was the rarest in our organisms.

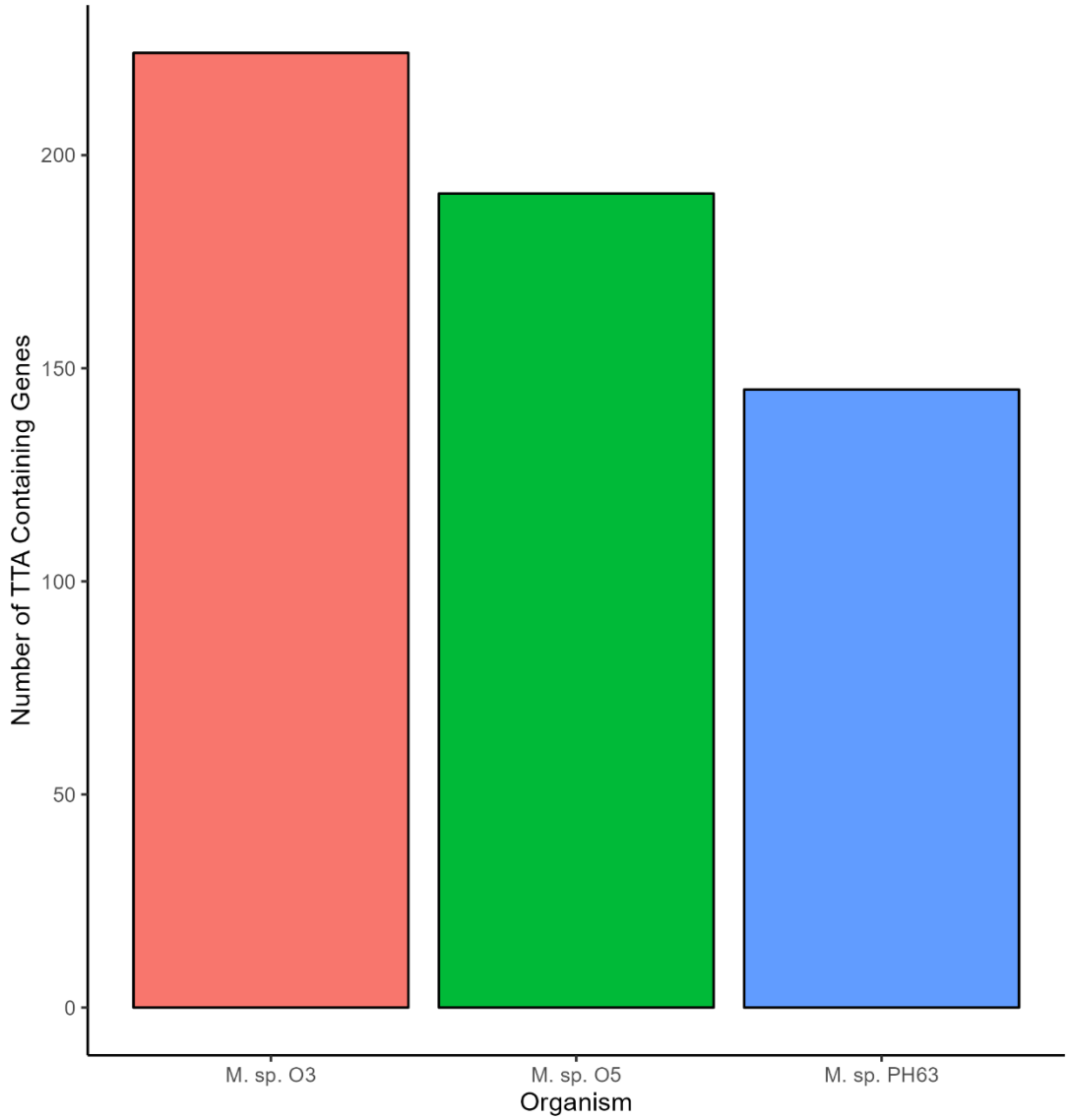


Figure 4-6 Number of TTA Containing Genes in Chilean *Micromonospora* isolates.

All three strains employed TTA encoding

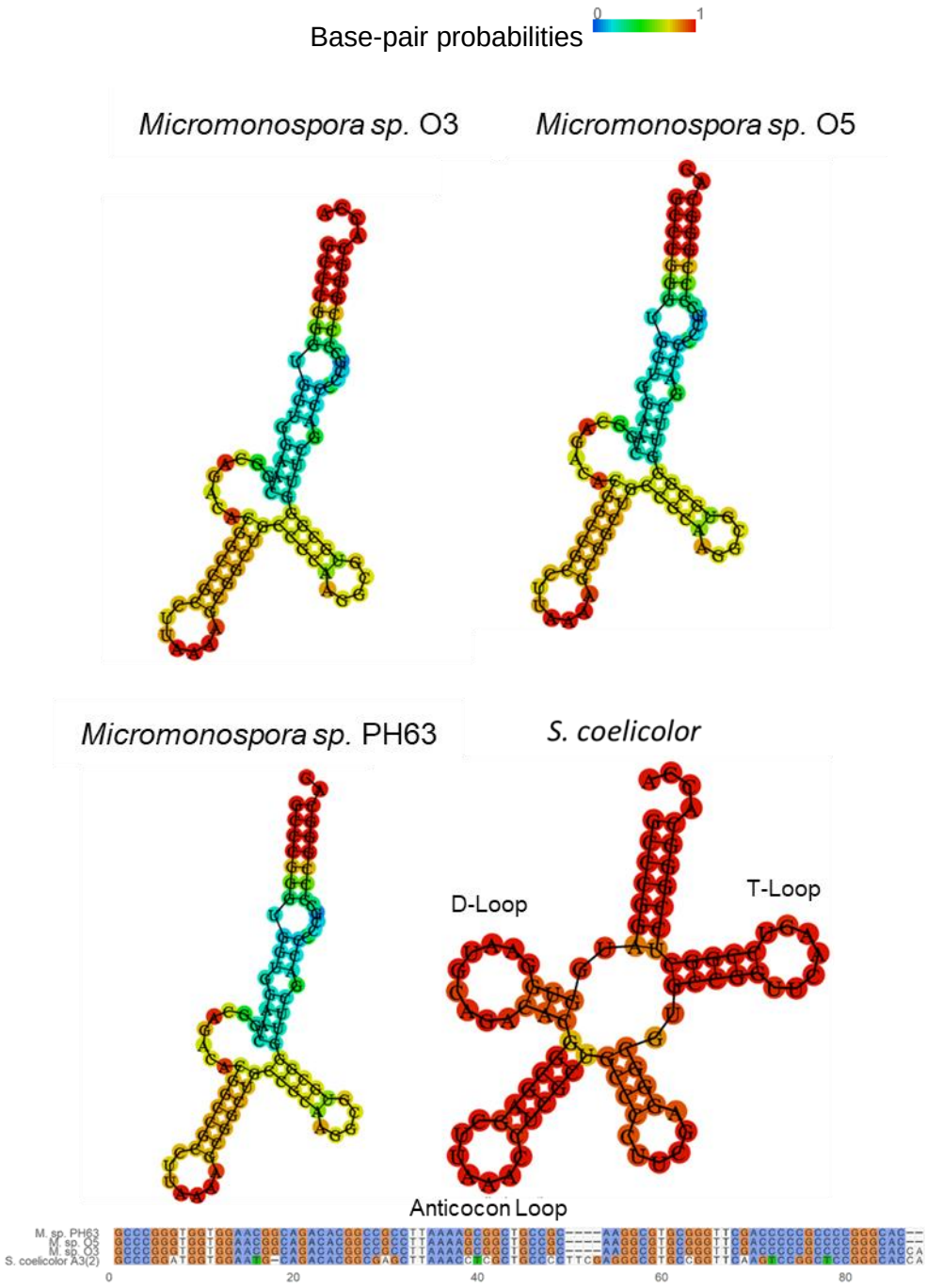


Figure 4-7 BldA in the Chilean *Micromonospora* is predicted to be non-functional.

RNAFold predictions of BldA secondary structures in our organisms versus that of the BldA of *S. coelicolor* A3 (2) indicated the D-loop and T-loop of the BldA of our strains of *Micromonospora* does not fold correctly. A clustalW alignment of the *bldA* genes indicates that this is because of truncation of the *Micromonospora bldA.*

4.2.5. *Micromonospora* sp. O3, *Micromonospora* sp. O5, and *Micromonospora* sp. PH63 chromosomes contain a diverse array of specialised metabolites

To assess the biosynthetic potential of our organisms, and gain insight into what may be driving the antibiotic activity against *P. aeruginosa,* we submitted their assemblies to antiSMASH for annotation. This predicted a total of 50 Biosynthetic Gene Clusters between our three organisms – nine for *Micromonospora* sp. O3, 21 for *Micromonospora* sp. O5, and 19 for *Micromonospora* sp. PH63. Of these 50, 18 were types present only in one organism (Fig. 4.8.). The most common class of BGC was terpenes, with each of our organisms possessing four of these clusters. All three organisms possessed a single T3PKS type cluster. All three organisms employ BldA-mediated translational regulation of some BGCs, and the majority of their clusters did not generate informative ClusterBLAST hits (i.e. hits with a similarity similarity score >50%, meaning more than 50% of the genes in the putative BGC generated significant hits) – this suggests that their secondary metabolism is largely uncharacterised. The BGC content of each of the three organisms is discussed below. *Micromonospora* sp. O3 was notable as encoding relatively few BGCs despite being a *Micromonospora*.

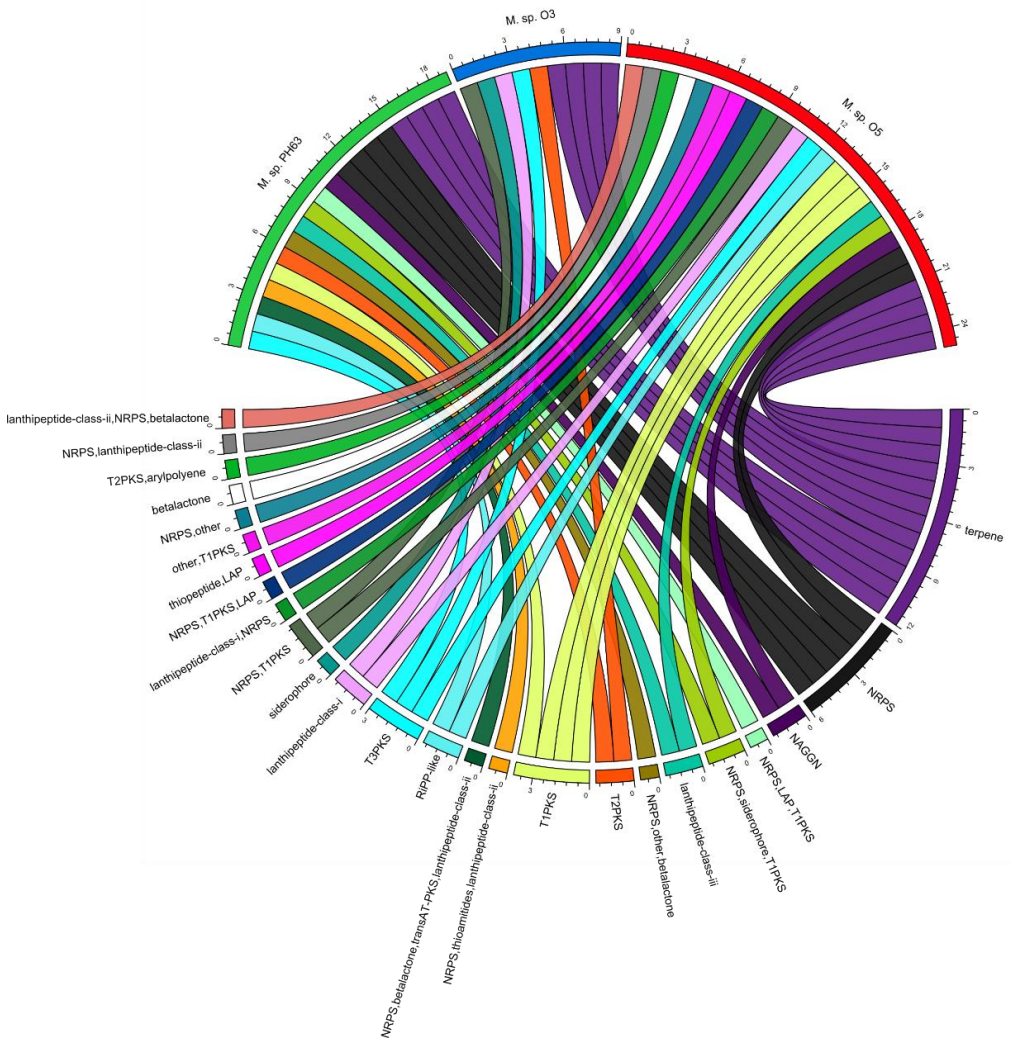


Figure 4-8 Chilean *Micromonospora* isolates encode a diverse array of specialised metabolites.

Chord diagram depicting counts and classes of BGCs present in our Micromonospora. The most common class were terpenes, with each strain possessing four of these.

4.2.6. BGC content of *Micromonospora* sp. O3

Micromonospora sp. O3 contained the fewest predicted BGCs of our organisms, nine. It has four terpene BGCs, two of which had no hits on ClusterBLAST (Table 4.1.). The only ClusterBLAST hits to known BGCs with >50% similarity were the T3PKS cluster, o3-9 and siderophore cluster, o3-6 which aligned to the alkyl-o-dihydrogeranyl-methoxyhydroquinone and desferrioxamine B BGCs, respectively. This suggests the majority of *Micromonospora* sp. O3's BGCs represent uncharacterised metabolites. four of the clusters contained TTA codons, suggesting they are at least partially under control of BldA.

Table 4.1 BGCs of *Micromonospora* sp. *O3*

Region	Type	From	To	Most similar known cluster		Similarity
o3-1	terpene	218,217	238,666		NA	
o3-2	T3PKS	1,677,104	1,718,153	alkyl-O-dihydrogeranyl-methoxyhydroquinones	Terpene + Polyketide	71%
o3-3	terpene	2,244,564	2,263,685	Meridamycin	NRP + Polyketide	10%
o3-4	terpene	2,333,050	2,353,361	Phosphonoglycans	Saccharide	3%
o3-5	lanthipeptide-class-i	4,529,845	4,554,458		NA	
o3-6	siderophore	5,157,746	5,169,611	desferrioxamin B	Other	100%
o3-7	terpene	5,814,773	5,836,101		NA	
o3-8	NRPS,T1PKS	5,902,230	5,999,236	microsclerodermin M	NRP + Polyketide:Modular type I	27%
o3-9	T2PKS	6,506,169	6,544,888	frankiamicin	Polyketide	28%

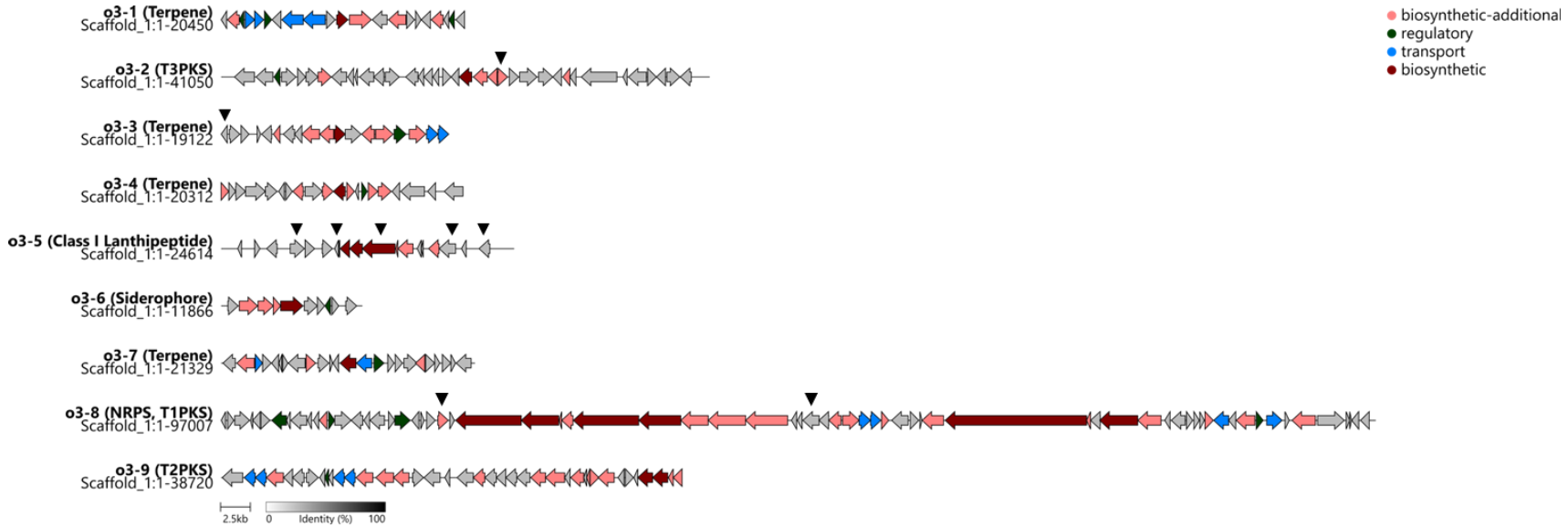


Figure 4-9 Map of the BGCs of *Micromonospora* sp. *O3*.

Illustration generated using clinker and coloured by antiSMASH predicted biosynthetic functions.

4.2.7. BGC Content of *Micromonospora* sp. PH63

Micromonospora sp. PH63 contained 19 BGCs, the most common of which were terpenes at four clusters. The N-acetylglutaminylglutamine amide cluster, ph63-4 and terpene clusters ph63-9 and ph63-17 were the only clusters to not have any known ClusterBLAST hits (Table 4.2.). Only three clusters generated ClusterBLAST hits with similarity scores >50%: ph63-8, which encodes the SapB lanthipeptide; ph63-13, a region encoding for NRPS, thioamitide, and Class II lanthipeptide biosynthesis; and ph63-18 which contains the T3PKS responsible for biosynthesising alkyl-o-dihydrogeranyl-methoxyhydroquinones. The BGCs of *Micromonospora* sp. PH63 were rich in TTA codons, with nine of the 19 BGCs containing at least one (Fig. 4.10.).

Table 4.2 BGCs of *Micromonospora* sp. *PH63*

Region	Type	From	To	Most similar known cluster		Similarity
ph63-1	terpene	228,927	248,871	meridamycin	NRP + Polyketide	13%
ph63-2	terpene	334,532	352,391	phosphonoglycans	Saccharide	3%
ph63-3	NRPS	684,574	726,946	azicemicin B	Polyketide	13%
ph63-4	NAGGN	1,187,316	1,202,032		NA	
ph63-5	NRPS,LAP,T1PKS	1,308,750	1,425,550	sipanmycin	Polyketide	22%
ph63-6	NRPS,siderophore,T1PKS	2,499,731	2,729,179	chaxamycin A / chaxamycin B / chaxamycin C / chaxamycin D	Polyketide	41%
ph63-7	NRPS	3,080,184	3,123,286	muraymycin C1	NRP + Polyketide	5%
ph63-8	lanthipeptide-class-iii	3,233,139	3,255,739	SapB	RiPP:Lanthipeptide	100%
ph63-9	terpene	3,445,591	3,465,727		NA	
ph63-10	NRPS,other,betalactone	3,526,864	3,603,582	melithiazol A	NRP + Polyketide	10%
ph63-11	T2PKS	3,847,040	3,919,558	paramagnetoquinone 1 / paramagnetoquinone 2	Polyketide	25%
ph63-12	T1PKS	4,063,329	4,109,151	sporolide A / sporolide B	NRP + Polyketide:Enediyne type I	36%
ph63-13	NRPS,thioamitides,lanthipeptide-class-ii	4,174,948	4,348,966	rimosamide	NRP	50%
ph63-14	NRPS,betalactone,transAT-PKS,lanthipeptide-class-ii	4,350,314	4,467,160	oxalomycin B	NRP + Polyketide	9%
ph63-15	NRPS	4,485,633	4,533,110	malacidin A / malacidin B	NRP:Ca+-dependent lipopeptide	5%
ph63-16	RiPP-like	4,736,200	4,747,018	lymphostin / neolymphostinol B / lymphostinol / neolymphostin b	Polyketide + NRP	30%
ph63-17	terpene	4,882,743	4,903,696		NA	
ph63-18	T3PKS	6,404,959	6,446,011	alkyl-O-dihydrogeranyl-methoxyhydroquinones	Terpene + Polyketide	71%
ph63-19	NRPS	6,746,966	6,794,932	catenulisporolides	NRP + Polyketide	3%

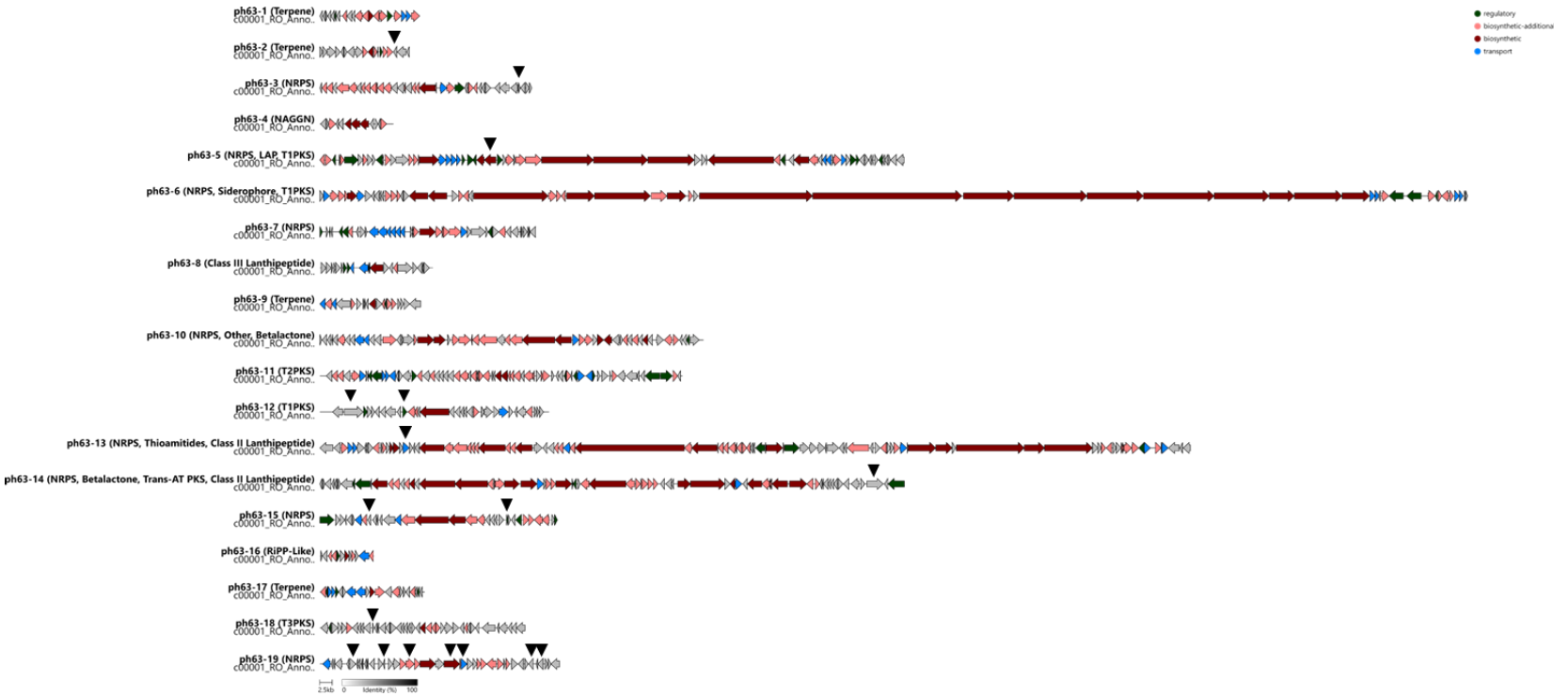


Figure 4-10 Genetic Map of the BGCs of *Micromonospora* sp. *PH63*.

Illustration generated using clinker and coloured by antiSMASH predicted biosynthetic functions.

4.2.8. BGC content of *Micromonospora* sp. O5

Micromonospora sp. O5 posessed the richest suite of BGCs amongst the three Chilean strains, totalling 25. Four of its BGCs generated no ClusterBLAST hits: o5-4, o5-12, o5-22, and o5-25. Only two clusters of the 25 generated similarity scores >50% - o5-11 which encodes SapB biosynthesis, and o5-24 which encodes for the biosynthesis of alkyl-o-dihydrogeranyl methoxyhydroquinones. 12 of *Micromonospora* sp. O5's BGCs contained TTA codons (Fig. 4.11.).

Table 4.3 BGCs of *Micromonospora* sp. *O5*

Region	Type	From	To	Most similar known cluster		Similarity
o5-1	terpene	250,976	271,063	meridamycin	NRP + Polyketide	13%
o5-2	terpene	352,069	372,190	phosphonoglycans	Saccharide	3%
o5-3	lanthipeptide-class-i,NRPS	715,376	804,970	azicemicin B	Polyketide	11%
o5-4	NAGGN	1,075,005	1,089,787		NA	
o5-5	NRPS,T1PKS,LAP	1,474,251	1,590,572	maytansine / ansamitocin P-3	Polyketide	12%
o5-6	T1PKS	1,875,524	1,922,226	carrimycin	Polyketide	6%
o5-7	thiopeptide,LAP	2,784,890	2,812,439	lactazole	RiPP:Thiopeptide	22%
o5-8	NRPS,siderophore,T1PKS	2,860,071	3,104,616	rifamycin	Polyketide	21%
o5-9	other,T1PKS	3,417,201	3,464,953	cyphomycin	Polyketide	6%
o5-10	NRPS,other	3,529,893	3,593,143	himastatin	NRP	12%
o5-11	lanthipeptide-class-iii	3,637,185	3,659,785	SapB	RiPP:Lanthipeptide	100%
o5-12	terpene	3,850,084	3,870,711		NA	
o5-13	betalactone	3,946,149	3,970,813	ketomemicin B3 / ketomemicin B4	Other	33%
o5-14	NRPS,T1PKS	4,041,356	4,093,112	BE-43547A1 / BE-43547A2 / BE-43547B1 / BE-43547B2 / BE-43547B3 / BE-43547C1 / BE-43547C2	NRP:Cyclic depsipeptide + Polyketide:Modular type I	13%
o5-15	T2PKS,arylpolyene	4,235,469	4,307,987	pradimicin-A	Polyketide	25%
o5-16	T1PKS	4,449,701	4,493,209	sporolide A / sporolide B	NRP + Polyketide:Enediyne type I	34%
o5-17	NRPS,lanthipeptide-class-ii	4,570,760	4,640,995	atratumycin	NRP	10%
o5-18	NRPS	4,641,282	4,716,891	rimosamide	NRP	42%
o5-19	lanthipeptide-class-ii,NRPS,betalactone	4,719,074	4,798,086	cyclomarin D	NRP	17%
o5-20	NRPS	4,996,415	5,079,915	malacidin A / malacidin B	NRP:Ca+-dependent lipopeptide	7%
o5-21	RiPP-like	5,237,065	5,247,883	lymphostin / neolymphostinol B / lymphostinol / neolymphostin b	Polyketide + NRP	30%
o5-22	terpene	5,390,801	5,411,754		NA	
o5-23	T1PKS	5,954,485	6,001,945	lydicamycin	NRP + Polyketide:Modular type I	36%
o5-24	T3PKS	6,890,532	6,931,584	alkyl-O-dihydrogeranyl-methoxyhydroquinones	Terpene + Polyketide	71%
o5-25	lanthipeptide-class-i	7,272,240	7,297,412		NA	

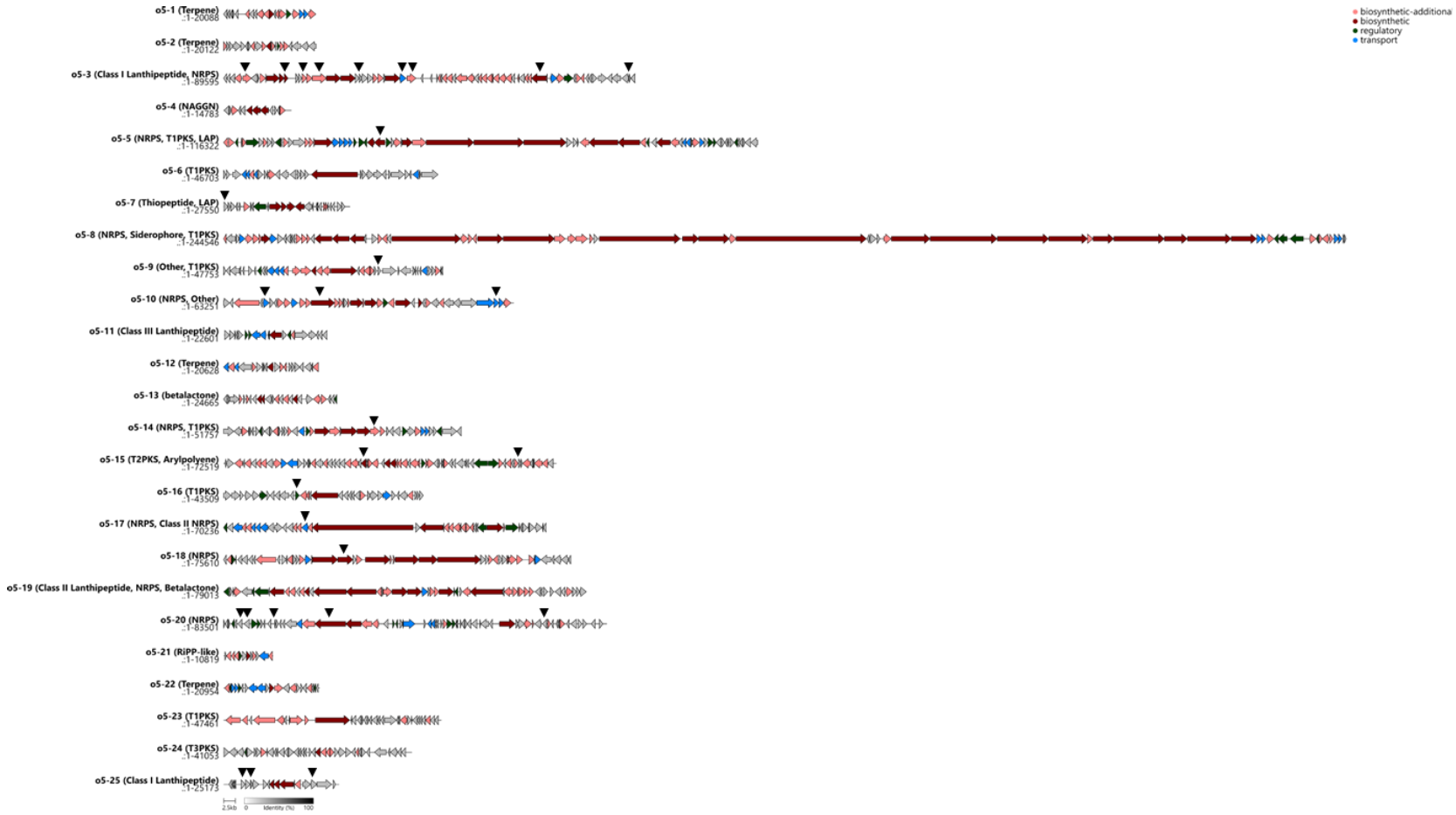


Figure 4-11Genetic Map of the BGCs of *Micromonospora* sp. *O5*.

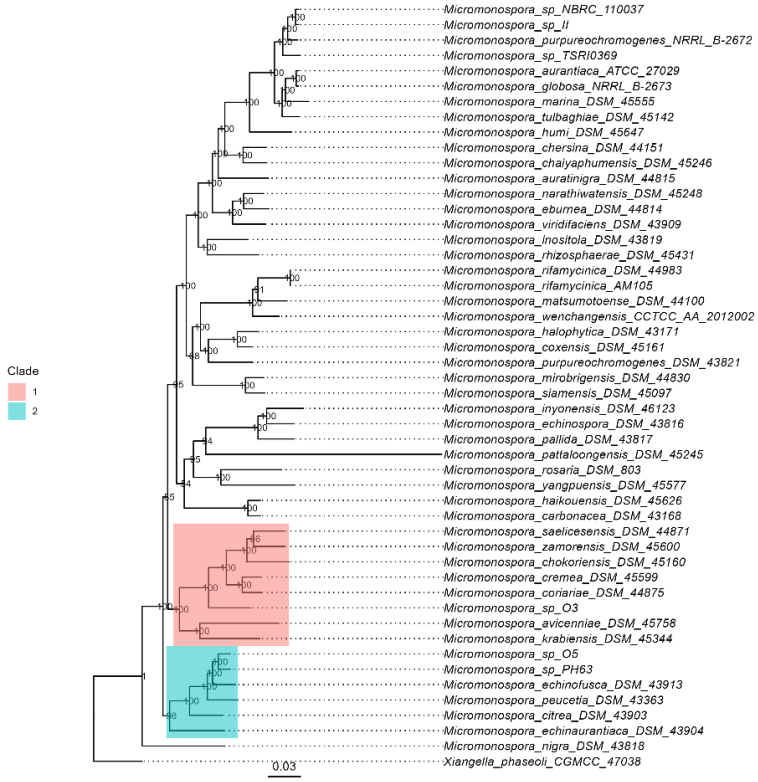
Illustration generated using clinker and coloured by antiSMASH predicted biosynthetic functions.

4.2.9. *Micromonospora* sp. O3, O5, and PH63 are novel species of *Micromonospora*

To enable assembly improvement and confirm their phylogeny we submitted our assemblies to the autoMLST pipeline to place them within the *Micromonospora* and identify closely related strains to them. The tree generated by the core-genome alignment autoMLST uses places our three strains into two distinct monophyletic clades (Fig. 4.9. A). *Micromonospora* sp. O3 was placed into a clade by itself (Clade 1), neighboured by *M. zamorensis* DSM 45600, *M. rosaria* DSM 803, *M. saelicesensis* DSM 44871, *M. cremea* DSM 45599, *M. coriariae* DSM 44875, *M. krabiensis* DSM 45344, and *M. avicenniae* DSM 45758. Clade 2, which contained *Micromonospora* sp. O5 and *Micromonospora* sp. PH63 as well as *M. echinofusca* DSM 43913, *M. peucetia* DSM 43363, *M. citrea* DSM 43903, and *M. echinaurantiaca* DSM 43804.

ANI comparisons indicated that all three of our organisms did not share a species with their nearest neighbours and thus are likely novel (Fig. 4.9. B). *Micromonospora* sp. O5 and *Micromonospora* sp. PH63 are 95.5% identical to each other, and so are also likely distinct species from each other.

A



B

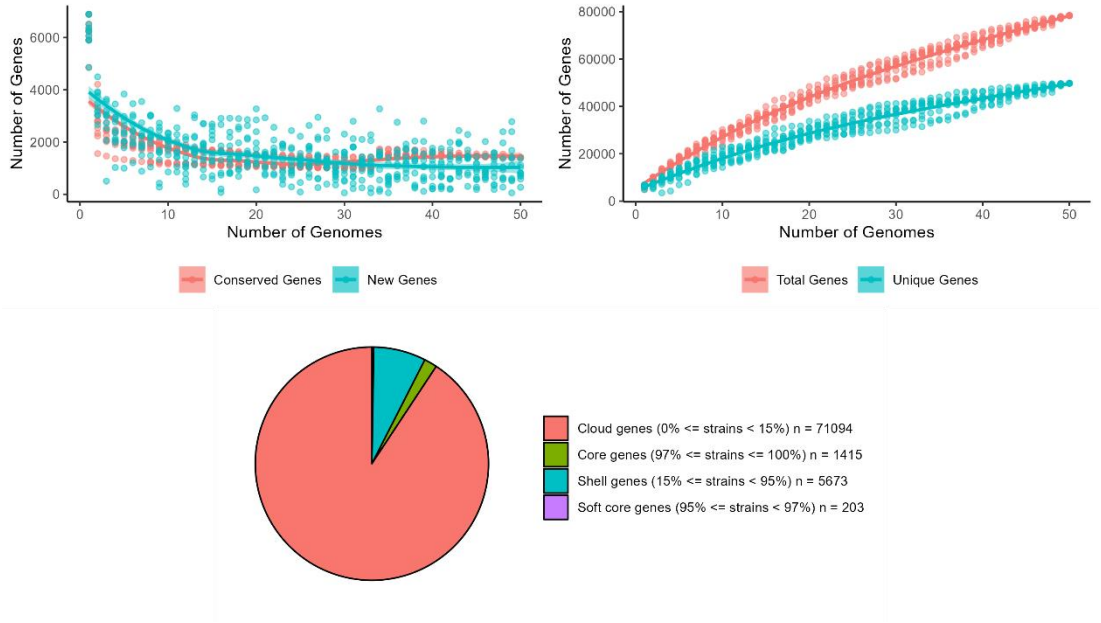
Figure 4-12 The Chilean *Micromonospora* are novel species.

(**A**) ANI values of *Micromonospora* sp. O3, O5, and PH63. With the exemption of ANI calculated against themselves, no members of the tree generated ANI values >= 96%, as indicated by the vertical red line (**B**) Phylogenetic tree generated by AutoMLST, with *Xiangella phaseoli* automatically selected as an outgroup. Scale bar = substitutions per site.

4.2.10. The *Micromonospora* pangenome is open

After using autoMLST to obtain a set of 50 *Micromonospora* related to our strains, we chose to explore the pangenome of these organisms to gain an understanding of their gene content. Using the Roary pipeline, we found that the *Micromonospora* pangenome is open. As we were using a small number of organisms for within-genus comparisons instead of within species comparisons, we set the amino acid identity cutoff for clustering to 80% and the cutoff for the number of organisms a gene must be in to be considered core to 97% [237]. Apart from this, all settings were default. This predicted the *Micromonospora* pangenome to be large, composed of 78,385 clusters (Fig. 4.10. B). The majority of these genes were cloud genes (71,094), present in <15% of the organisms in the tree. This left a shell of 5673 genes, soft-core of 203 genes, and core of 1415 genes (Fig. 4.10. A). As this pangenome was so large, there was a possibility of inappropriate placement of a non-*Micromonospora* into the autoMLST tree. We chose to further investigate this.

A



B

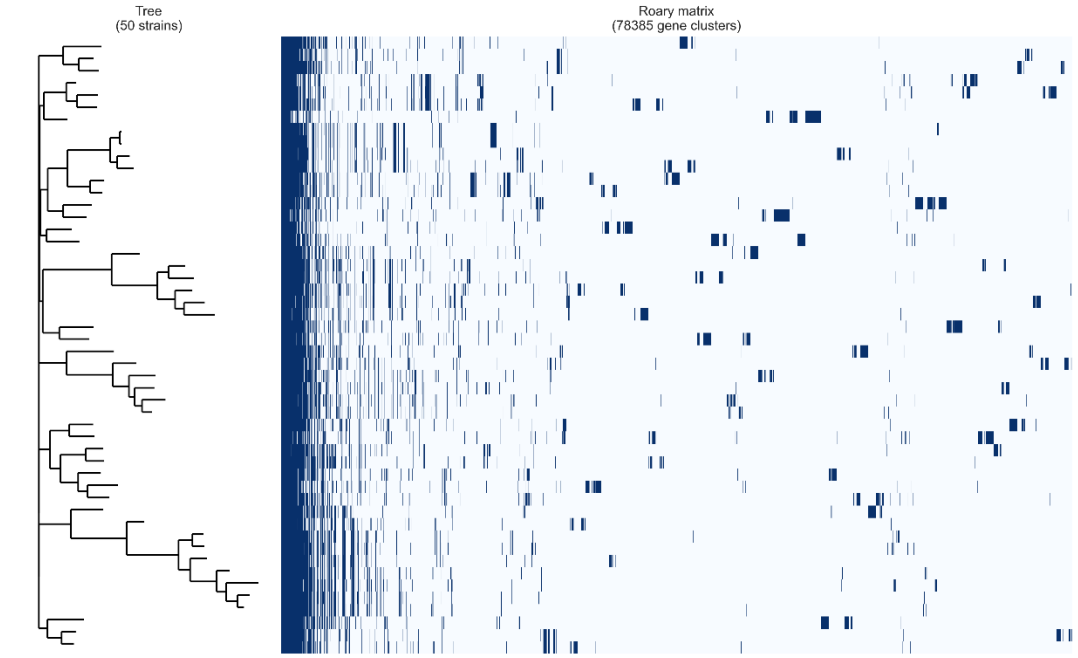


Figure 4-13The *Micromonospora* pangenome is open.

(**A**) Proportion of genes present in the core, soft-core, shell, and cloud *Micromonospora* pangenome. (**B**) Presence-absence table of genes in the *Micromonospora pangenome.*

4.2.11. AutoMLST may incorporate non-*Micromonospora* as part of its pipeline

As the pangenome generated from the strains in the autoMLST tree was remarkably large, we chose to revisit how closely the organisms which populated it are related. To calculate ANI, the FastANI algorithm fragments a query genome and maps that onto the reference which ANI is being calculated against. Calculating the fraction of the fragments which mapped between organisms suggested that *M. pataloongensis* had been misassigned as a *Micromonospora*, (Fig. 4.14.). Despite this, the organisms that fell alongside Clade 1 and Clade 2 still clustered closely together through heatmap2's default clustering (Fig. 4.15.), and so we maintained that *Micromonospora* sp. O3, *Micromonospora* sp. O5, and *Micromonospora* sp. PH63 are still novel species of *Micromonospora* and closely related to the organisms within their clades.

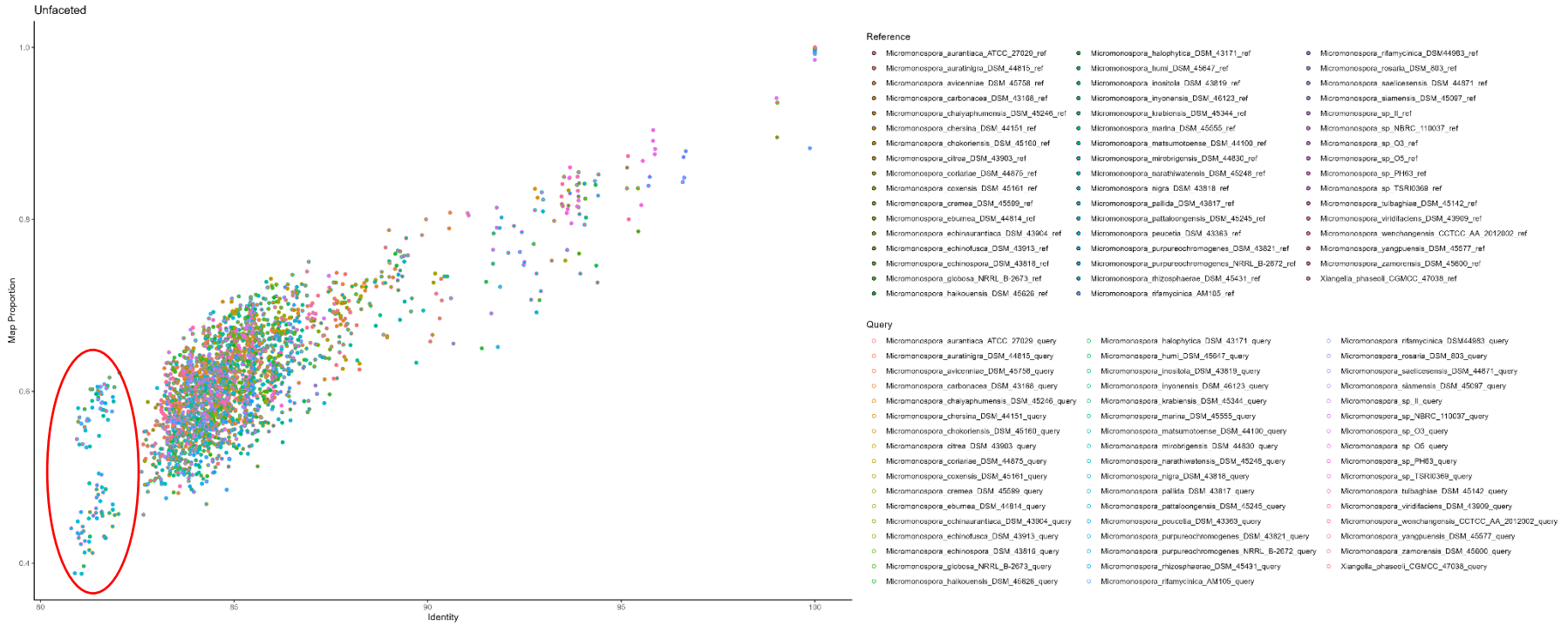


Figure 4-14 *Micromonospora pataloongensis* DSM 45245 may be misassigned *Micromonospora*.

Scatterplot of FastANI Identity against proportion of fragments mapped between query and reference genomes. The two highlighted clusters separate from the central mass of points are populated by estimates where *M. pataloongensis* is either the reference or query sequence.

A

B

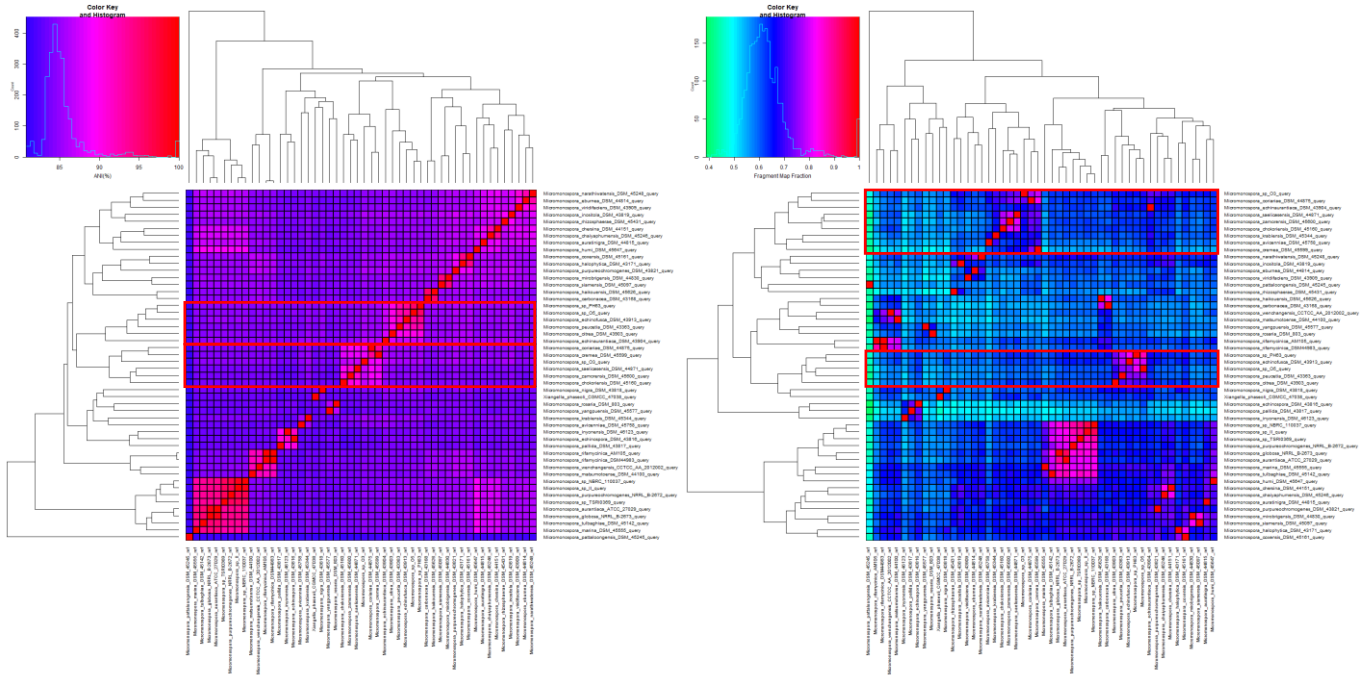


Figure 4-15 FastANI analysis cooroborates AutoMLST analysis on the placement of Clades 1 and 2 within the genus *Micromonospora*.

Heatmaps of ANI (**A**) and proportion of mapped fragments (**B**) for members of the tree generated by AutoMLST, clustered by heatmap2's default clustering. Clade 1 and 2 still emerge from this data, however the positions of the clades do not agree with the autoMLST phylogeny.

4.2.12. *M. pataloongensis*' inclusion inflated the predicted *Micromonospora* pangenome

With *M. pataloongensis* identified as unlikely to be a *Micromonospora*, we recalculated the pangenome of the members of the autoMLST phylogeny after its removal. In addition, we compared the effect of splitting paralogous genes (the default option in Roary) in the pangenome versus clustering them together.

As expected, the smallest genome with the largest core was that generated after excluding *M. pataloongensis* and clustering paralogs (Fig. 4.16). Whether paralogs were clustered together or not, *M. pataloongensis* contributed disproportionately to the cloud genes of the pangenomes – without clustering, there were 71,094 cloud genes with *M pataloongensis* and only 68,372 when it was excluded.

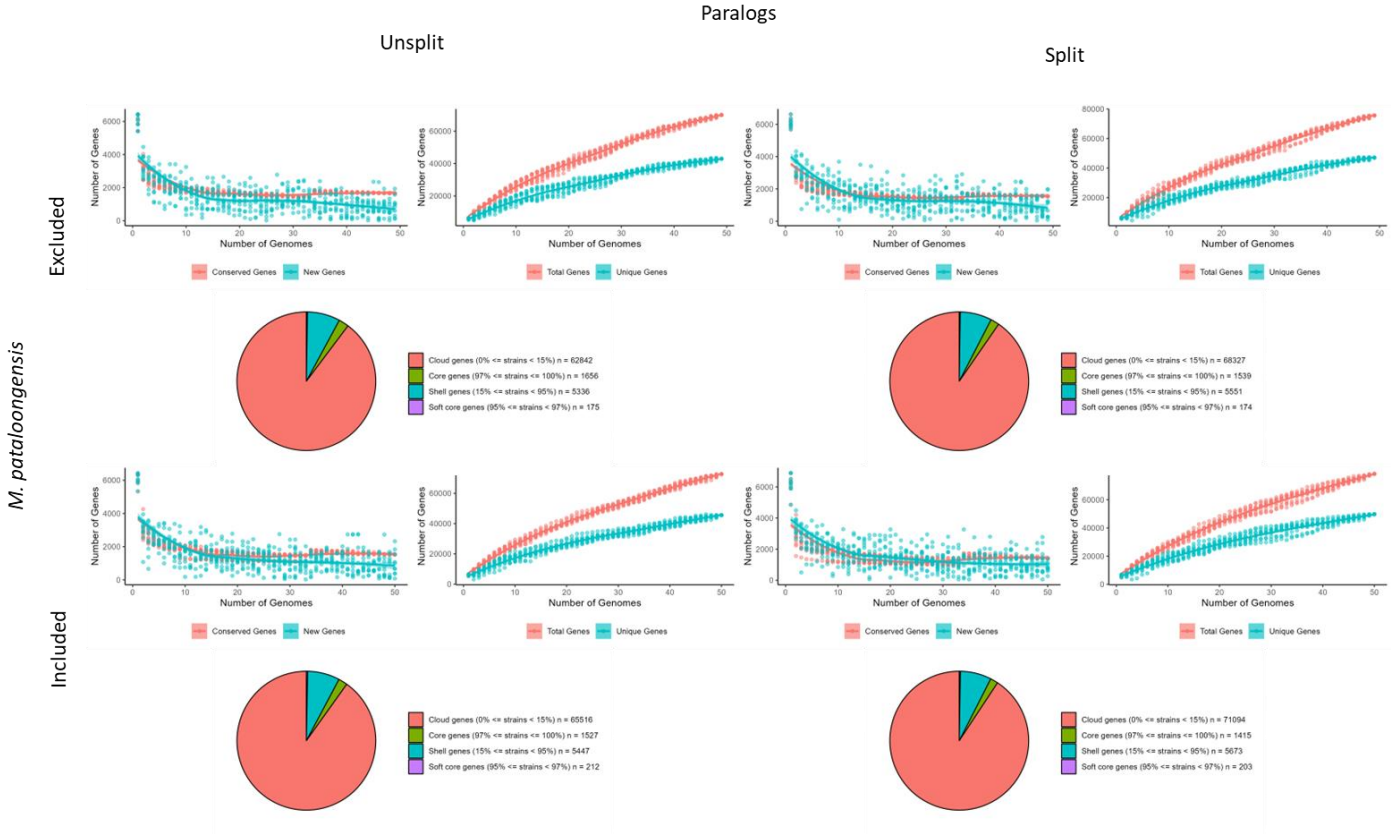


Figure 4-16 A. Effect of paralog splitting and M. pataloongensis inclusion on pangenome prediction.

Exclusion of *M. pataloongensis* resulted in the largest set of core genes and reduced the cloud gene group by ~3000-~4000 genes (unsplit and split paralogs, respectively).

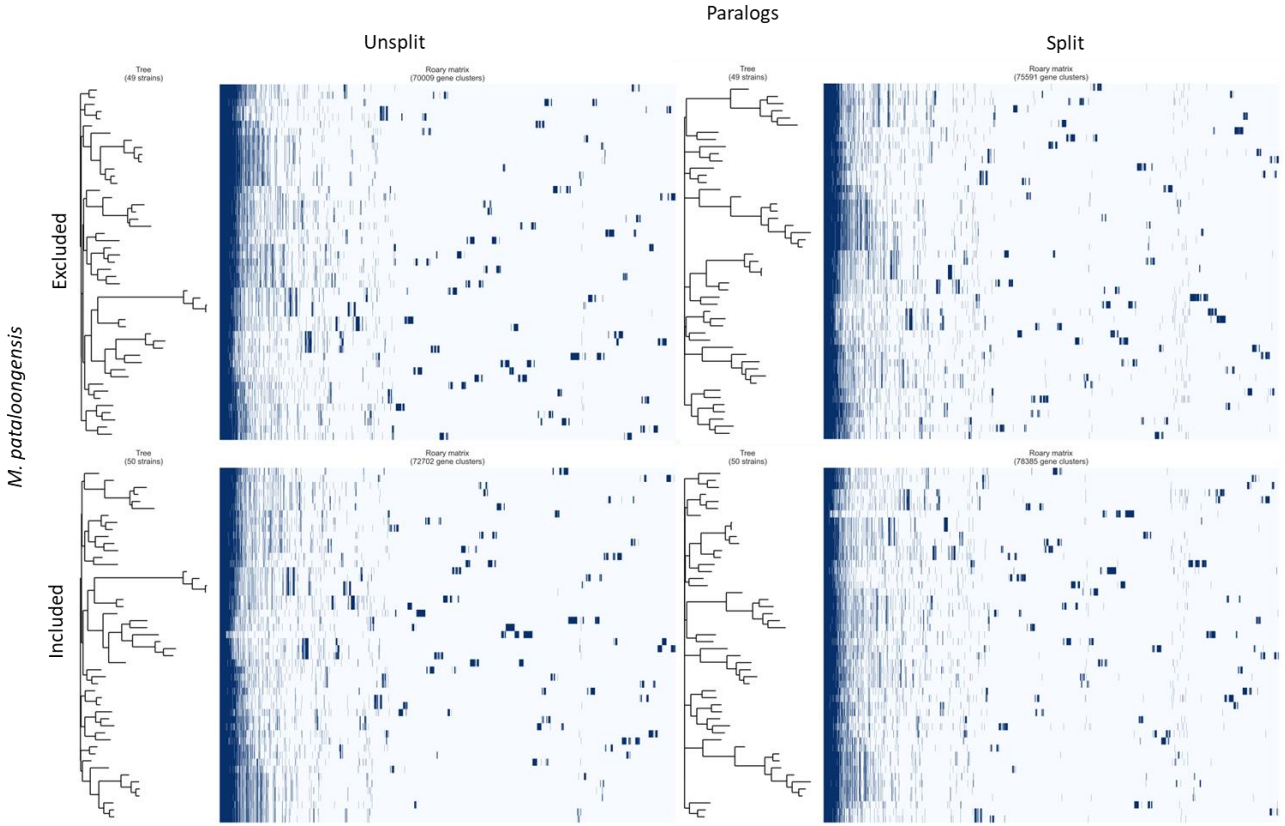


Figure 4.16. B Effect of paralog splitting and M. pataloongensis inclusion on pangenome prediction.

Presence absence heatmaps of pangenomes. Excluding M. pataloongensis and not splitting paralogous genes resulted in the fewest clusters and largest core genome.

4.2.13. Closely related species of *Micromonospora* sp. O3, *Micromonospora* sp. O5, and *Micromonospora* sp. PH63 are globally distributed

As our *Micromonospora* strains were isolated from hyper-arid soil, we hypothesised that the phylogenetically related strains would have been isolated from similar habitats. To assess this, we reviewed the literature surrounding these organisms that identified their isolation conditions. Despite *Micromonospora* sp. O5 and *Micromonospora* sp. PH63's origins, the rest of Clade 2 were isolated from a diverse suite of environments. *M. citrea* was isolated from lake mud; *M. echinofusca* from chukar (*Alectoris chukar*) faeces, and *M. peucetia* and *M. echinaurantiaca* from unspecified soil [238]. Clade 1 was notable for being populated by endophytic or rhizospheric *Micromonospora*. *M. cremea* and *zamorensis* were isolated from the rhizosphere of *Pisum sativum* [228], and *M. saelicesensis* likewise from *Lupinus angustifolius* [62]. *M coriariae* is a root-dwelling *Micromonospora* isolated from a root nodule of a coriaria plant. Whilst *M. cremea, zamorensis, coriariae,* and *saelicesensis* were isolated in Spain, *M. avicenniae* was isolated from roots of the mangrove *Avicennia marina*. *M. rosaria* and *M. krabiensis* were the only members of Clade 2 not associated with plants – the former was isolated from Texan soil, whist the latter is a marine Thai strain. The geographic diversity of these organisms suggests that *Micromonospora* are widely globally distributed, and that their phylogeny does not necessarily follow their geography.

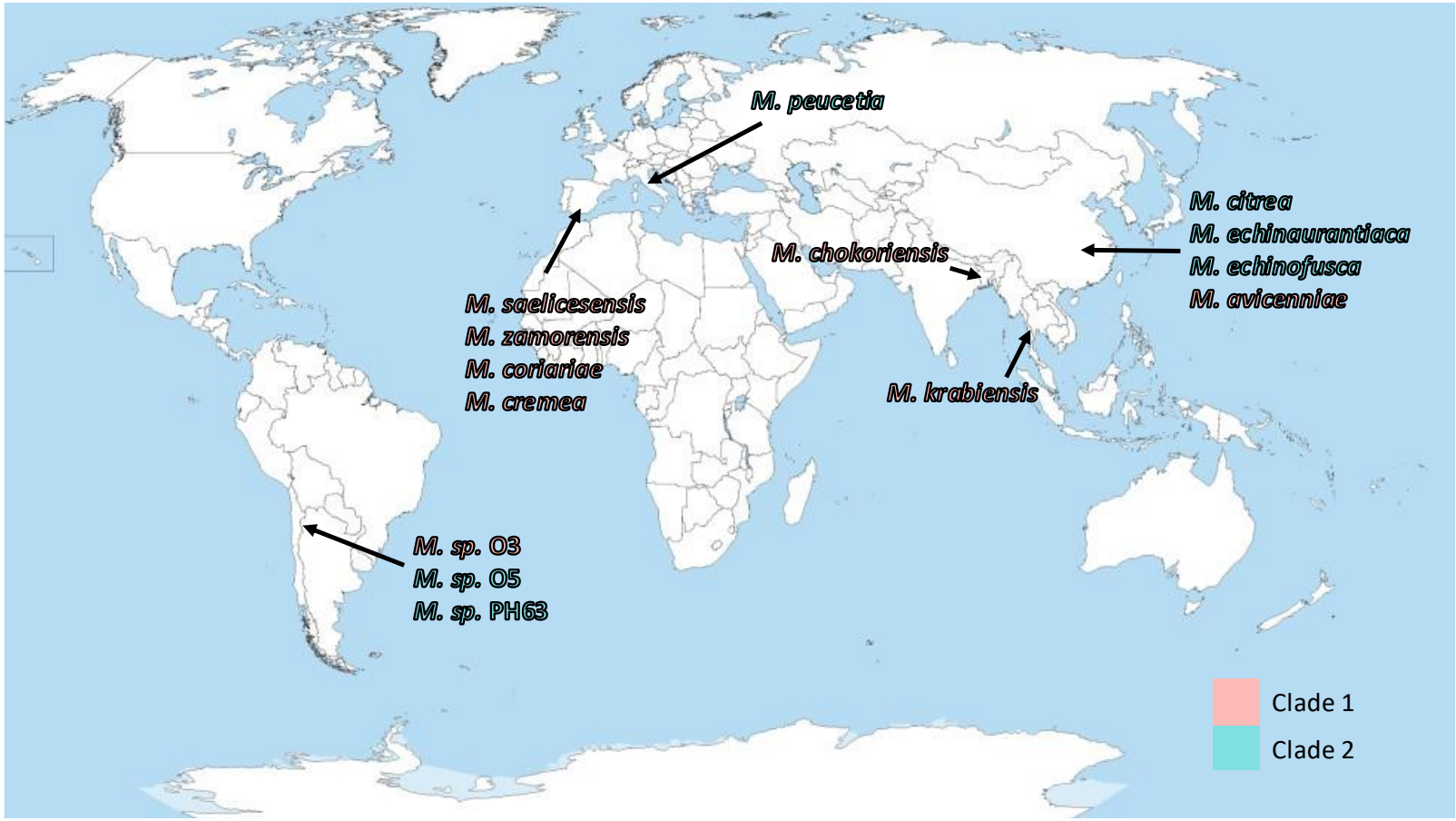


Figure 4-17 The phylogenetic neighbours of *Micromonospora* sp. *O3, O5*, and *PH63* are globally distributed

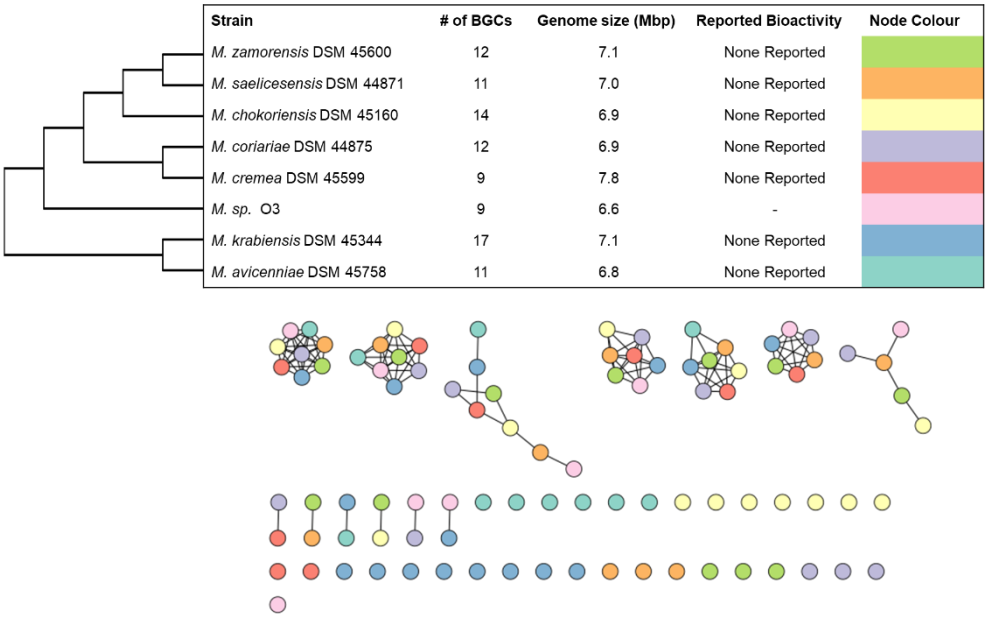
4.2.14. *Micromonospora* sp. O3, *Micromonospora* sp. O5, and *Micromonospora* sp. PH63 share BGCs with their phylogenetic neighbours

As the neighbours of our *Micromonospora* were broadly distributed geographically we next sought out to examine the BGCs present in our organisms. We analysed the total BGC content of each clade, as identified by antiSMASH, and then clustered using BiG-SCAPE. In both clades, all strains had at least one singleton cluster.

Clade 1 (4.18. A) contained three eight-membered networks, with a single representative from each strain. It also contained six BGC pairs, which largely followed the ancestry of their host organisms. The exceptions to this were a pair of clusters from *Micromonospora* sp. O3 and *M. krabiensis*, and a pair from *M. zamorensis* and *M. chokoriensis*, without a member from *M. saelicesensis. Micromonospora* sp. O3 contained one singleton node, corresponding to the hybrid NRPS,T1PKS cluster o3-8.

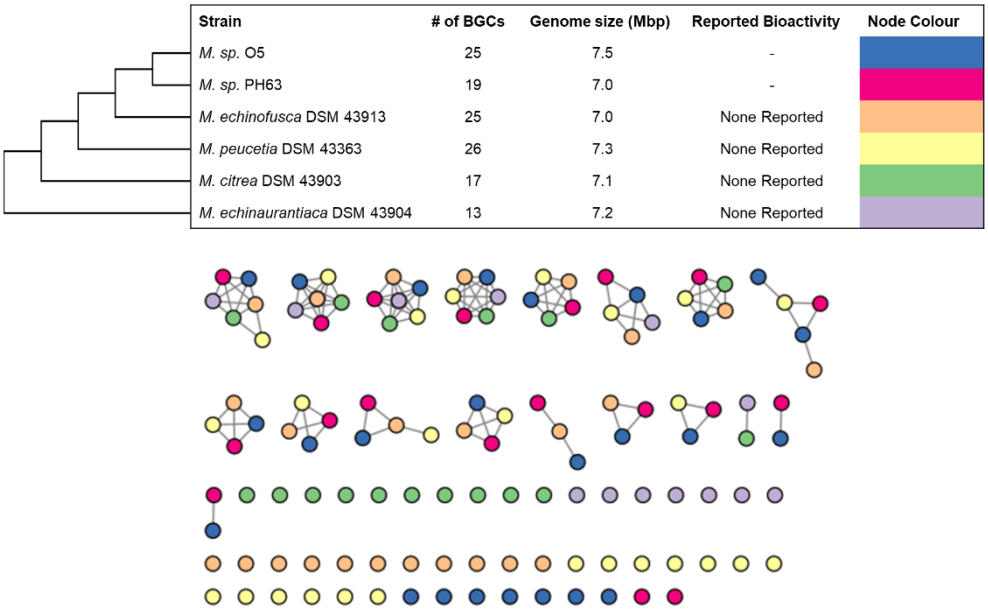
Clade 2 (4.18. B) had four six-membered networks, populated by one BGC from each organism in the clade, whilst Clade 2 had two eight-membered networks representing the same. Despite being closely related by ANI, *Micromonospora* sp. O5 and PH63 had nine singleton nodes between them (seven belonging to *Micromonospora* sp. O5 and 2 to *Micromonospora* sp. PH63 respectively). In *Micromonospora* sp. O5 these correspond to BGCs o5-6, o5-7, o5-9, o5-10, o5-14, o5-23, and o5-25. In *Micromonospora* sp. PH63 these corresponded to ph63-7 and ph63-19. Like in clade 1, clustering of the BGCs corresponded to phylogenetic closeness – such as the node pairs between *Micromonospora* sp. O5 and *Micromonospora* sp. PH63, and between *M. citrea* and *M. echinaurantiaca.*

A



Strain	# of BGCs	Genome size (Mbp)	Reported Bioactivity	Node Colour
M. zamorensis DSM 45600	12	7.1	None Reported	
M. saelicesensis DSM 44871	11	7.0	None Reported	
M. chokoriensis DSM 45160	14	6.9	None Reported	
M. coriariae DSM 44875	12	6.9	None Reported	
M. cremea DSM 45599	9	7.8	None Reported	
M. sp. O3	9	6.6	-	
M. krabiensis DSM 45344	17	7.1	None Reported	
M. avicenniae DSM 45758	11	6.8	None Reported	

B



Strain	# of BGCs	Genome size (Mbp)	Reported Bioactivity	Node Colour
M. sp. O5	25	7.5	-	
M. sp. PH63	19	7.0	-	
M. echinofusca DSM 43913	25	7.0	None Reported	
M. peucetia DSM 43363	26	7.3	None Reported	
M. citrea DSM 43903	17	7.1	None Reported	
M. echinaurantiaca DSM 43904	13	7.2	None Reported	

Figure 4-18 Closely related *Micromonospora* contain shared and unique BGCs.

BiG-SCAPE networks generated from the BGCs of Clade 1 (A) and Clade 2 (B). All three Chilean strains contained at least one unique BGC in their clade.

4.2.15. *Micromonospora sp.* O3, O5, and PH63 encode numerous UV resistance genes

As all three strains produced UV-absorbing pigments, and owing to the high ambient UV exposure of the Atacama Desert, we checked for the presence of genes associated with UV radiation resistance – orthologs of the *uvr* genes present in *E. coli*

4.3. Discussion

4.3.1. *Micromonospora* sp. O3, *Micromonospora* sp. O3, and *Micromonospora* sp. PH63 bear the hallmarks of actinomycete genomes

The first step of genome mining is generating a high-quality assembly of the genome to be mined. *Micromonospora* sp. PH63 and *Micromonospora* sp. O5 benefitted from the incorporation of long-read sequencing in the assembly process, generating fewer contigs during assembly pre scaffolding compared to *Micromonospora* sp. O3 [167]. All three assemblies are large and high GC (>70%) [49, 52], as well as biosynthetically rich (discussed in 4.12.). Recapitulating the chromosomes in as few contigs as possible was critical to genome mining, as the repetitive sequence present in many core biosynthetic genes is difficult to reconstitute and risks fragmentation of BGCs during assembly [93].

4.3.2. Differing strategies for developmental regulation

As part of the work presented here, we have screened our organisms' genomes for conserved regulators of development (i.e. the transition from vegetative hyphal growth to sporulation initiation and completion) between our organisms and *S. coelicolor*. *S. coelicolor* was selected owing to its place as the model actinomycete, and the subsequent level of detail in which its development has been studied [72]. Expectedly, there was only partial conservation between *S. coelicolor* and the three Chilean strains. All three organisms encode a BldD ortholog. BldD acts as a master repressor of development, mediated by c-di-GMP binding [77]. In *Actinoplanes*, it has been shown to play a role in proper sporangium development [81]. All organisms also had WhiA orthologs – this regulator is implicated in completion of spore development and proper chromosome segregation [206] and so possibly shares a similar role in the Chilean strains.

What stood out is that not all three strains had the same conserved regulators versus *S. coelicolor*. *Micromonospora* sp. O3 lacked WhiB (although the best BLASTp hit was predicted to be a WhiB type regulator) and the sigma factor BldN. This suggests diversity of development within the genus. This is not surprising, with the most famous example of this being *S. coelicolor's* inability to complete its life cycle in liquid culture compared to *S. venezuelae's* ability to do so [71].

All three strains lacked AdpA, AmfR, BldM, RsbN – implicated in aerial hyphae development – and WhiG, and WhiI – which are implicated in spore maturation. None

of the Chilean strains were predicted by antiSMASH to biosynthesise gamma-butyrolactone molecules, and so the lack of AdpA suggests they do not employ A-factor mediated developmental regulation [135]. AmfR, BldM, and RsbN are responsible for developing aerial hyphae in *Streptomyces* – a phenotype that *Micromonospora* lack [49] and so their absence is also not surprising. On the opposite end of development, they also lacked regulators involved with spore maturation [209]. Morphological development is regulatorily intertwined with specialised metabolism in actinomycetes, and so understanding one helps us understand the other.

What stood out was how heavily the Chilean strains are predicted to rely on translational regulation, through the action of BldA – a leucyl tRNA whose accumulation is developmentally regulated [239] and thus restricts cellular translation of transcripts bearing BldA's cognate codon . This is reflected in the rarity of the TTA codon in all three strains – present in 2.4-3.7% of genes. Despite this, the BldA of the three Chilean strains was predicted to be non-functional. Their *bldA* genes carry a four bp deletion (ΔT50-G54) compared to *S. coelicolor*. Although it maintains the anticodon, it lacks the D-arm required for charging of the tRNA with leucine [240]. Heterologous introduction of foreign tRNAs into actinomycetes for improved antibiotic production is a developing field [241, 242] and it stands to reason that the Chilean strains could benefit from such experimentation.

4.3.3. Genome mining and BGC prediction

Molecule rediscovery plagues the search for novel antimicrobials [243], reducing the yield of an already low-throughput an labour intensive process. The ease of access of contemporary genome sequencing helps to alleviate this, by allowing for predictions of interesting organisms' biosynthetic capacity [93, 215, 244]. One of the most popular and supported tools for this endeavour, antiSMASH [181], is the tool we selected to analyse the Chilean *Micromonospora*. In addition to being well maintained over the decade since its first release [179], antiSMASH is integrated into and supported by various other bioinformatic tools [182, 183, 187] and thus was well suited for this project.

4.3.4. Searching for biosynthetic novelty

All three of the *Micromonospora* strains we sequenced were rich in BGCs. That terpenes were common is unsurprising, orthologous clusters in *Salinispora* contribute

to proper pigmentation [227]. Likewise, the conserved T3PKS is orthologous to the alkyl-o-dihydrogeranyl methoxyhydroquinones of *Actinoplanes* [245].

Outwith the organisms' terpenes and T3PKS clusters, the antiSMASH results suggest they possess uncharacterised diversity within their secondary metabolism. The antiSMASH ClusterBLAST results compare predicted BGCs to the BGCs present in the MIBiG database [181] and return the best hit from it. Multiple BGCs from the Chilean strains returned no hits from this database, and many of those that did are likely false-positives (egregious examples include o5-3, encoding class I lanthipeptide and NRPS biosynthetic machinery, and clusters o3-2, o5-2, and ph63-2 which were all called with 3% similarity to the saccharide-class phosphonoglycan BGC). The similarity values need to be considered carefully – the values are calculated as the number of genes in a query cluster that had hits to a reference cluster, as a percentage of the number of genes in that cluster [179].

As biological diversity and chemical diversity go hand-in-hand [246, 247], it was important to establish if the Chilean isolates were novel species of *Micromonospora* or not. we submitted them to autoMLST for confirmation that they are, indeed, *Micromonospora*, and to identify related species that we can compare them to. AutoMLST has the advantage of selecting organisms that are likely to be related to a query and selecting conserved single-copy-orthologs for alignment and tree building [178] and so provided information as to which strains are close relatives of the Chilean isolates. FastANI suggests our organisms represent two species of *Micromonospora*. *Micromonospora* sp. O5 and *Micromonospora* sp. PH63 are very closely related, at ~95.5% ANI. Despite this, *Micromonospora* sp. O5 possesses a considerably larger chromosome and distinct footprint of BGCs – whether or not these two organisms truly are distinct species is a matter of debate and other algorithms may disagree with FastANI's identity estimation [248].

Knowing that they were likely to be distinct species, we then chose to investigate whether the BGCs our Chilean strains encode were unique to them or shared with their phylogenetic neighbours. BiG-SCAPE, which generates networks of BGCs based on sequence similarity [182], allowed us to identify that members of both Clade 1 (*Micromonospora* sp. O3 and neighbours) and Clade 2 (*Micromonospora* sp. PH63, *Micromonospora* sp. O5, and neighbours) possess both conserved and strain-unique BGCs. This suggests that both vertical inheritance and horizontal gene transfer are driving the evolution of secondary metabolism within these organisms. Both vertical

[143] and horizontal transfer [147, 249] are known to contribute to secondary metabolism evolution within actinomycetes, in agreement with this observation. Without knowledge of the chemical nature of the antibacterial metabolites produced by our organisms, it is difficult to infer which biosynthetic gene cluster is responsible for observed bioactivity

Clade 1 (Fig. 10) contained *Micromonospora* sp. O5 and *Micromonospora* sp. PH63, as well as *M. echinofusca, M. peucetia, M. citrea,* and *M. echinaurantiaca*. These strains were isolated from a number of sources: lake mud (*M. citrea*), chukar excrement (*M. echinofusca*), and unspecified soil (*M. peucetia* and *M. echinaurantiaca*) [238]. It is interesting that none of these strains were apparently isolated from hyper-arid soil as was the case with *Micromonospora* sp. O5 and *Micromonospora* sp. PH63, raising questions about the biogeographic distribution of the *Micromonospora* and their ability to survive a variety of environmental pressures. The pattern of broad biogeographic distribution was repeated with Clade 2 (Fig. 11) and the neighbours of *Micromonospora* sp. O3 – *M. chokoriensis* was isolated from soil adjacent to a Bangladeshi waterfall [250]; *M. cremea*, *M. zamorensis*, and *M. saelicesensis* are Spanish rhizosphere isolates – the former two from *Pisum sativum*, [228] the latter from *Lupinus angustifolius* [62]. *M. coriariae* was isolated, again in Spain, from the root nodule of a *Coriaria* plant. *M. krabiensis* represents a marine isolate from Thailand [251], and *M. avicenniae* an endophyte isolated from the roots of the mangrove *Avicennia marina* [252]. That members of Clade 1 and Clade 2 have been isolated from a diverse range of habitats spanning the globe raises questions about what role *Micromonospora* play in their natural habitat: are they generalists, adapted to thriving in a variety of stressful environments? Or is it the case instead that spores of the genera are widely dispersed throughout the globe and survive through dormancy until isolation?

4.3.5. Accurately estimating the *Micromonospora* pangenome

A pangenome is the total set of genes present in a set of organisms [84]. For meaningful information to be derived from a pangenome, it is necessary to understand the relationship between the organisms from which the pangenome is being estimated. We chose the strains selected by autoMLST as a starting point, as it preselected strains based off of 100 conserved single-copy orthologs and thus likely only included *Micromonospora.* The core of the first pangenome was 1415 genes, with 71,094 cloud genes composing the majority of the accessory genome. This core

was comparable in size to that observed in *Saccharomonospora,* which had 978 genes between 19 strains [253], and the family *Micromonosporaceae* (1,670 between seven strains) [254]. Approximately 71,000 genes in the accessory genome is large, although perhaps not unexpected when sampled from many members of a genus with an open pangenome – it suggests around a quarter of the M*icromonospora* chromosome is variable. Nevertheless, we also undertook steps to optimise pangenome prediction. First, we sought to ensure that all the members of the autoMLST were *Micromonospora*.

The concepts of phylogeny, taxonomy, and nomenclature – although closely related – are all distinct concepts. Microbial taxonomy and nomenclature are frequently being revised - aided by the abundance of genomic data. When the fraction of genome fragments that mapped between organisms in the tree was plotted, *M. pataloongensis* clearly separated from the rest of the members of the tree. This, along with the long branch with *M. pataloongensis* at the tip justified its exclusion when regenerating the pangenome. Nevertheless, the pangenome was still large - suggesting that *Micromonospora* does possess a large and open pangenome.

4.4. Summary

The work presented here was carried out to answer three questions. Firstly, by assembling the genomes of the three Chilean *Micromonospora*, we have illustrated that they possess a wide range of BGCs suggestive of excellent biosynthetic potential. By screening them for conserved regulators responsible for control of morphological development and control of specialised metabolism in *Streptomyces*, we have identified partial conservation of development between the two. We have also identified that, despite utilising TTA codons for BldA-mediated translational regulation, the organisms do not possess a functional BldA. Thirdly, we have identified that the three strains are novel species of *Micromonospora* that show semi-conservation of BGCs. We have also shown that the genus *Micromonospora* possesses a large and open pangenome, suggestive of large intra-genus diversity.

Chapter 5 The *Micromonospora* chromosome contains spatially distinct BGC populations

5.1. Introduction

In the previous chapter, through assembling the genomes of three strains of *Micromonospora* isolated from hyper-arid Chilean soil and comparing them to their phylogenetic neighbours, we demonstrated further evidence of the genus' prodigious capacity for natural product biosynthesis. The wide pangenome of the organisms identified as phylogenetic neighbours of ours, as well as the pattern of BGC conservation between closely-related strains stood out to us. In this chapter, we sought out to investigate *Micromonospora* genomics through a lens of secondary metabolism, and to seek answers to three questions:

1. "Is there conserved chromosome architecture in *Micromonospora*?"
2. "Are the BGCs of *Micromonospora* conserved within the genus?"
3. "Does chromosome architecture play a role in the conservation of BGCs in *Micromonospora*?"

5.2. Results

5.2.1. Preparation of a dataset of well-assembled *Micromonospora* chromosomes

Our dataset began with 30 *Micromonospora* chromosomes, covering complete and chromosome level assemblies available on NCBI. We began by using Nucmer to generate all v all pairwise alignments to identify misassembles present in this set, after which we discarded *Micromonospora* sp. B006 and *Micromonospora* sp. L5 (Fig. 1, B and C, respectively), reducing the number of assemblies to 28. We discarded *Micromonospora* sp. B006 because its alignments suggested either misassembly or multiple chromosomal rearrangements that were absent in the other assemblies. *Micromonospora* sp. L5 was discarded owing to a large inversion in its assembly which, if true, would bring the mid-chromosome implausibly close to the origin of replication. Table 2.2. contains the members of our set and their accession numbers. In addition, the Nucmer alignments indicated a conserved architecture in *Micromonospora* chromosomes, with high levels of synteny close to the chromosomal origin of replication contrasted by a less well-structured mid-chromosome (Fig. 1A).

Having discarded poorly assembled genomes from our set, we next set out to ensure that all our assemblies belonged to the same genus they were reported as. Using FastANI, we achieved this – the minimum ANI value was 82.713% between *M. chokoriensis* DSM 45160 and *M. echinospora* DSM 43816, and the maximum between two different assemblies was 98.903% between *M. aurantiaca* DSM 27029 and *M. aurantiaca* 110B (Fig. 5.2). This confirmed that all the members of our dataset are members of *Micromonospora* [193].

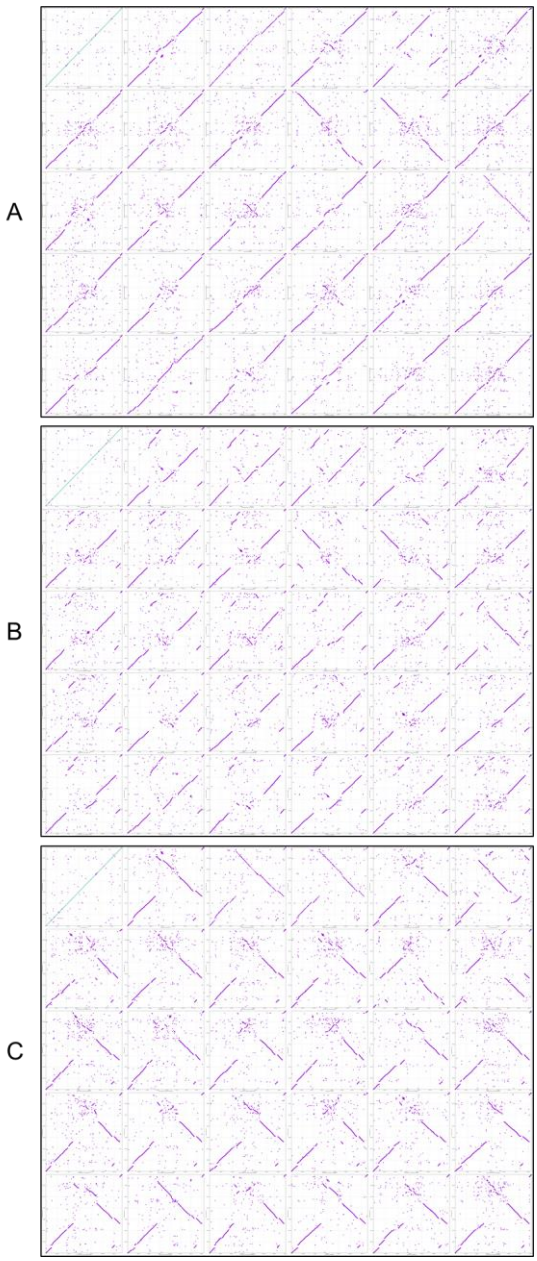


Figure 5-1 *Micromonospora* sp. L5 and *Micromonospora* sp. B006 are probably misassembled.

(5) Nucmer alignment of *M. aurantiaca* 27029 with the other 29 single-contig assemblies of *Micromonospora* chromosomes shows their conserved synteny. However, the assembly of *Micromonospora* sp. B006 (**B**) contains several spurious transpositions in the core chromosome, and the assembly of *Micromonospora* sp. L5 (**C**) contains a large inversion which brings the midchromosome close to the origin of replication, which is incompatible with proper DNA replication. For these reasons, *Micromonospora* sp. B006 and

L5 were excluded from our analysis. All chromosomes were aligned with *dnaA* at the beginning of their assembly, in the bottom right of each tile.

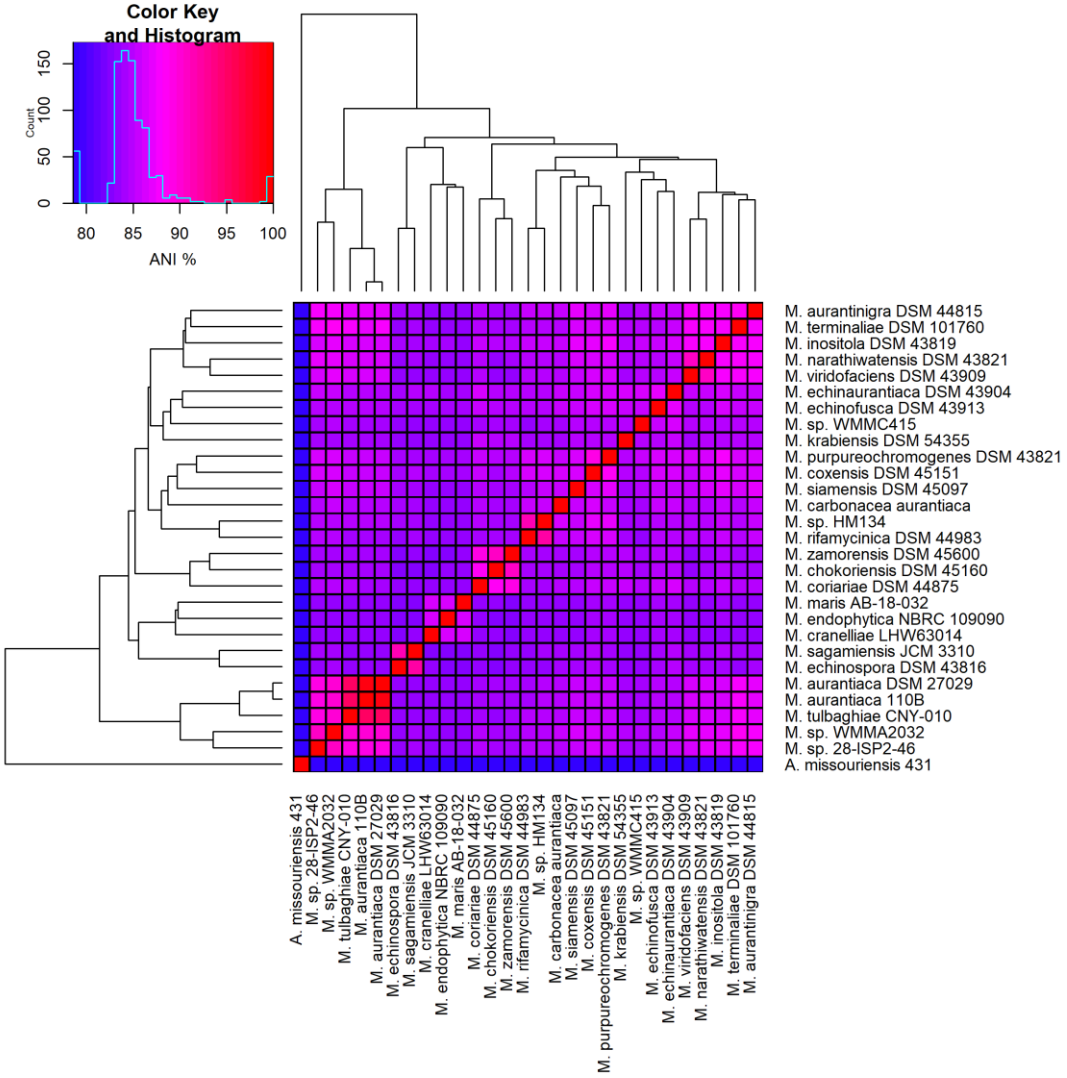


Figure 5-2 Analysed assemblies are correctly assigned as Micromonospora.

Heatmap of ANI values calculated by the FastANI algorithm. None of the assemblies in our dataset had ANI values <80%, confirming they belong to a single genus.

5.2.2. The *Micromonospora* chromosome has conserved architecture

Upon examination of the alignments generated by Nucmer, we observed conserved chromosomal architecture. Specifically, the origin-proximal region of the chromosome is highly conserved, whereas the opposite pole shows much less conservation (Fig. 5.3.). This suggests that the terminus-proximal region of the *Micromonospora* chromosomes acts as a hotspot of recombination compared to the origin. We also observed inversions in the origin islands of *M. cranelliae* LHW63014 and *M. echinaurantiaca* DSM 43904 (Fig. 5.3.) (Available in high-resolution at this thesis' GitHub repository). That large inversions only occurred in two of the organisms in our set made us curious about leading/lagging strand bias in gene distribution and so we also tested this. The *Micromonospora* strains in our data set have significantly more genes on the leading strand of either replichore, however there was no difference between the number of genes on the top and bottom strand of DNA in their chromosomes (Fig 5.4.) - in other words, genes are preferentially oriented in-line with the replication fork.

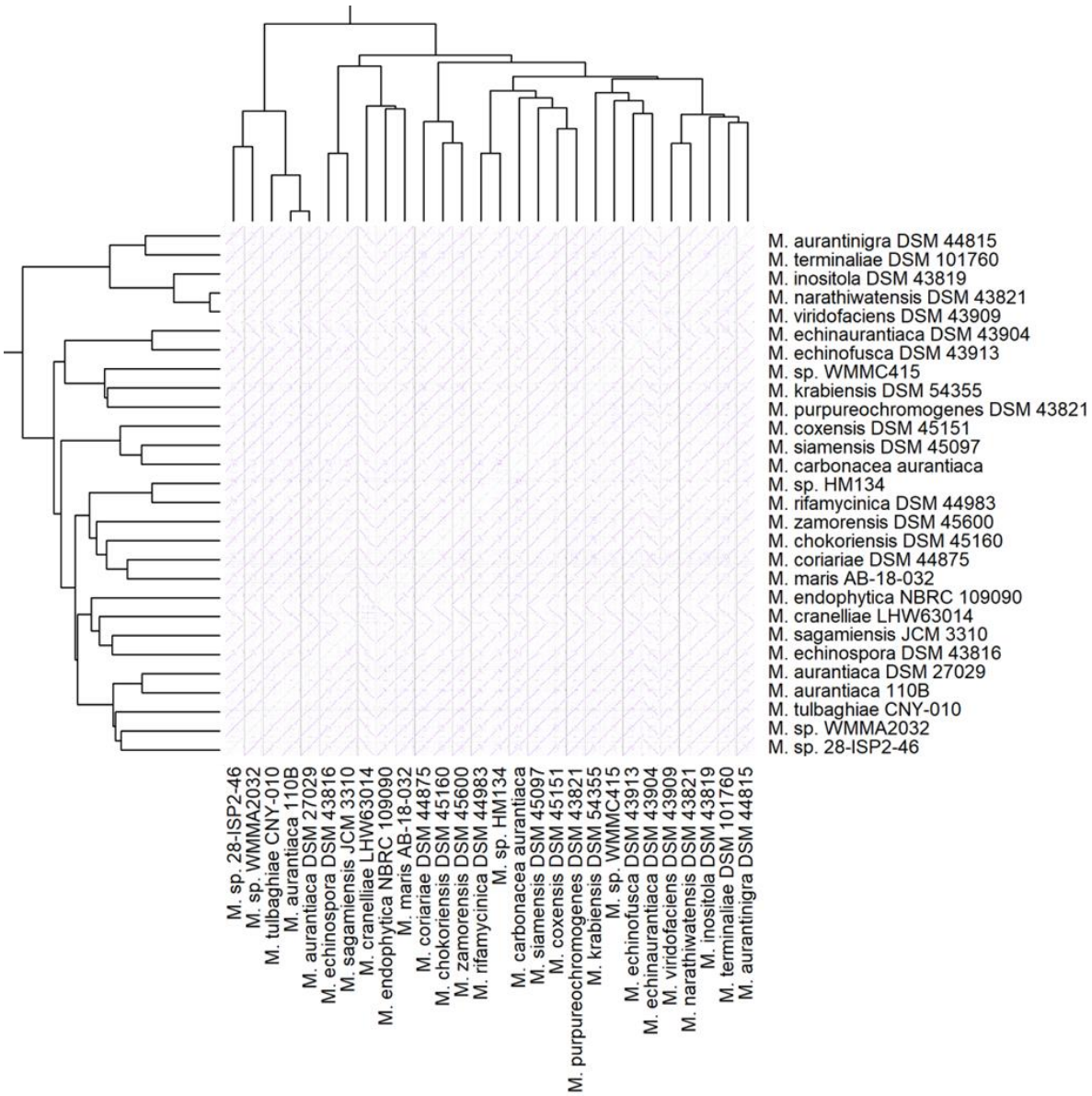


Figure 5-3 *Micromonospora* chromosomes possess conserved chromosomal architecture.

28x28 grid of Nucmer plots, ordered by clustering of the ANI heatmap. *Micromonospora* chromosomes are highly conserved in the origin proximal region, however the midchromosome region is highly diverse.

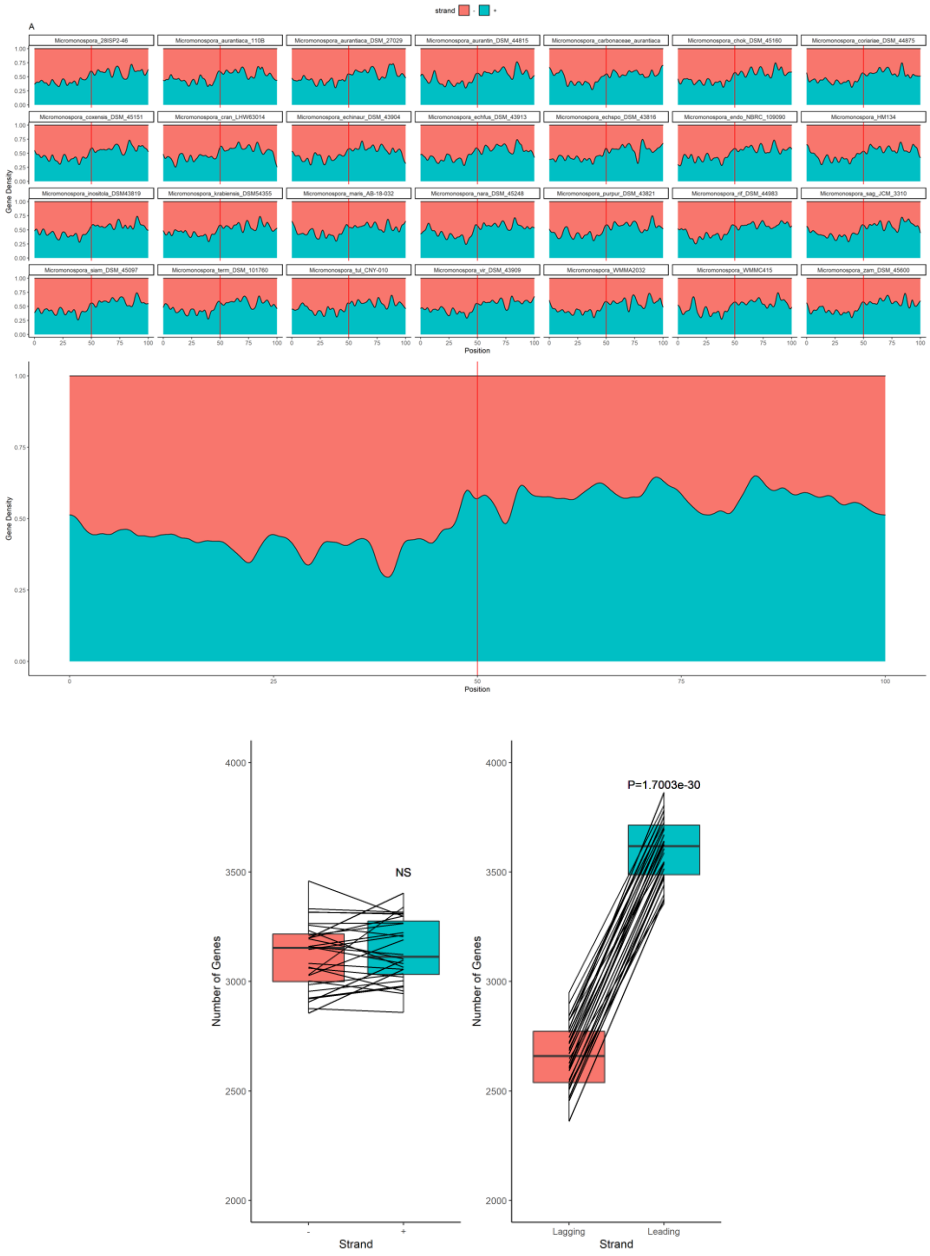


Figure 5-4 Strand bias in *Micromonospora*.

Top panel: Kernel density estimates of bias towards the top (+) and bottom (-) strands of the chromosome in *Micromonospora* per assembly in our dataset, with *oriC* indicated by red vertical lines as position 50. Middle panel: Kernel density estimate for every gene in our dataset, indicating strand switching occurs at the origin and midchromosome in *Micromonospora.* Bottom panel: Comparison of strand bias in our organisms. There is no difference between the number of genes in the top and bottom strands, however there were significantly more genes in the leading strand than the lagging strand.

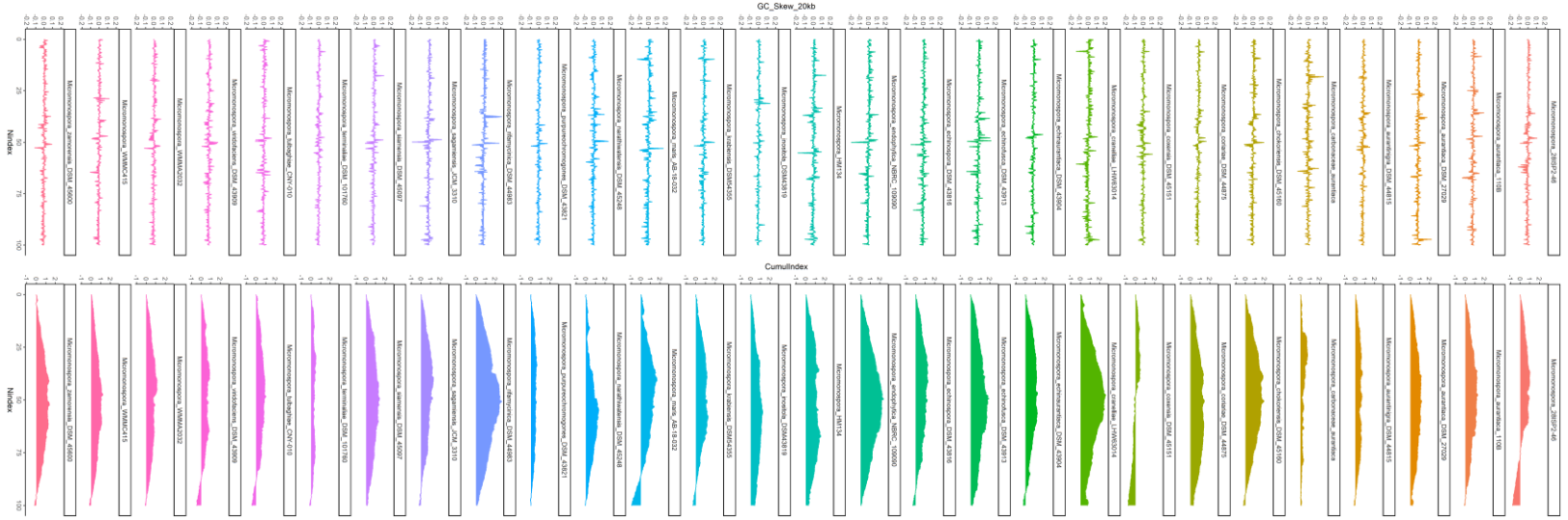


Figure 5-5 GC Skew in single-contig *Micromonspora* chromosomes.

Rolling (Left) and Cumulative (Right) GC Skew across chromosomes analysed. In most strains, skew reversed around 50 units from the origin of replication.

5.2.3. *Micromonospora* possess a rich and diverse repertoire of BGCs

Knowing that our dataset was populated only by fully assembled *Micromonospora* chromosomes, we then passed them on to antiSMASH to elucidate the number and nature of biosynthetic gene clusters possessed by them. This revealed that the genus to be rich in biosynthetic gene clusters, possessing 99 different types between them and 511 BGCs total, dominated by terpene, PKS, and NRPS classes (Fig 5.5.). The mean chromosome size in the assemblies was 6,914,258 bp and the mean number of BGCs carried was 18.25 (Fig. 5.6.), with wide variation between organisms.

Figure 5-6 Predicted biosynthetic repertoire of the *Micromonospora* genome sequences analysed.

Micromonospora predominantly carry terpene class BGCs, however there was a large number of singleton BGCs amongst our assemblies.

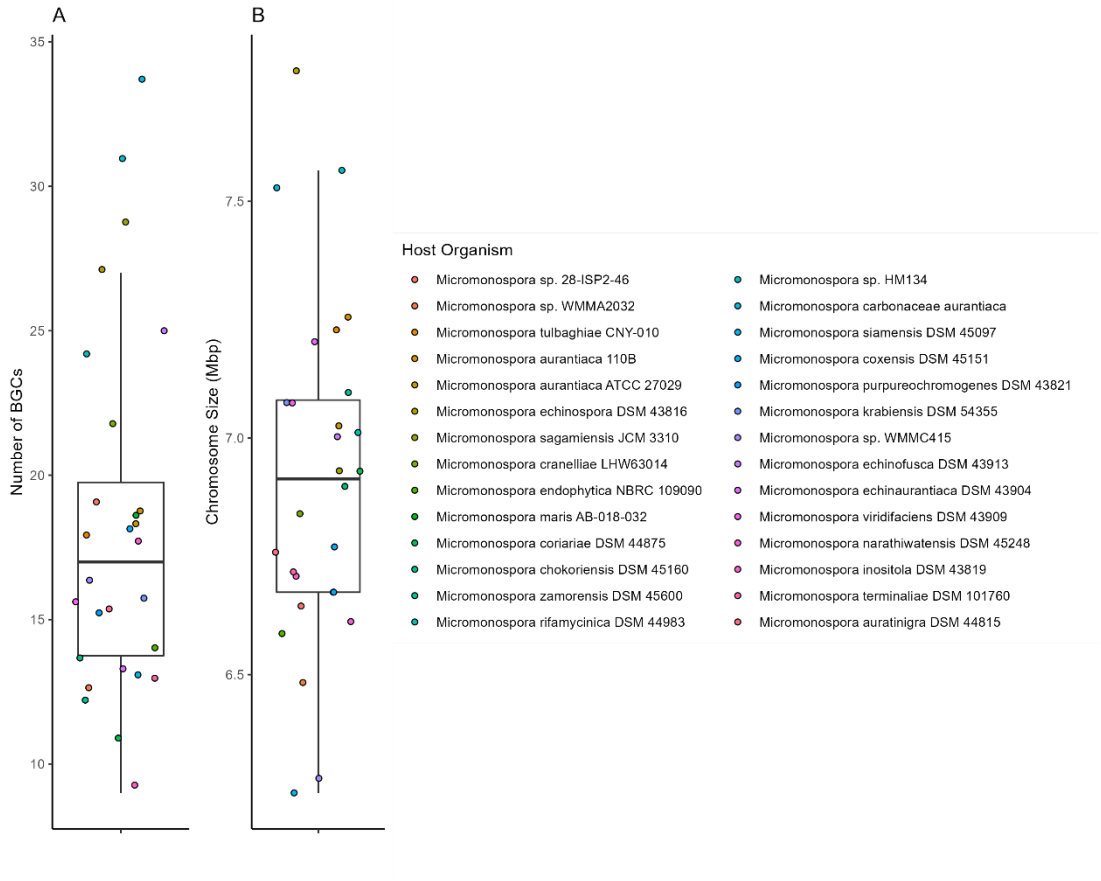


Figure 5-7 Chromosome size and BGC counts of analysed *Micromonospora*.

(**A**) Number of BGCs and (**B**) Chromosome size of *Micromonospora* analysed. Characteristic of *Micromonospora*, our organisms possess large chromosomes with many BGCs.

5.2.4. BGC Content Contributes to Chromosome Size in *Micromonospora*

After noting the variability between *Micromonospora* chromosome sizes and the number of BGCs they possessed, we sought out to see if the two values were correlated. To do this, we performed a linear regression between the number of BGCs present in our organisms and the size of their chromosomes (Fig. 4A). Understanding that BGCs vary in size, we also sought to see if there was a correlation between the percentage of the chromosome occupied by BGCs and chromosome size (Fig. 4B). We found a weak but statistically significant positive correlation for both cases, suggesting that BGCs both contribute to genome growth in *Micromonospora* and that larger chromosomes have more genomic space devoted to secondary metabolism. In addition, the proportion of the chromosome occupied by BGCs was highly variable and ranged from <5% of DNA content to >20% (Fig. 4B).

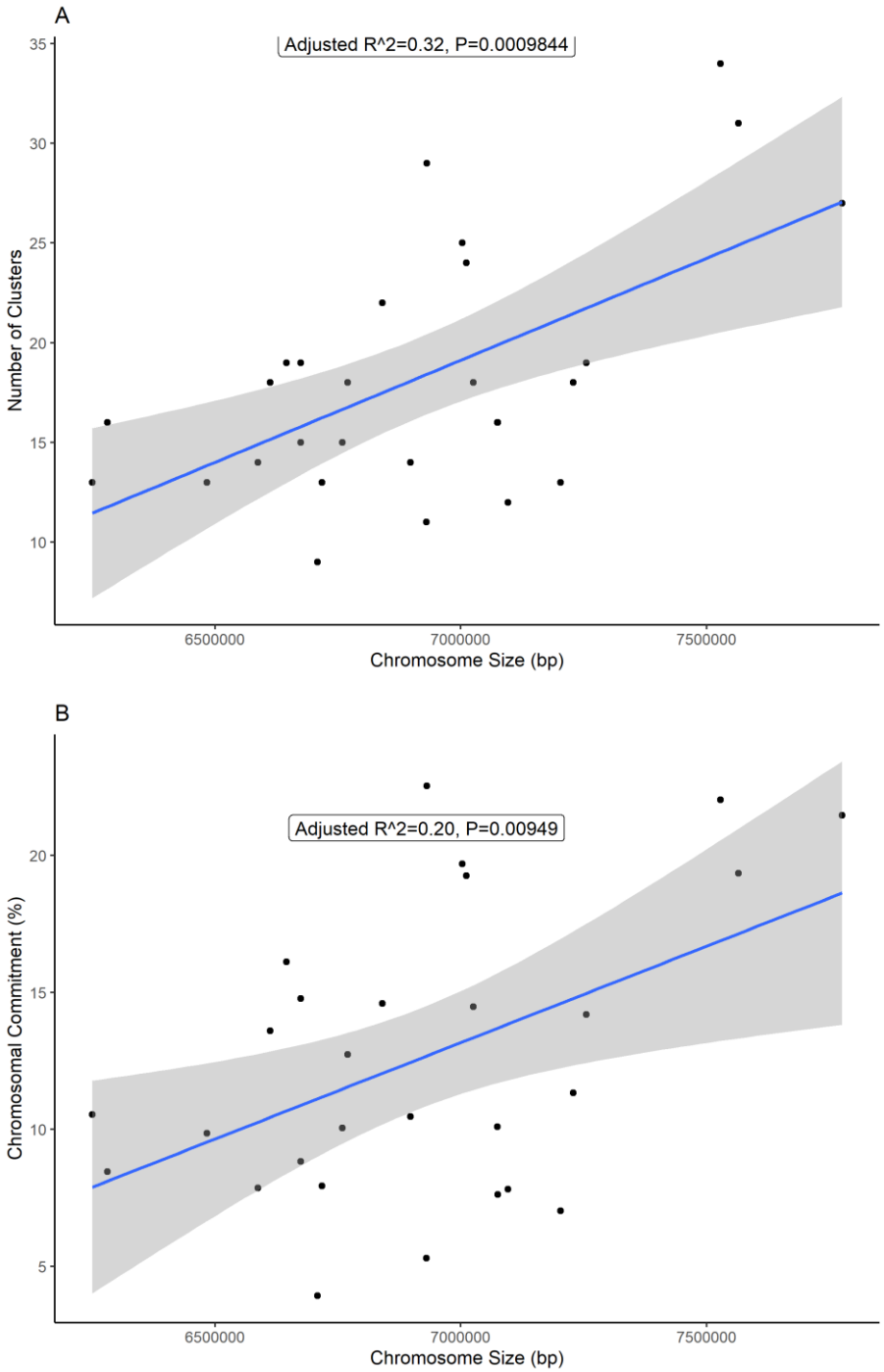


Figure 5-8 BGC accumulation drives the size of Micromonospora chromosomes.

(**A**) Linear regression of chromosome size against number of BGC clusters and of (**B**) Chromosomal Commitment against chromosome size indicates that committing genomic space to BGCs is a partial driver of chromosome growth in *Micromonospora*.

5.2.5. Biosynthetic Gene Clusters are present in both the core and variable regions of the chromosome

Whilst performing quality control on our dataset we observed that *Micromonospora* chromosomes have conserved architecture with syntenic regions close to the origin of replication and less synteny towards the midchromosome. This led us to compare the loci of BGCs in *Micromonospora* chromosomes. To achieve this, we normalised the loci of the biosynthetic gene clusters and plotted them on a linear pseudochromosome. This revealed that there are two hotspots at opposite poles of *Micromonospora* chromosomes where biosynthetic gene clusters accumulate: one at the origin of replication, and one at the terminus (Fig. 5.8.). This was the case for all 28 *Micromonospora* included in our dataset. From this, we can conclude some chromosomal loci are favoured over others for BGC accumulation. Of note was that, at the organism level, BGC accumulation was favoured at one arm of the ori-distal pole, rather than symmetrically distributed across the pole.

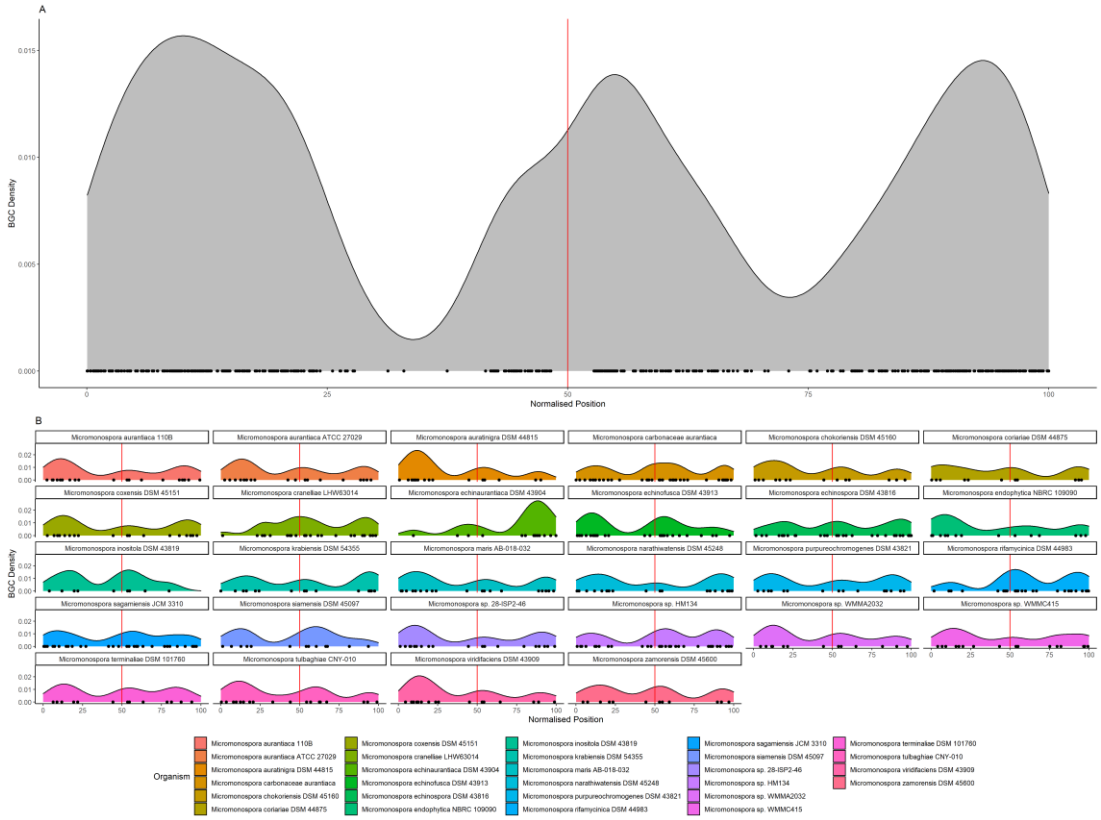


Figure 5-9 BGC distribution is conserved in Micromonospora chromosomes.

(**A**) Kernel density estimate of chromosomal position of BGCs belonging to *Micromonospora,* with the origin of replication at position 50. (**B**) Kernel density estimates of BGC distribution by species. These indicate that *Micromonospora* have two BGC dense regions, at opposite poles of their chromosome.

5.2.6. The distribution of BGCs in *Micromonospora* chromosomes is analogous to, but distinct from, that of *Streptomyces*

In our previous analysis, to enable comparability of the *Micromonospora* pseudochromosome to the linear chromosome of streptomycetes, the origin of replication was located at position 50 with regions 0 and 100 representing the opposite pole of the chromosome, as *Streptomyces* are the most prolific actinomycete natural product producers. We compared the distribution of BGCs in our dataset to the distribution of BGCs in 20 members of the *Streptomycetaceae* with their linear chromosomes sequenced from telomere to telomere [90]. In these organisms, BGCs are known to accumulate in the arms of the chromosomes. Analogous to this is the accumulation of BGCs in the midchromosome of *Micromonospora.* What was notably different, however, was an accumulation of BGCs close to the *Micromonospora* origin of replication that was absent in the linear chromosomes (Fig. 5.9.). From this, we concluded that although origin-distal BGC accumulation occurs in both the linear chromosomes of streptomycetes and the circular chromosomes of *Micromonospora*, there is also a driver of origin-proximal BGC accumulation in *Micromonospora*. The normalised position of these origin-proximal BGCs was within a narrow range of each other, and so we sought out to test if these BGCs served a conserved role, and whether there was a difference between the BGCs in the midchromosome versus the origin.

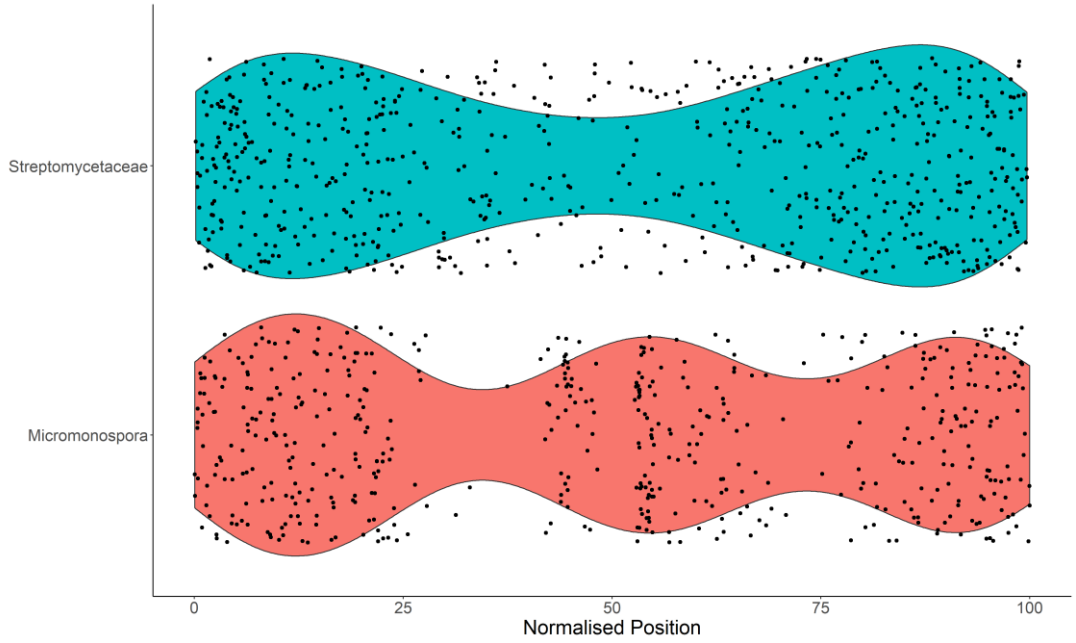


Figure 5-10 *Micromonospora* genomes encode a distinct BGC distribution from Streptomyces.

Violin plots of BGC distribution in *Streptomyces* and *Micromonospora* with the *Micromonospora* origin of replication at position 50. *Micromonospora* are distinct by having an *oriC*-proximal suite of BGCs that streptomycetes lack.

5.2.7. Different BGC classes are not distributed evenly across *Micromonospora* chromosomes

After showing that chromosomal regions are rich in biosynthetic gene clusters, we then sought to examine the distribution of classes of BGCs across *Micromonospora* chromosomes. We found that some classes of BGC are enriched at chromosomal loci whereas others were predominantly towards the midchromosome (Fig. 5.10.) For example, T3PKS, terpene, and NAGGN clusters were mostly located close to *oriC*, with NAGGN clusters having a median location at 14.5% (normalised distance) from the start of *dnaA* and T3PKS containing clusters at 6%. On the other hand, clusters containing NRPS, Siderophore, T1PKS, and T2PKS biosynthetic genes had median distances of 37.97%, 41.4 %, 37.61%, and 41.8%. Interestingly, terpene containing clusters appeared to exist as three different populations – with one close to the origin of replication, one towards the midchromosome, and one halfway between the two. From this, we concluded that BGC type affects whether that cluster lies on the chromosome. However the terpene clusters having three distinct populations demonstrate that this cannot be the only driver. Interestingly, deletion of the *oriC*-proximal Terp2 terpene cluster in *Salinispora* results in an apigmented phenotype owing to deletion of precursor biosynthesis, whereas disruption of the *oriC*-distal clusters only disrupts pigmentation due to deletion of pigment-modifying enzymes [227].

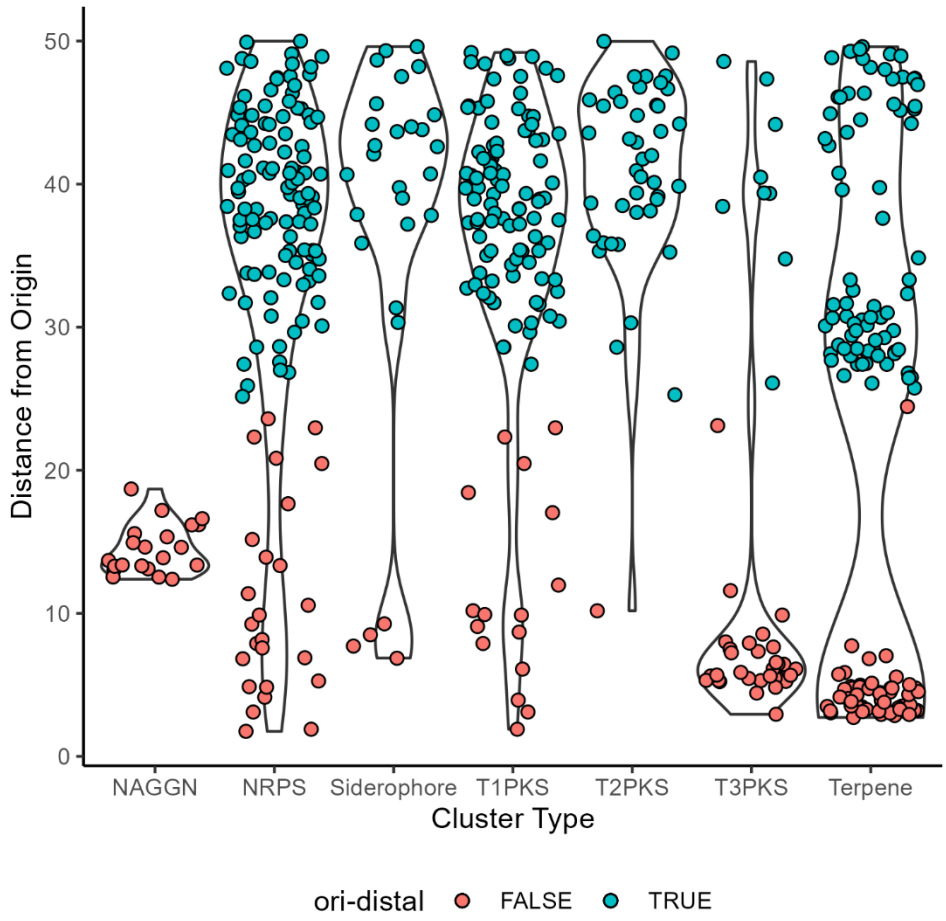


Figure 5-11 BGC class affects chromosomal location.

Violin plots of selected classes of BGC. NAGGN and T3PKS containing BGCs are notable as being present predominantly in the ori-proximal region (Distance <25) whereas NRPS, T1/2 PKS, and Siderophore clusters are predominantly in the ori-distal region (Distance >25). Terpenes are present in three groups in both the origin proximal and distal regions.

5.2.8. Homologous BGCs are syntenic across *Micromonospora* chromosomes.

Observing that there was a pattern of distribution where BGCs of given classes appeared to localise at particular regions of the *Micromonospora* chromosome, we hypothesised that these were in fact homologous BGCs. To test this, we generated a BiG-SCAPE network to group similar BGCs present in our dataset, annotated by the position of the BGCs on the chromosome. This network partially confirmed our hypothesis – BGCs at the origin of replication shared networks (Fig. 5.11.). We also observed that some BGCs were placed in networks of otherwise syntenic BGCs – these clusters belonged to organisms which Nucmer analysis suggested a historical inversion of the origin of replication and, thus, whether the BGC was located up or downstream of the origin. We also observed that our dataset contained a large number of singleton BGCs, not associated with a network. These singletons mostly existed away the origin of replication and thus we sought out to see if the origin-proximal region of the chromosome and the midchromosome contained different populations of BGCs.

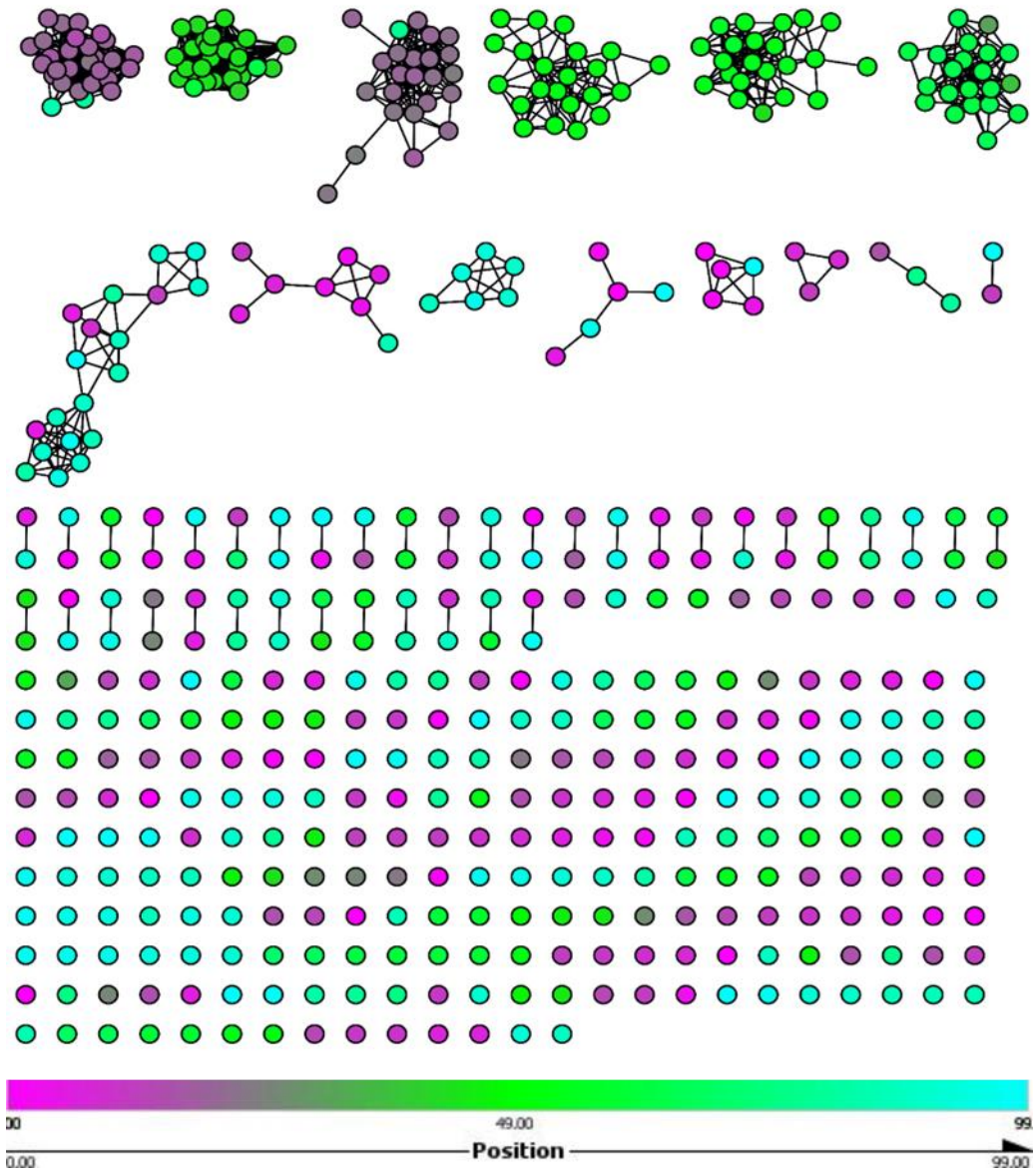


Figure 5-12 Homologous clusters are syntenic in *Micromonospora*.

BiG-SCAPE network of BGCs coloured by chromosomal position, with the *dnaA* at position 50 (green) and the midchromosome at positions 0 and 99 (magenta and cyan). Homologous BGCs in our assemblies are syntenic and largely represented in the origin proximal region. On the other hand, small gene cluster families and singletons were mostly located in the origin distal region of the chromosome.

5.2.9. Classification of BGCs into Gene Cluster Families Suggests BGC Gene Content is Dynamic

After using BiG-SCAPE to generate networks of closely related BGCs, we sought to investigate the contents of these networks. By aligning the BGCs present in the larger networks we found evidence that BGC content and the genes contained within those BGCs in *Micromonospora* is dynamic. For example, ori-proximal terpene clusters were a mean of 2.9 (Fig. 5.12. A) and 4.6 (Fig. 5.12 B) normalised units from *oriC*. The core and accessory biosynthetic genes, as well as regulatory genes in these clusters are conserved and syntenic, however transporter-encoding genes were less conserved and did not show synteny in the clusters. Two other large networks were populated by terpene clusters, one which was populated by 24 clusters at a mean of 28.5 units from *oriC* and contained a highly conserved and syntenic core, accessory, regulatory, and transport genes (Fig. 5.13). The second contained 28 members – a mix of RiPP like and terpene, RiPP-like clusters – that were a mean of 30 units from *oriC*. (Fig. 5.14). All members of this second network contained a RiPP-like set of core biosynthetic genes, whilst 18 members of the network also possess core terpene biosynthetic enzymes.

The T3PKS clusters identified in our *Micromonospora* formed a 28 membered network at with a mean distance of 7 units from *oriC*. (Fig. 5.15.), members of which showed synteny across core, accessory, regulatory, and transport genes. Cluster 24 of *M. echinofusca* stood out in this network, with an insertion of core genes associated with thiopeptide and LAP biosynthesis. We also investigated a cluster containing NAGGN type BGCs, responsible for osmoprotection (Fig. 5.16.) – this cluster was notable for having conserved core biosynthetic genes but differences in accessory genes. Representatives of NAGGN were not present in every organism we analysed despite the metabolite type's high level of conservation – the network was composed of 21 BGCs that were on average 21 units of distance from *oriC*.

Finally, we examined a three-membered network composed of hybrid trans-AT NRPS-PKS clusters (Fig. 5.17.). These clusters belonged to the two *M. aurantiaca* species as well as the *M. tulbaghiae* species present in our dataset. These three strains represent one species by ANI, and so it was surprising that they only had one network to themselves. This offers an insight into the evolution of *Micromonospora* specialised metabolism, showing that even rare BGCs are maintained across species boundaries.

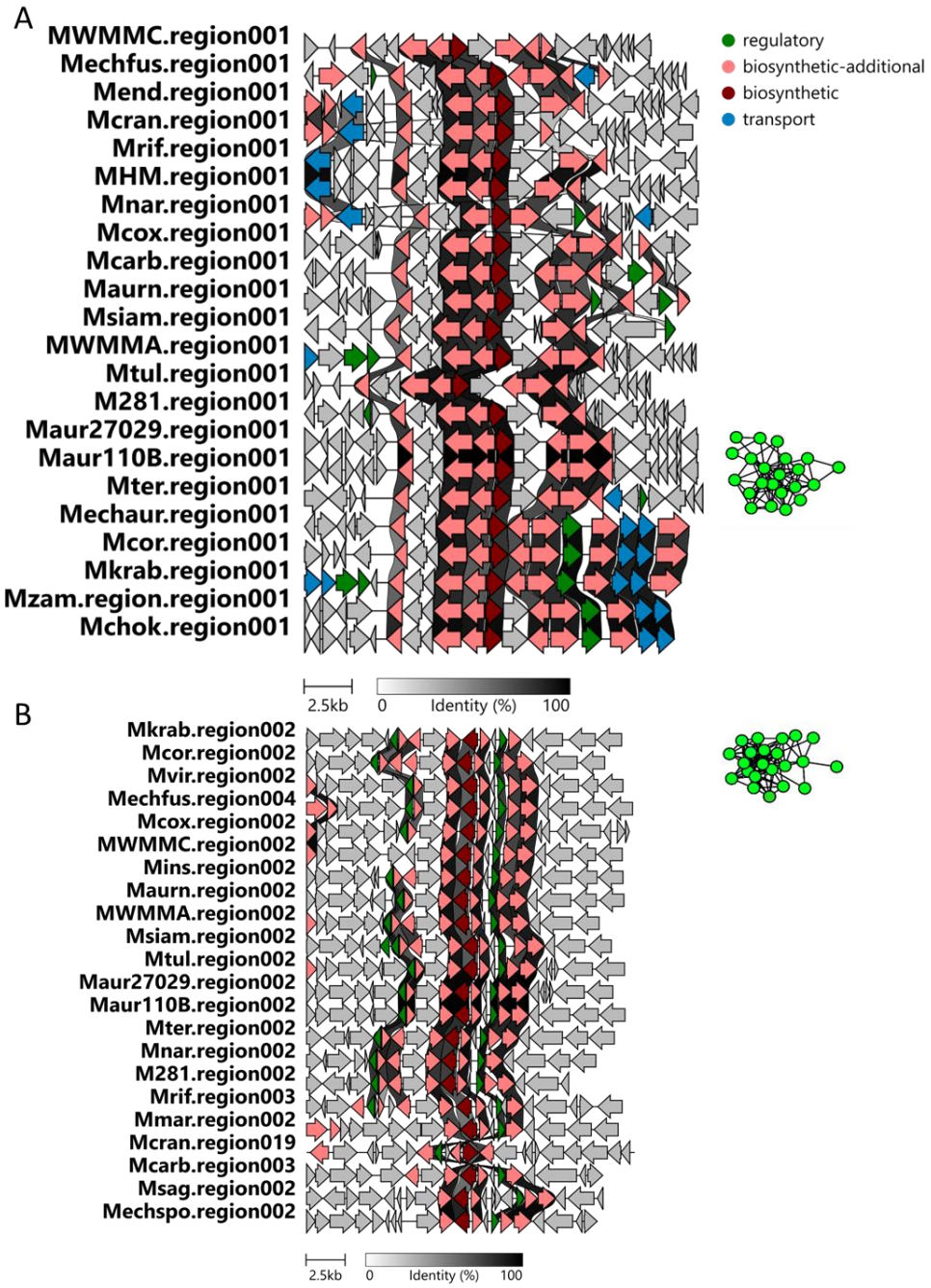


Figure 5-13 *oriC*-proximal terpene clusters are conserved in Micromonospora.

Alignment generated by clinker using the BGCs present in the indicated network and coloured by role in specialised metabolism as predicted by antiSMASH. (A) Alignment of phytoene synthase containing terpene BGCs. B) Alignment of terpene synthase containing terpene BGCs. The core and accessory biosynthetic genes in these clusters were syntenic.

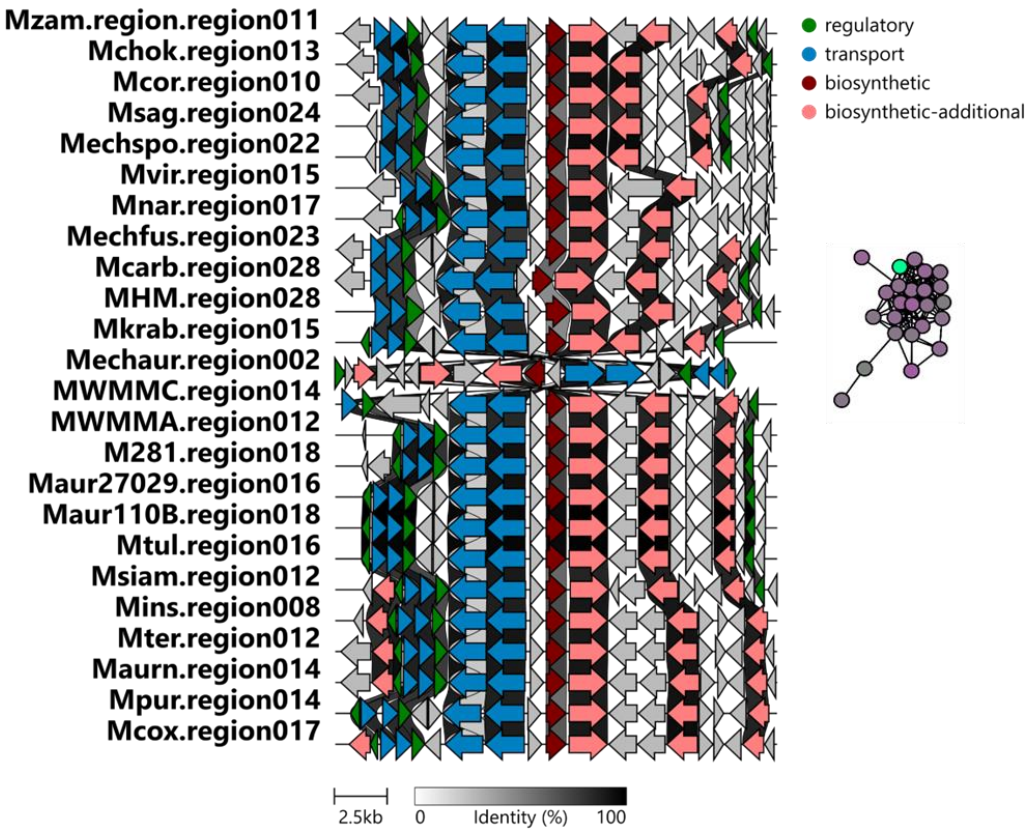


Figure 5-14 Conserved *oriC*-distal terpene clusters of Micromonospora.

Alignment generated by clinker using the BGCs present in the indicated network and coloured by role in specialised metabolism as predicted by antiSMASH. These clusters contain highly conserved regulators and transporters

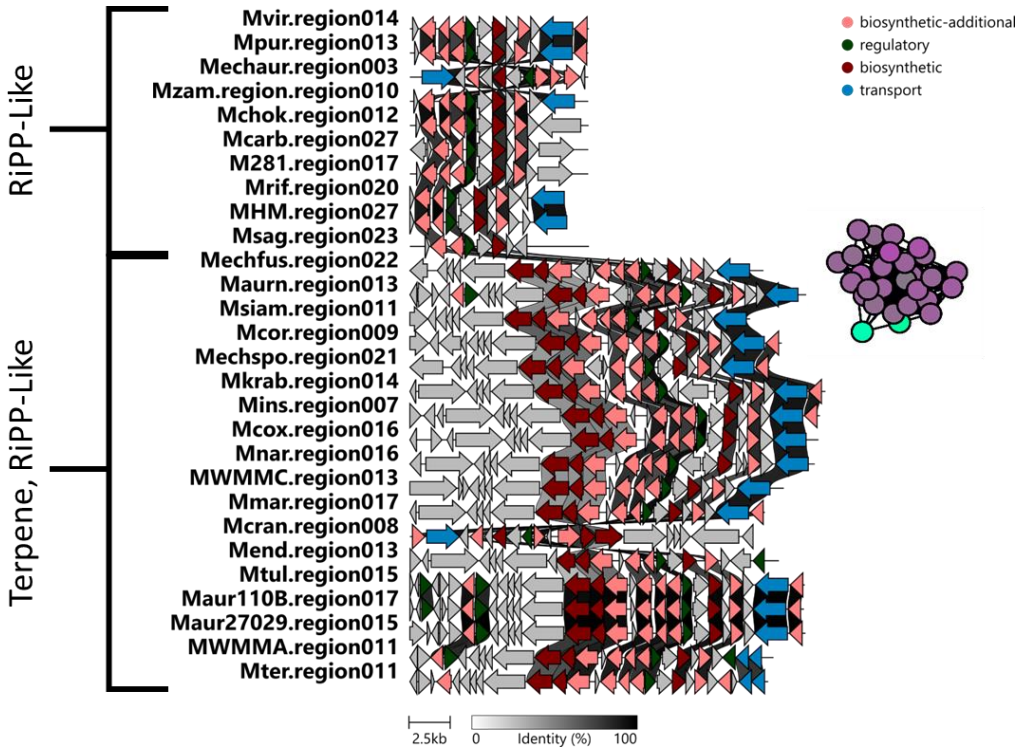


Figure 5-15 BGC network containing *oriC*-distal RiPP-Like and Terpene-RiPP like Hybrid Clusters from complete Micromonospora chromosomes.

Alignment generated by clinker using the BGCs present in the indicated network and coloured by role in specialised metabolism as predicted by antiSMASH.

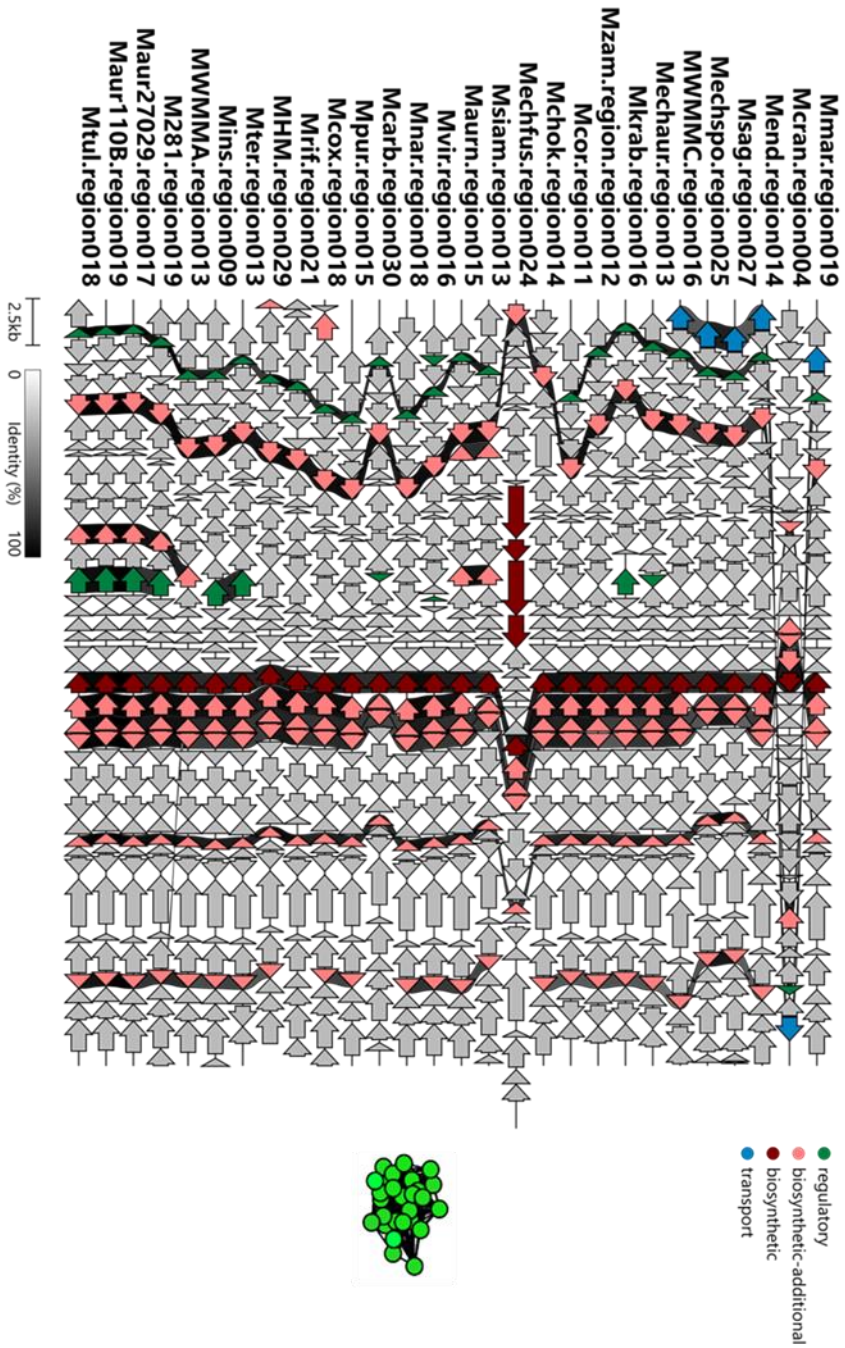


Figure 5-16 Conserved *oriC*-proximal T3PKS clusters from complete Micromonospora.

Alignment generated by clinker using the BGCs present in the indicated network and coloured by role in specialised metabolism as predicted by antiSMASH. *M. echinofusca* (Mechfus.region024) has historically experienced and insertion of thiopeptide synthesising core biosynthetic genes at this locus.

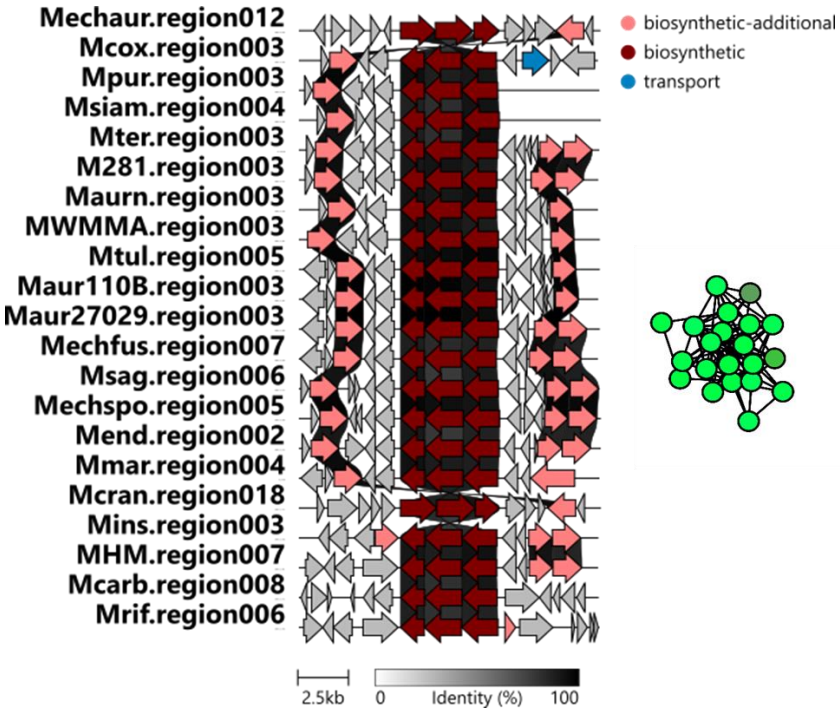


Figure 5-17 BGC Network of ori-proximal NAGGN clusters from complete *Micromonospora*.

Alignment generated by clinker using the BGCs present in the indicated network and coloured by role in specialised metabolism as predicted by antiSMASH. The clusters in this network were notable for having inconsistent accessory biosynthetic gene content around the core NAGGN assembly genes.

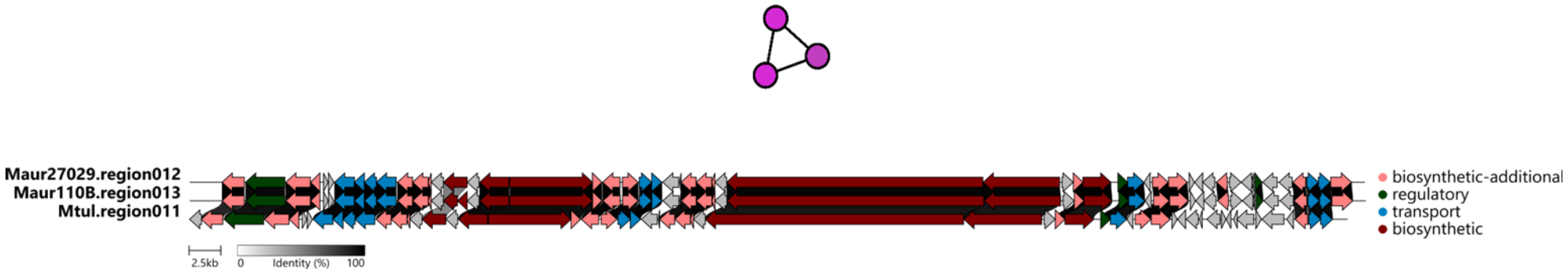


Figure 5-18 BGC Network Containing the conserved hybrid Trans-AT NRPS-PKS of *M. aurantiaca* and *M. tulbaghiae*.

Alignment generated by clinker using the BGCs present in the indicated network and coloured by role in specialised metabolism as predicted by antiSMASH. The clusters in this network were notable as being unique to these three closely related strains.

5.2.10. *OriC*-Distal Biosynthetic Gene Clusters show greater diversity than origin-proximal clusters

After noting that the largest networks of BGCs predominantly occurred close to the origin of replication, we hypothesised that BGCs close to the origin and those in the *oriC*-distal region could be described as distinct populationsof BGC. To test the hypothesis that different regions of the *Micromonospora* chromosome contained different populations of BGCs, we divided the chromosome into two distinct regions: BGCs belonged to the origin region if their normalised locus was higher than 25 but less than 75, else they were designated as ori-distal BGCs. Using the Gene Cluster Families identified by BiG-SCAPE in our previous network analysis, we showed that the origin region contained less BGC diversity than the Mid-Chromosome by Shannon, Simpson, and Inverse-Simpson diversity indices. This indicates the ori-distal half of the chromosome has a more diverse BGC population than the origin of replication. The only strain exempt from this trend was *M. carbonaceae*, which incidentally encodes the greatest number of BGCs in our dataset.

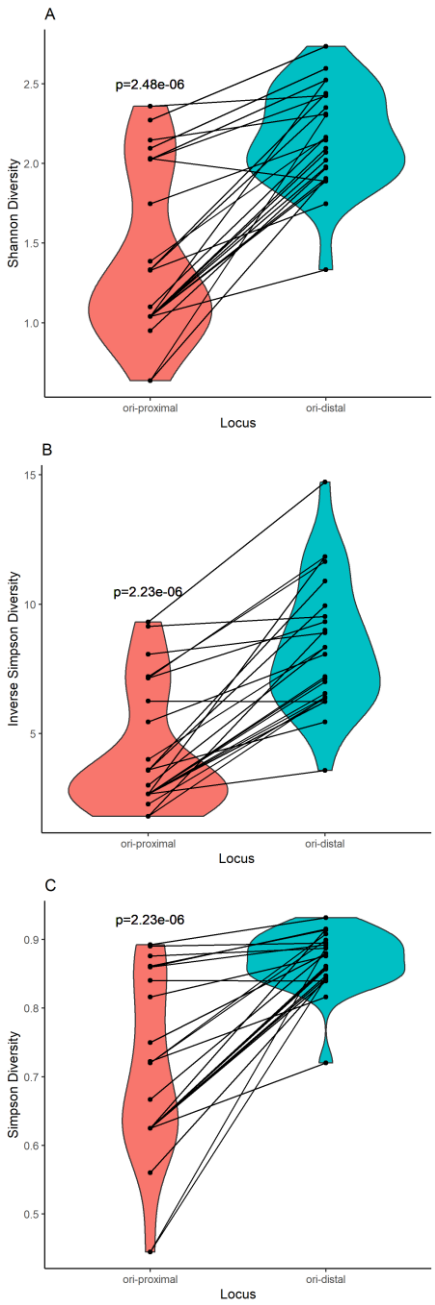


Figure 5-19 The *Micromonospora* mid-chromosome contains a greater diversity of BGCs than the origin of replication.

Shannon (**A**), Inverse Simpson (**B**), and Simpson (**C**) diversities of the origin and mid-chromosome regions of *Micromonospora* chromosomes. For all measures, the mid-chromosome was more diverse than the origin. P values were calculated by Wilcoxon's signed-rank test with continuity correction.

5.2.11. The *oriC*-distal region of the *Micromonospora* chromosome contains a greater density of BGCs than the origin region

One possible reason for the greater diversity when antiSMASH was used to identify multiple putative BGCs near each other, they are grouped into a single region of potentially hybrid clusters. This identification of regions containing proximal-clusters allows for the mapping of BGC-dense regions in *Micromonospora* chromosomes. Mapping these clusters revealed that hybrid BGCs were predominantly found in the midchromosome (Fig. 5.20. A). However, at an organismal level, there was no significant difference between the number of hybrid BGCs present at the midchromosome compared to the origin (Fig. 5.20. B). There was, however, a significant difference between the number of non-hybrid BGCs between the two regions (Fig. 5.20. C). This suggests that hybrid BGCs are not the contributing factor to the greater diversity observed in the *Micromonospora* midchromosome.

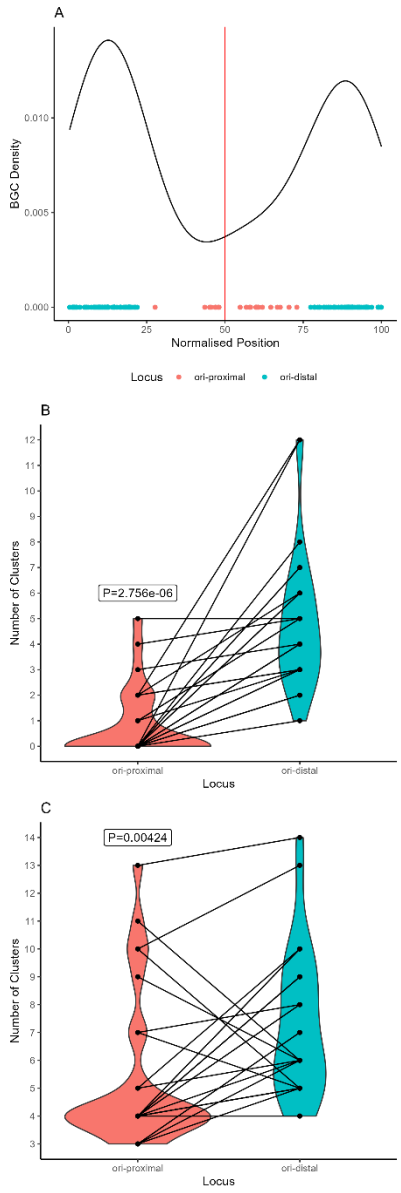


Figure 5-20 Hybrid BGCs occur more commonly at the mid-chromosome.

(**A**) Kernel density estimation of hybrid BGC loci in Micromonospora suggests that hybrid BGCs occur more commonly towards the mid-chromosome (**B**) Violin plot of the number of hybrid BGCs in the origin vs the mid-chromosome of each member of our dataset (**C**) Violin plots of the number of non-hybrid BGCs present in the organisms in our dataset, P values were calculated by Wilcox's signed rank test with continuity correction.

5.3. Discussion

5.3.1. Quality control, Chromosomal Triage, and Architectural insights

The aim of the work described in this chapter was to characterise the genomics of secondary metabolism in *Micromonospora*, by utilising high-quality genome assemblies to support mapping the relative loci of BGCs within chromosomes. We firstly sought to collect a dataset of high-quality *Micromonospora* assemblies.

What is “high-quality” is a subjective matter – here we defined it as an assembly contained in a single contig whose predicted physical structure agreed with other published *Micromonospora* genomes. As actinomycete genera contain conserved origin islands with a high degree of synteny [90] we believe that these criteria were sufficient to exclude large-scale, biologically implausible misassemblies from our dataset whilst not rejecting datasets based on events such as putative genome rearrangements. We chose to exclude *Micromonospora* sp. L5 owing to a rearrangement in its assembly which would unbalance DNA replication by bringing the terminus of replication in the midchromosome adjacent to the origin of replication. We chose to remove *Micromonospora* sp. B006 as there were large, conserved, regions that were asyntenic with other members of our dataset. As the positions of BGCs were calculated relative to each organisms’ origin of replication, the calculated loci from these organisms would be spurious. ANI analysis supported that our members are all *Micromonospora*, with a minimum identity of 82.713% between organisms. This step was important to resolve the taxonomic identity of our organisms and ensure that they were related enough for comparing the loci of their BGCs to be worthwhile, as well as by mitigating the possibility that their taxonomy had been misassigned [256].

The Nucmer alignments we performed as part of the quality control process also allowed insight into the ultraarchitecture of *Micromonospora* chromosomes, revealing the ori-distal pole of the chromosome to be poorly conserved compared to the ori-proximal region, as well as highlighting the inverted origin islands of *M. cranelliae* and *M. echinaurantiaca*. We also examined our organisms for strand bias, a frequently observed phenomenon across bacteria where genes are preferentially located on the leading strand of the chromosome [257]. In agreement with this, our organisms preferentially encoded genes on the leading strand of their chromosome. Upstream of *oriC*, genes were preferentially encoded on the minus strand of the chromosome, whereas downstream genes were preferentially encoded on the minus strand. GC

skew also matched this trend. This suggests that the midchromosome of *Micromonospora* also serves as the site of termination replication [258].

5.3.2. Genome Mining and Locus Mapping

With our dataset collected our first question was how committed are *Micromonospora* to specialised metabolism. In line with other actinomycetes, the *Micromonospora* in our sample possessed large chromosomes that were rich in BGCs. What was surprising however was the variability in how much “genomic real estate” the genus commits to secondary metabolism, ranging from 3% to 22% of their chromosome. This contrasts with closely related genera such as *Salinispora* which devote ~10% of their genome to specialised metabolism [121].

Having established this, we next questioned whether BGCs are uniformly distributed across the chromosome or not. In *Streptomyces*, for example, BGCs are predominantly found in the telomeres of the organisms’ linear chromosomes [49]. The BGCs of our *Micromonospora* were distributed across two loci – the ori-distal region contained a rich and diverse set of BGCS, in a way analogous to linear-chromosomed streptomycetes. Different, however, was the fixation of ori-proximal clusters. We observed that the location of BGCs was partially driven by the class of molecule encoded by that BGC. Type I and II polyketides, Nonribosomal peptides, and siderophores (the majority of which were desferrioxamines) were found in the ori-distal region.

The BiG-SCAPE generated network of Gene Cluster Families present in our *Micromonospora* showed that homologous BGCs are syntenic within the genus (or reverse syntenic in strains with reversed origin islands), which confirms that synteny is maintained through vertical inheritance, even in small GCFs . Previous work has explored the genus level distribution of BGCs in *Amycolatopsis* [91] and *Salinispora* [143], this work builds on it by introducing evidence of a core set of BGCs in *Micromonospora*, as well as providing evidence that the nature of a BGC partly determines its fixation into the core set. Further comparison of BGC distribution in other genera of actinomycetes, as well as successful natural producers of other bacterial families will further shore up our understanding of BGC evolution and what drives incorporation of a BGC and its cognate natural product(s) into an organism’s core suite. Contrary to the core BGCs are the diverse set of BGCs present in the ori-distal region of the *Micromonospora* chromosome. This was illustrated by the large number of singletons gene cluster families present in the *oriC*-distal region.

We demonstrated that, when split into two regions – the *oriC* proximal region which describes the chromosome half on either side of the chromosome, and the ori distal region which describes the other half – the ori distal region consistently possesses a greater diversity of BGC content. Although not organisms *per se*, these ecologic measurements are proxies for entropy – how difficult it is to predict a sample from a population [259] and so were appropriate for us to employ, treating different regions as analogous to habitats occupied by BGCs. It could be argued that bisection of the chromosome is a crude way of dividing it, missing nuance of different genomic islands – however, despite the crudeness, we were able to detect a difference in populations between the two regions.

5.3.3. Drivers of BGC fixation

The question stands: what is the driving factor in fixation of BGCs in *Micromonospora* chromosomes, and the partitioning of BGCs into different populations? The fixed clusters orbited the origin of replication, and encoded functions such as compatible solute production [109], and pigment development [227]. NAGGN type clusters are responsible for the biosynthesis of the compatible solute NAGGN, a dipeptide derived from two units of glutamine. NAGGN is overwhelmingly represented amongst members of the Gram-negative *Pseudomonas* and *Sinorhizobium*, as well as members of the *Micromonosporacea* such as *Salinispora*. This raises the possibility of an ancient horizontal gene transfer event mediating the acquisition of NAGGN clusters. In terpene-class BGCs, which were distributed across the chromosome, disruption of the ori-proximal BGCs has been shown to have the greatest downstream impact on pigmentation [227].

What stands out about ori-proximal clusters is that the molecules they synthesise protect against environmental stressors – such as the NAGGN clusters which protect against dessication or pigmentation involved terpene clusters which protect against ultraviolet radiation. This may partially explain the differences between the two populations, and it is easy to hypothesise that organisms which form quiescent spores as part of their life cycle stand to benefit from being able to weather harsh abiotic factors. The accumulation and loss of ori-distal BGCs may then reflect transient usefulness in the evolutionary history of their hosts.

BGCs are constantly evolving genetic entities [89], Their sheer size and energetic costs of maintenance represent a considerable investment to the organisms that host them. Through the small molecules they encode generate fitness benefits to this host.

They also contribute to diversification of their hosts, to the point where differences in BGC content between two related organisms can predict interstrain antagonism [260]. Their maintenance depends on occupying a niche within that organism – a function that the molecule they encode fulfils. This has been demonstrated in siderophores in *Salinispora* where some strains have independently lost desferrioxamine biosynthesis in favour of salinichelins [261]. Therefore, BGCs must be under extraordinary selection pressure to maintain their existence. By migrating to the core of the *Micromonospora* chromosome, sharing space with essential genes in the chromosome [262], the core BGC suite in our organisms have become incorporated into the conserved core of the organisms and the protection against deletion that this core's existence implies. This strategy is not guaranteed to preserve the BGC, however, shown by the introduction of core thiopeptide biosynthetic enzymes into the T3PKS cluster of *M. echinofusca*.

5.3.4. From where BGCs are, to where they are not

What was conspicuous in the comparison between the BGCs of linear chromosomes and those of the single-contig *Micromonospora* analysed here was the absence BGCs at the chromosomal equator – halfway between the origin of replication and the ori-distal pole. The hyper-variable ori-distal region of the chromosomes seems to be the most likely site of BGC insertion into the chromosome, with BGCs migrating to the core over evolutionary history. An analogous phenomenon has been proposed in *Streptomyces* [89]. The question then stands of where are the BGCs between these two regions? Two hypotheses may explain this: firstly, BGC incorporation into the ori-proximal region from the ori-distal region is rare and happens rapidly, with useful BGCs spending little, if any time at the chromosomal equator. The other, more interesting hypothesis, is that incorporating BGCs into the equator is selected against, thus explaining the absence.

5.4. Summary

Going into this chapter, we set out to answer three questions. For the first, we have shown that *Micromonospora* do possess a conserved architecture. This takes the form of an ori-proximal core, with a hyper-variable region at the opposite pole. This pole likely is likely where termination of DNA replication occurs – it is where strand switching between gene density occurs, as well as the GC skew switching. For the second question, we have shown that some BGCs within the genus are conserved, namely terpene, NAGGN and T3PKS clusters. This lead us into the answer for the

third question, which is yes – chromosome architecture does impact conservation of BGCs and the core suite of BGCs exist as highly syntenic members of the larger *Micromonospora* chromosomal core. The hyper-variable region is populated by a diverse suite of BGCs, implying high turnover of these clusters. What remains is a question of what exactly drives the discrimination between the core BGCs of the genus and other genes, and how they migrate towards the chromosomal core.

Chapter 6 Summary and Future Perspectives

6.1. Elucidating the mechanism of *Micromonospora* antibiotic activity

The aim of this thesis was to discover novel molecules by exploiting organisms isolated from the hostile environment of Atacama Desert soil. Unfortunately, an actual molecule responsible for the observed antipseudomonad activity of *Micromonospora* sp. O3 and *Micromonospora* sp. O5 was never elucidated. Great strides have been made towards pairing tandem mass spectrometry data and genomic data, such as the curated MiBiG database of BGCs [123], the dereplication tools available on GNPS [263], and the Paired-Omics data platform which aims to make identification of linked genomic and chemical datasets trivial [264].

With a comparative metabolomics experiment, using extracts derived from culture grown on antibiotic inducing and non-inducing media, the molecule responsible for that antibiotic activity may be elucidated [137, 265]. As *Micromonospora* sp. O5 and *Micromonospora* sp. PH63 are so closely related and share a number of BGCs, it is tempting to assume that one of the BGCs present in *Micromonospora* sp. O5 but not in *Micromonospora* sp. PH63 is the driver of *Micromonospora* sp. O5's antibiotic activity . However, the different responses of the two strains to nutrient rich media (with *Micromonospora* sp. PH63 displaying suppressed development where *Micromonospora* sp. O5 did not) means that we can not rule out the lack of activity as a product of differences in regulation rather than BGC content. Work in *Salinispora pacifica* highlights this issue, showing that even closely related strains express BGCs differentially owing to differences in regulation [266]. Thus, proper analytical chemistry is key to resolving this outstanding question. Should it be the case that *Micromonospora* sp. O5's bioactivity is derived from a molecule that it and *Micromonospora* sp. PH63 are both capable of producing, the outstanding question is what differences between the two are driving this observation?

All three strains contain BGCs which are

With an – ideally novel – candidate metabolite and cognate BGC identified, generation of a null or silenced strains of *Micromonospora* sp. O5 and *Micromonospora* sp. O3 and screening them for bioactivity and metabolite production would allow for confirmation of that metabolite and BGC's role in observed antibiotic activity. *Micromonospora* are notably genetically intractable organisms, and although method development is ongoing to develop genetic tools for them – success has been seen by employing CRISPR [102, 267], and through the development of phasmids [57]. Whilst still lagging behind the developments in *Streptomyces* genetics, it is very

possible that advances such as CRISPRi silencing and CRISPR-BEST gene editing [268] will soon be translated to other actinomycete families. Failing that, heterologous expression of the candidate BGC is also an option [269].

6.2. *Streptomyces* and *Micromonospora*: Same, but different

Highlighting the fundamental differences between the biology of *Streptomyces* and *Micromonospora* biology is the lack of conserved transcriptional regulators we observed between our strains and *S. coelicolor*. We chose these regulators as their expression and regulation covers the transition from vegetative growth to completion of sporulation, and coincident antibiotic production [72]. The only universally conserved elements were BldD, the master repressor of development [77, 78, 81], and WhiA, a regulator of spore maturation [206]. *Micromonospora* sp. O5 and *Micromonospora* sp. PH63 also possessed an ortholog of WhiB, whilst the closest hit from *Micromonospora* sp. O3 was a WhiB-like regulator, however, is unlikely to be an ortholog of the *S. coelicolor* WhiB. These results are concordant with previous work [270].

The events underpinning the developmental transition from vegetative to spore-bearing hyphae remain to be elucidated, and likely benefit from a spore-to-spore longitudinal study [128, 271]. Spore development in *Streptomyces* is characterized by simultaneous chromosome division and segregation, marking the transition from multinucleate aerial hyphae into unicellular spores [49, 75]. Whether or not the *Micromonospora* possess a developmental stage analogous to aerial hypha remains to be seen, however the dense clusters of spore bearing hyphae we observed microscopically suggest some level of coordination of sporulation exists. The application of bioinformatic tools such as OrthoFinder [272] might expand our understanding of the evolution of these organisms without relying on *a priori* knowledge of the orthologs to look for. Actinomycete biology has certainly benefited from the relative ease of working with *Streptomyces* compared to other organisms, however looking at actinomycetes through a lens of model *Streptomyces* biology may also limit our understanding of them.

6.3. The Potential of BldA as an Elicitor of Secondary Metabolism

Despite these limits, we can exploit knowledge garnered from *Streptomyces,* such as that of the function of BldA [239, 273]. Notable in all three of our *Micromonospora*

strains was the four bp deletion in the *bldA* gene, predicted to transcribe a non-functional tRNA. As previously discussed – developmental regulation of the expression of BldA in turn regulates all the TTA encoded genes of an organism by inhibiting their translation. Thus, it remains to be seen if, and how, these three *Micromonospora* handle TTA encoding genes. GFP reporters for BldA mediated translation have been successfully utilised in *S. lividans* [274]. This reporter could likewise be exploited to investigated whether or not TTA-mediated translation is occurring in the Chilean *Micromonospora* or not. If not, complementation with a functional copy of *bldA*, such as that of *S. coelicolor*, may lead to the production of novel metabolites [242].

Micromonospora sp. PH63 and *Micromonospora* sp. O5 are divergent from *Micromonospora* sp. O3, and this may be an indicator of a truncated BldA throughout the genus – if this were the case, it may represent an untapped wealth of secondary metabolites. It may be the case that there are other, as yet undiscovered mediators of secondary metabolism that could be exploited for intergenic metabolite elicitation.

6.4. Exploiting Genomics to Accelerate Drug Discovery

In addition to exploiting known genetics to elicit cryptic metabolites, we can also use the accumulation of sequence information in public databases to expand our understanding of secondary metabolism in actinomycetes. The work here has shown that high-quality, single contig assemblies of *Micromonospora* can be used to show the conserved loci of some BGCs. What remains to be seen is the extent of this conservation in other actinomycetes, and other families of BGC-rich bacteria. Expanding the number and diversity of well-sequenced bacterial genomes is key to this [247], with *Streptomyces* being the most covered group of actinomycete[171]. As important as expanding the number of high quality publicly available sequence data that exists is to improve the taxonomy of actinomycetes and improve our understanding of how phylogeny facilitates or limits horizontal gene transfer [275]. For example, *M. pataloongenis*, despite being included in the autoMLST tree which *Micromonospora* sp. O3, *Micromonospora* sp. O5, and *Micromonospora sp* PH63 were placed in, had a lower proportion of fragments mapped from and to its genome by FastANI than the rest of the *Micromonospora* in the tree. More robust methods of calculating ANI such as ANIm and ANIb [276] may better elucidate if the genus *Micromonospora* is a genus *sensu stricto* or in fact multiple closely related genera in a trenchcoat.

The logical next group to examine are the rest of the members of the *Micromonosporaceae*, and beyond that into the *Pseudonocardiaceae* [51]. One question that remains outstanding is what drives the difference we observed between the *Streptomyces* and *Micromonospora*, and in particular the *oriC* proximal suite of BGCs present in only *Micromonospora* Chromosome capture studies of *Streptomyces* have shown multiple conformational changes in the linear chromosome over the course of the organisms development [277, 278] – an analogous process may be driving the segregation of BGCs to different regions of the *Micromonospora* chromosome. It may be the differences in BGC distribution that we see are owed to the linearity of *Streptomcyes* chromosomes versus the circularity of *Micromonospora*.

6.5. The NAGGN Cluster

The NAGGN cluster present in the *Micromonospora* serves to biosynthesise compatible solutes in a number of Gram-negative bacteria [109]. Its presence and conserved location in the *Micromonospora* we analysed as well as its presence in the related genus *Salinispora* [279] but not other *Micromonosporaceae* suggest its acquisition by horizontal gene transfer, although when that is predicted to have occurred remains to be elucidated. Intergeneric conjunction into actinomycetes is frequently exploited in the laboratory environment [280], and so its occurrence in nature is not beyond the realm of possibility.

6.6. Concluding Remarks

This thesis lays bare the potential for comparative genomics to accelerate our understanding of bacterial natural products. It also highlights the advantages provided by high-quality sequencing data in terms of tracking the chromosomal locus of biosynthetic gene clusters. By expanding the diversity of actinomycetes that we sequence and accurately cataloguing both their biosynthetic capabilities and their phylogenetic history.

Chapter 7 Appendices

Appendix 1: ANI values of Chilean *Micromonospora* and their closest related strains

Table 7.1 ANI of *Micromonospora* sp. O3, O5, and PH63 against their phylogenetic neighbours

Reference Organism	Query Organism: M. sp. O3	M. sp. O5	M. sp. PH63
M. aurantiaca ATCC 27029	84.8016	84.9594	85.0382
M. aurantinigra DSM 44815	85.6888	85.2283	85.4061
M. avicenniae DSM 45758	84.9711	84.6282	84.4682
M. carbonaceae DSM 43168	85.0545	85.3005	85.3595
M. chaiyaphumensis DSM 45246	85.4583	85.1726	85.3188
M. chersina DSM 44151	85.5383	85.1969	85.5208
M. chokoriensis DSM 45160	86.8963	84.486	84.6621
M. citrea DSM 43903	86.1457	90.185	90.345
M. coriariae DSM 44875	88.9084	85.4706	85.594
M. coxensis DSM 45161	85.9136	86.0502	86.3275
M. cremea DSM 45599	89.0006	85.4254	85.7586
M. eburnea DSM 44814	84.8435	84.6885	84.7561
M. echinaurantiaca DSM 43904	86.5976	87.232	87.3898
M. echinofusca DSM 43913	85.9542	93.1022	93.4664
M. echinospora DSM 43816	83.8018	84.6563	84.5965
M. globosa NRRL B-2673	84.8384	84.9444	85.0837
M. haikouensis DSM 45626	85.097	85.288	85.6921
M. halophytica DSM 43171	85.5081	85.8152	85.9942
M. humi DSM 45647	84.9848	84.6047	84.7906
M. inositola DSM 43819	86.1906	85.7564	85.8609
M. inyonesnsis DSM 46123	83.7347	84.2489	84.0628
M. krabiensis DSM 45344	85.7132	85.1568	85.3506
M. marina DSM 45555	84.6216	84.7883	84.9941
M. matsumotoense DSM 44100	84.6698	84.3661	84.4988
M. mirobrigensis DSM 44830	85.6795	85.3424	85.4845
M. narathiwatensis DSM 45248	85.129	84.9705	85.05
M. nigra DSM 43818	84.4409	85.1549	85.2207
M. pallida DSM 43817	83.7841	84.3764	84.3866
M. pattaloongensis DSM 45245	81.6175	81.4905	81.58
M. peucetia DSM 43363	85.6383	92.3479	92.4293
M. phaseoli CGMCC 47038	83.8879	84.2792	84.2248
M. purpureochromogenes DSM 43821	86.3136	86.0427	86.1573
M. purpureochromogenes NRRL B-2672	84.803	84.9331	85.048
M. rhizosphaerae DSM 45431	85.6519	85.3492	85.5634
M. rifamycinica AM105	84.6927	84.3335	84.7422
M. rifamycinica DSM 44983	84.811	84.4038	84.752
M. rosaria DSM 803	84.17	84.409	84.4604
M. saelicesensis DSM 44871	87.4468	84.6093	84.7834
M. siamensis DSM 45097	85.6049	85.1712	85.2948
M. sp. II	84.6956	84.939	85.1087
M. sp. L5	84.7353	84.9319	85.1206
M. sp. NBRC 110037	84.7553	84.9545	85.1353
M. sp. O3	100	85.879	85.9378
M. sp. O5	85.9201	100	95.4018
M. sp. PH63	86.1611	95.3159	100
M. sp. TSRI0369	84.7411	84.9639	85.0945
M. tulbaghiae DSM 45142	84.7335	84.5209	84.7146
M. viridifaciens DSM 43909	84.9198	84.8911	84.9244
M. wenchangensis CCTCC AA 2012002	84.9776	84.5599	84.6898
M. yangpuensis DSM 45577	83.9692	84.061	83.9531
M. zamorensis DSM 45600	87.4603	84.5925	84.8142

Appendix 2: Alignments of *S. coelicolor* developmental regulatory proteins and best BLASTp Hits

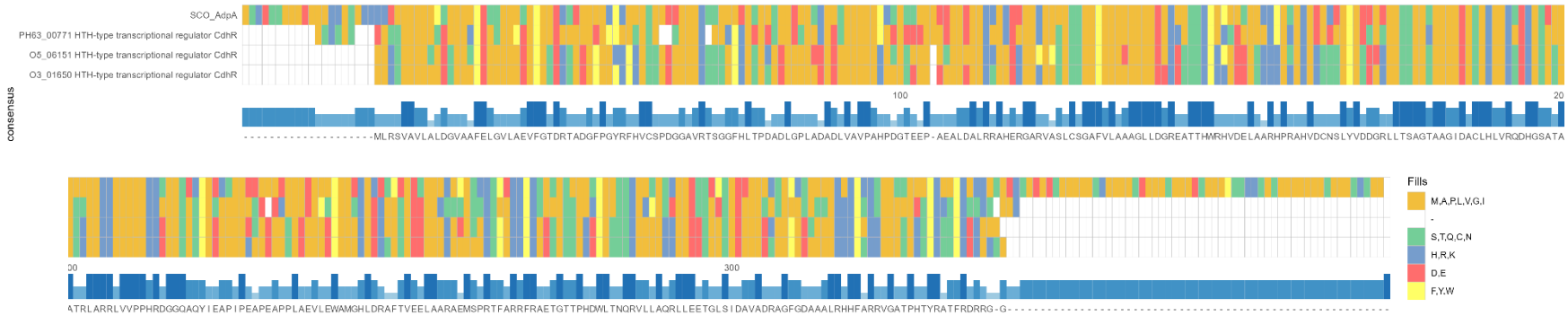


Figure 7-1 A. Alignment of *S. coelicolor* AdpA and closest related proteins

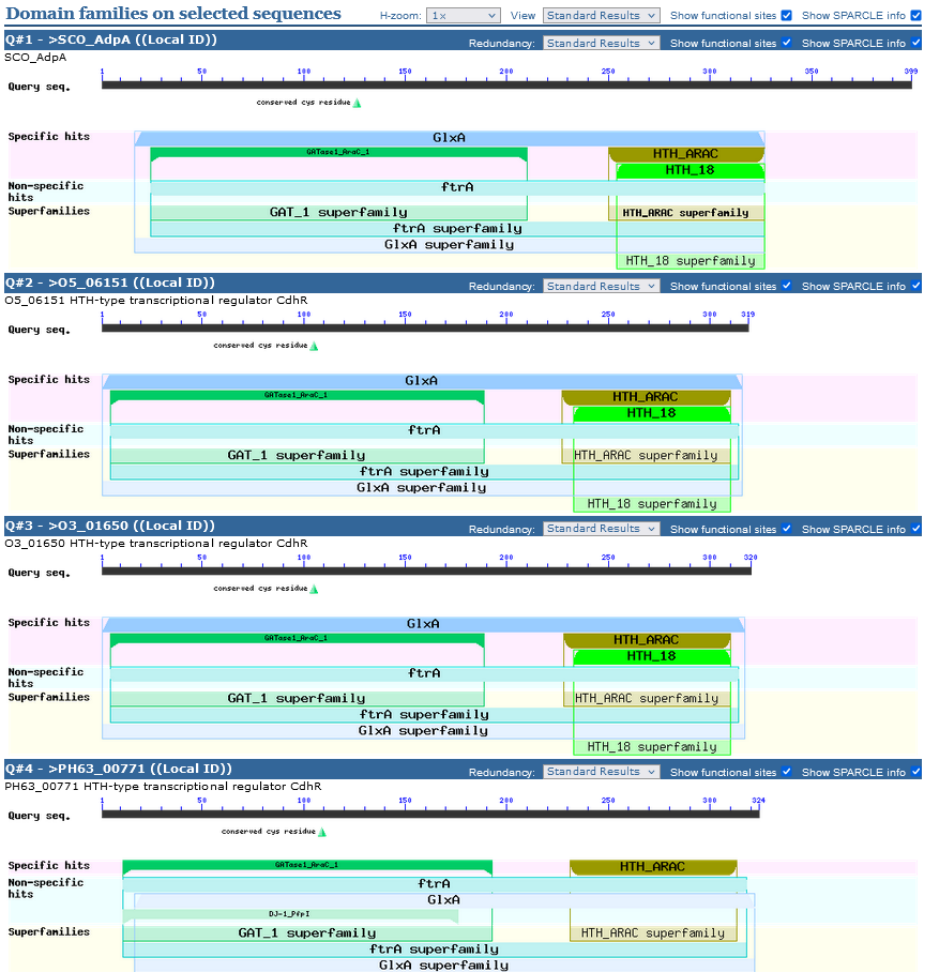


Figure 7-2. Conserved domains of *S. coelicolor* AdpA and closest related proteins

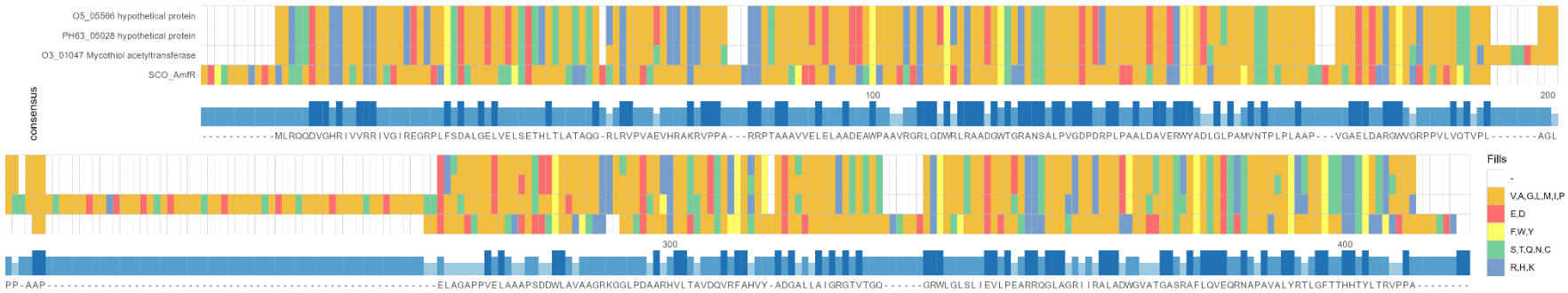


Figure 7-3. Alignment of AmfR and closest related proteins

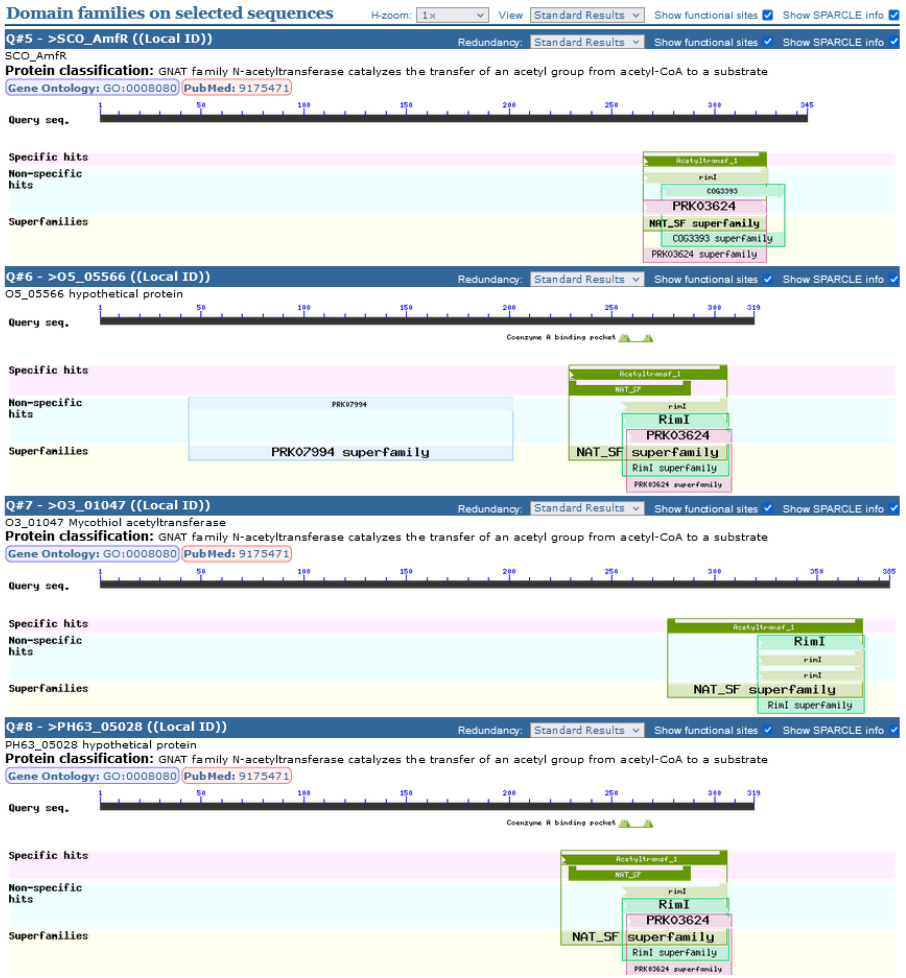


Figure 7-4. Conserved domains of *S. coelicolor* AmfR and closest related proteins

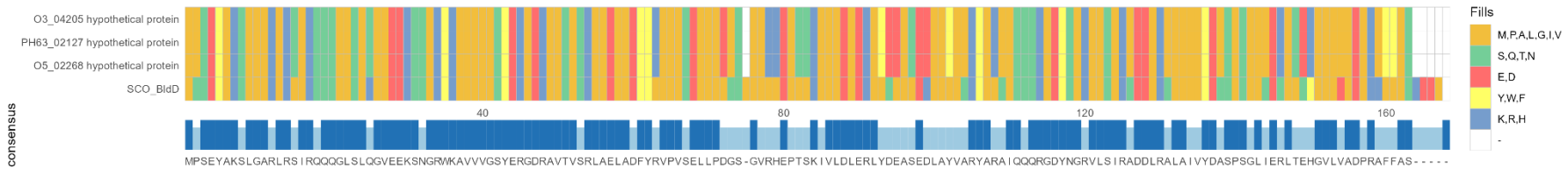


Figure 7-5 Alignment of BldD and closest related proteins

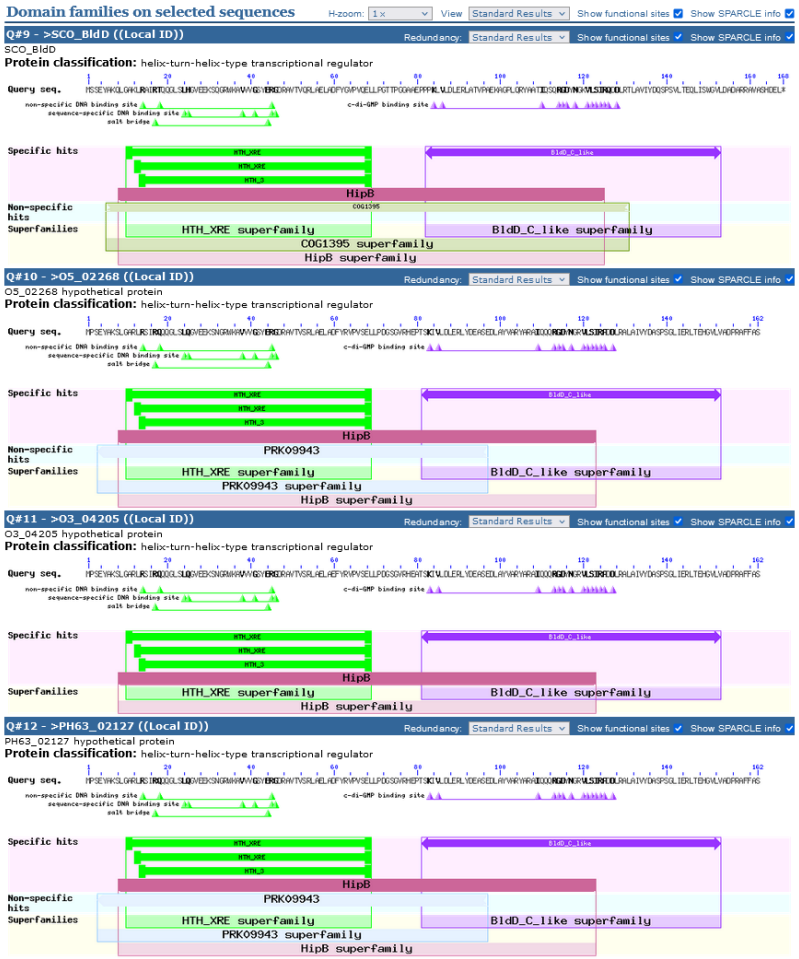


Figure 7-6 Conserved domains of *S. coelicolor* BldD and closest related proteins

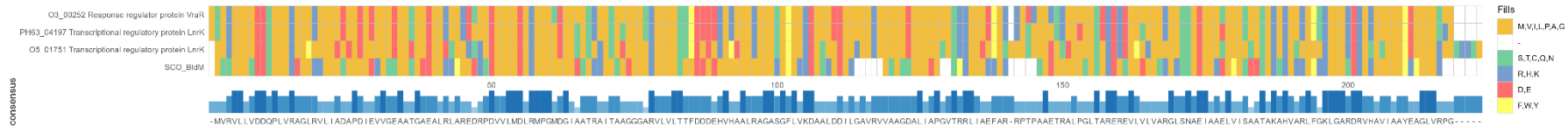


Figure 7-7 Alignment of BldM and closest related proteins

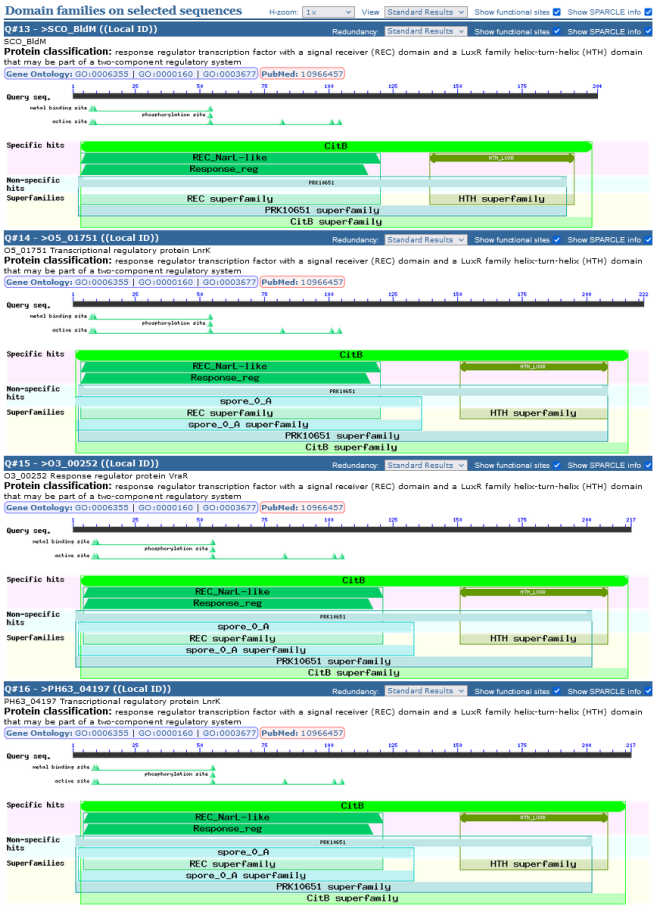


Figure 7-8 Conserved domains of *S. coelicolor* BldM and closest related proteins

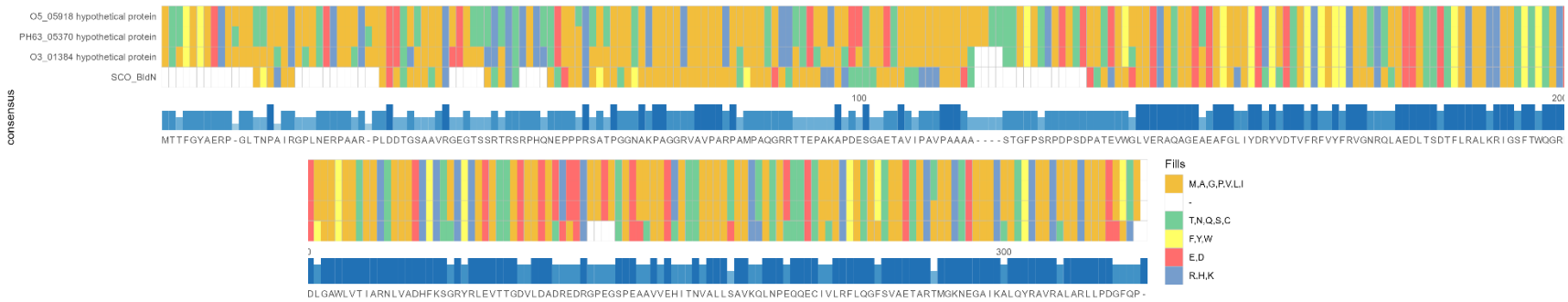


Figure 7-9 Alignment of BldN and closest related proteins

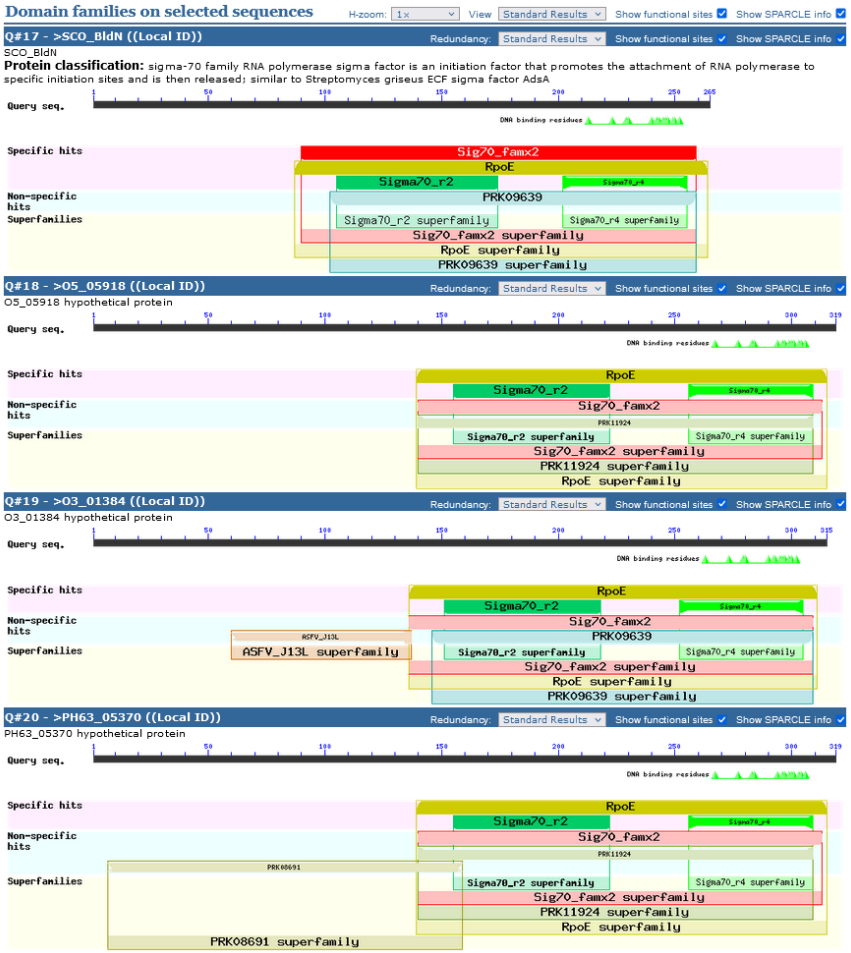


Figure 7-10 Conserved domains of *S. coelicolor* BldN and closest related proteins

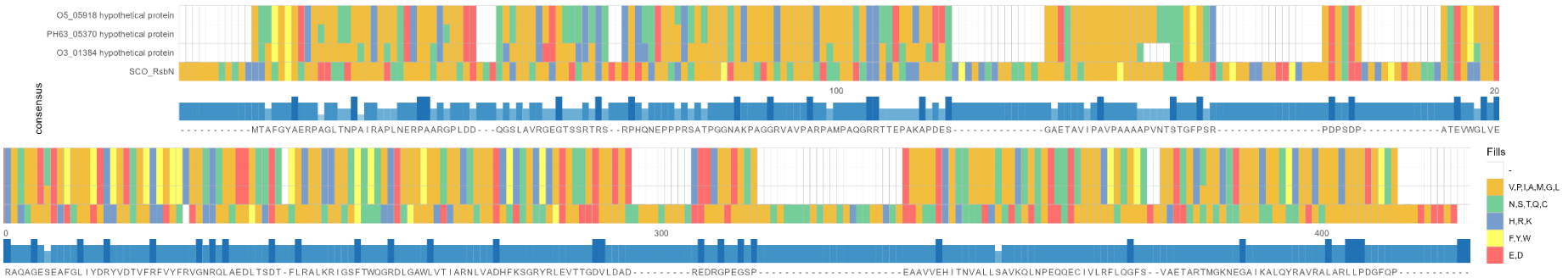


Figure 7-11 Alignment of RsbN and closest related proteins

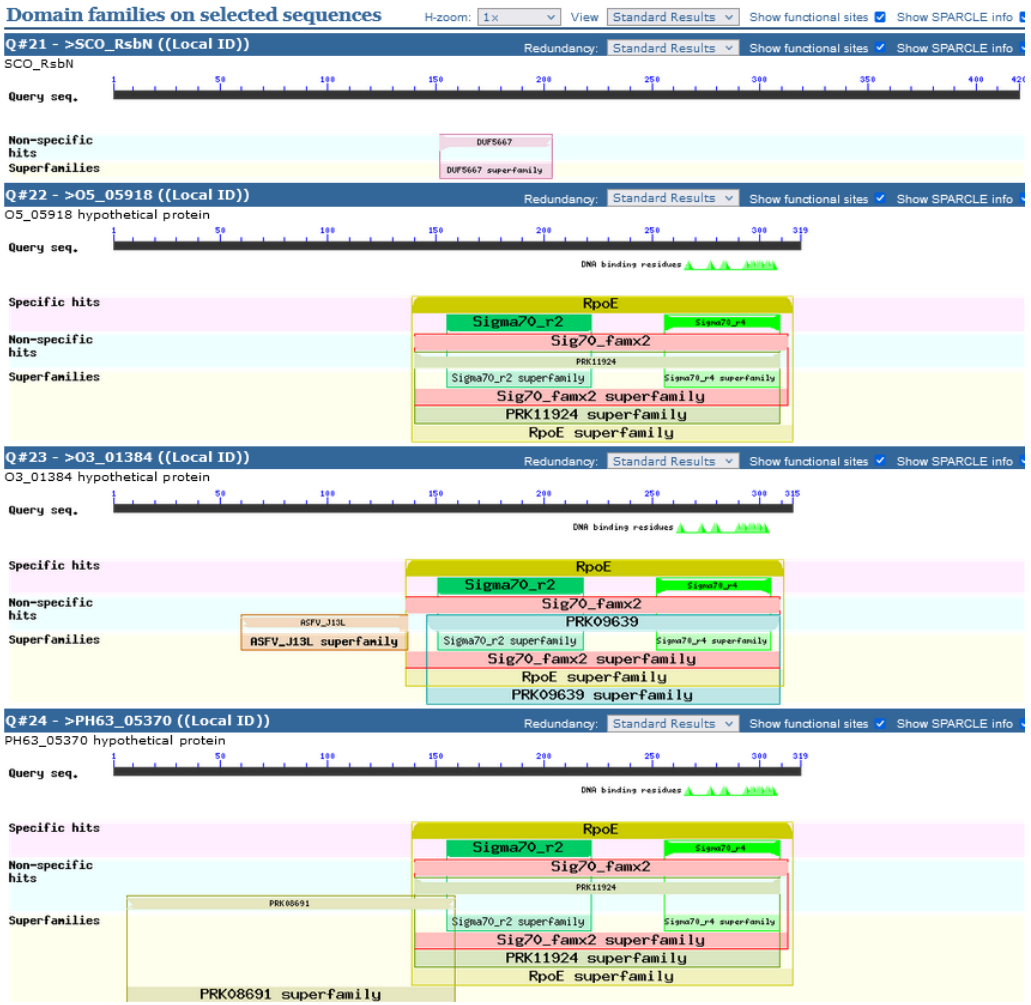


Figure 7-12 Conserved Domains of RsbN and closest related proteins

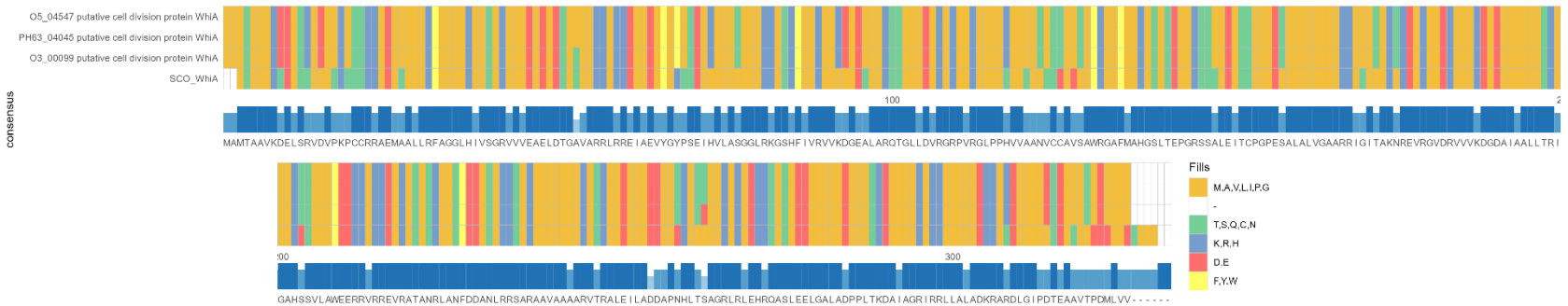


Figure 7-13 Alignment of WhiA and closest related proteins

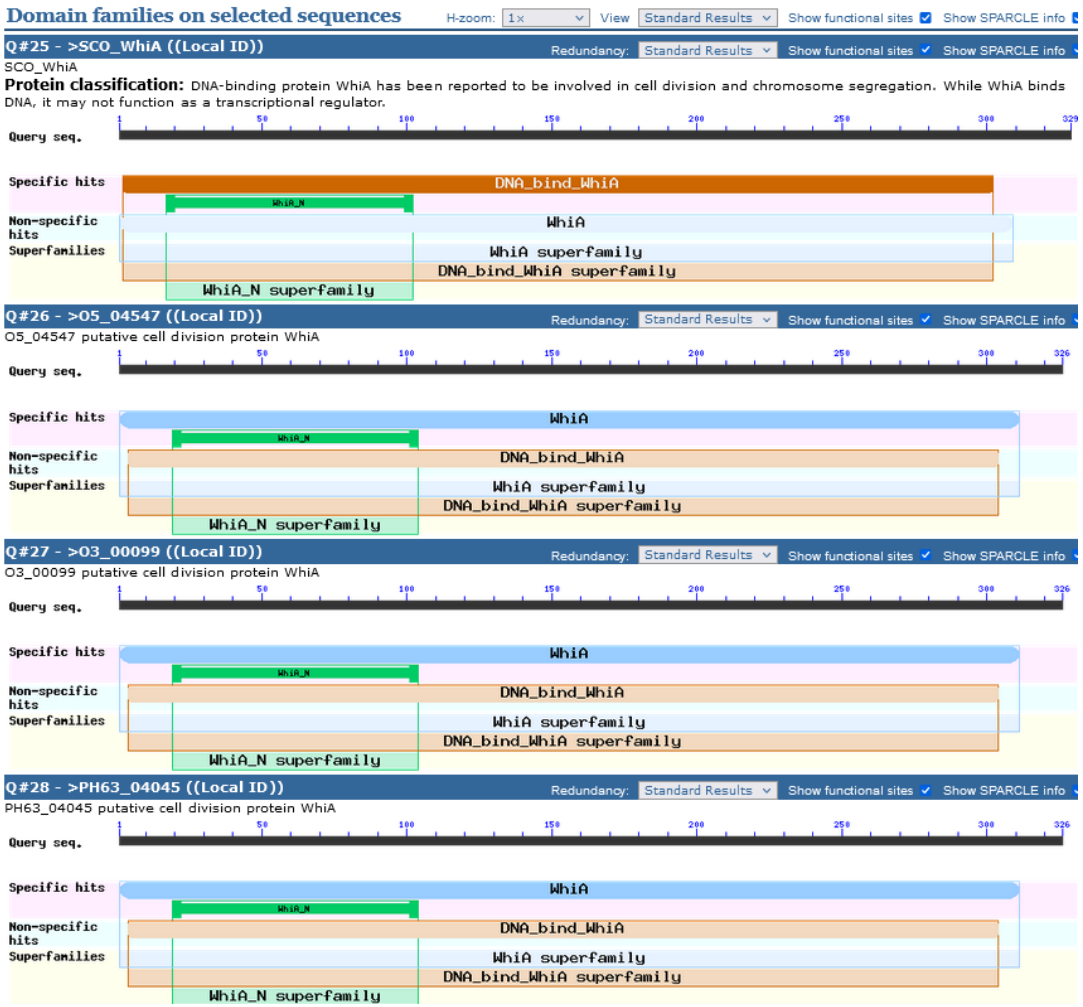


Figure 7-14 Conserved domains of WhiA and closest related proteins

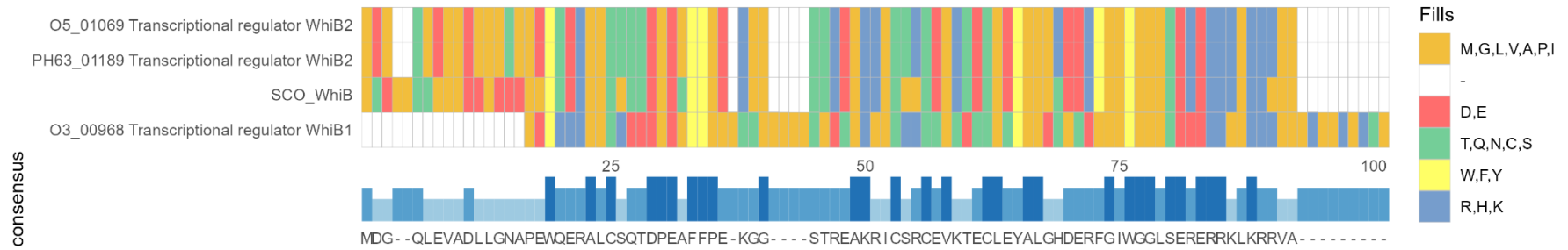


Figure 7-15 Alignment of WhiB and closest related proteins

Figure 7-16 Conserved Domains of WhiB and closest related proteins

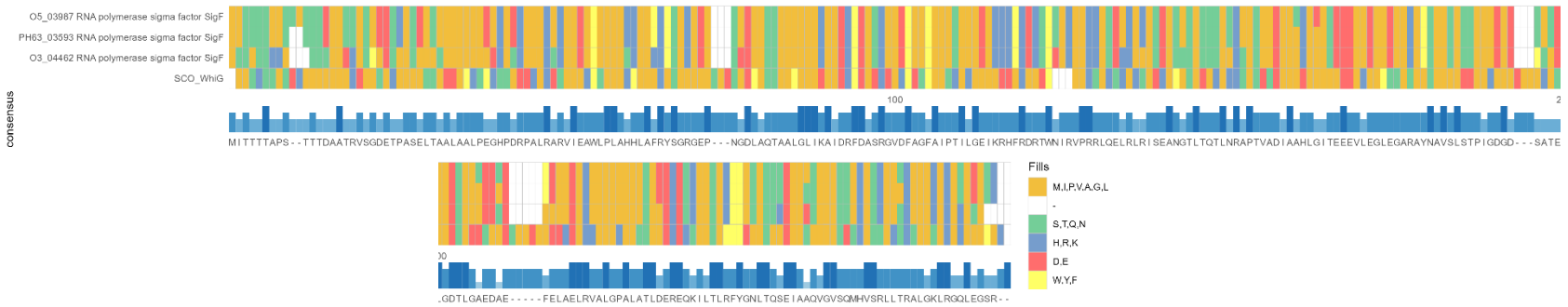


Figure 7-17 Alignment of WhiG and closest related proteins

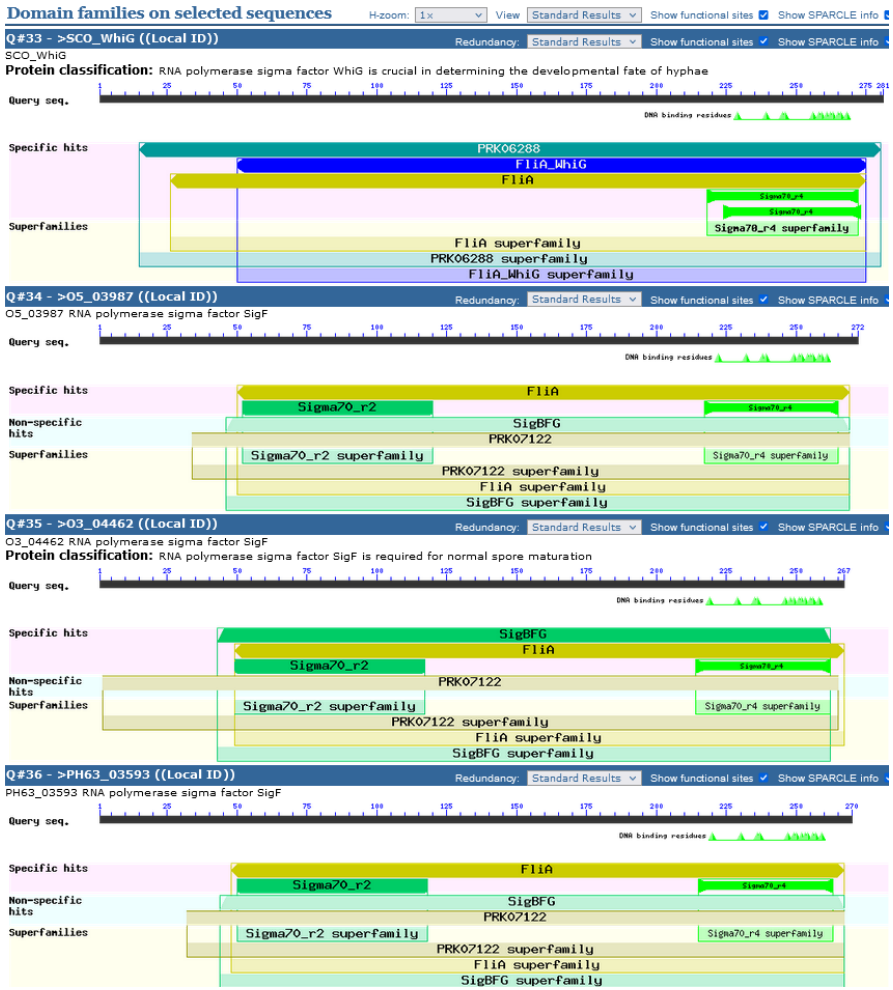


Figure 7-18 Conserved Domains of WhiG and closest related proteins

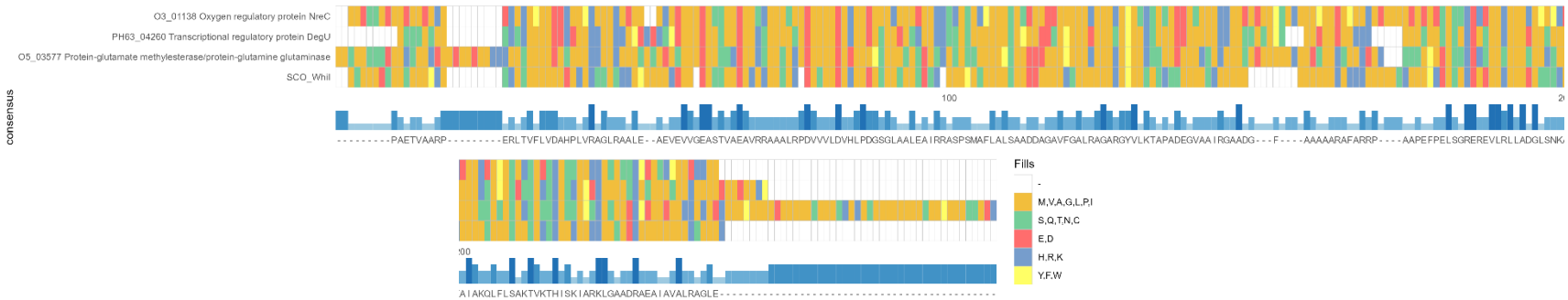


Figure 7-19 Alignment of WhiI and closest related proteins

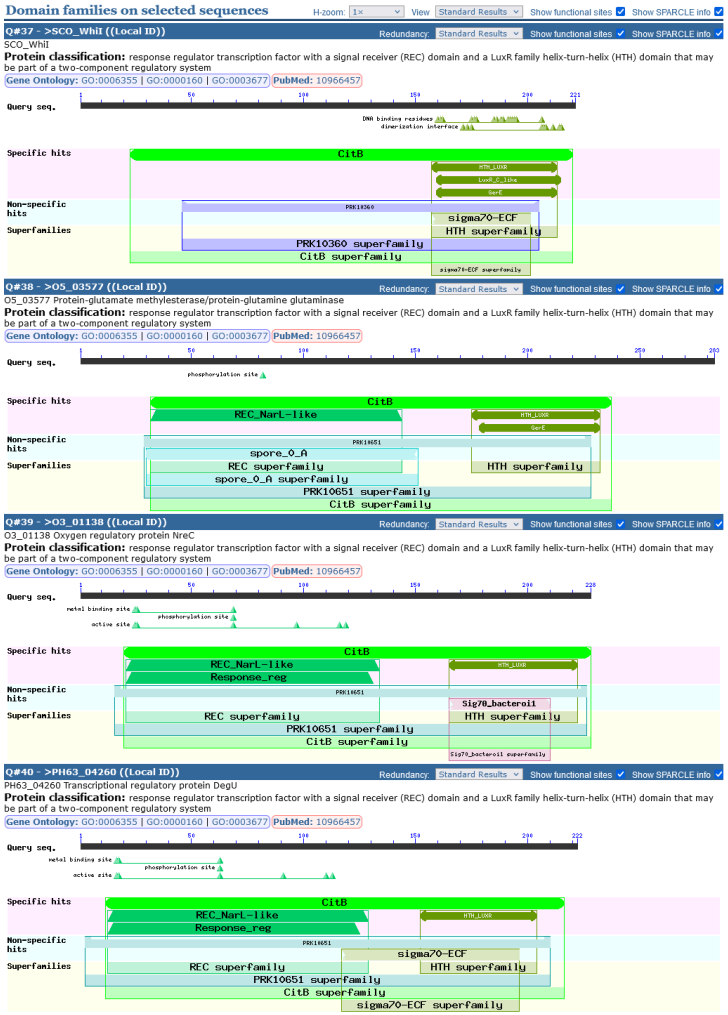


Figure 7-20 Conserved Domains of WhiI and closest related proteins

Chapter 8 References

1. O'Neill, J., *Tackling Drug-Resistant Infections Globally: Final Report and Recommendations.* 2016.
2. Larsen, J., et al., *Emergence of methicillin resistance predates the clinical use of antibiotics.* Nature, 2022.
3. Van Boeckel, T.P., et al., *Global antibiotic consumption 2000 to 2010: an analysis of national pharmaceutical sales data.* The Lancet Infectious Diseases, 2014. **14**(8): p. 742-750.
4. Davies, J. and D. Davies, *Origins and evolution of antibiotic resistance.* Microbiol Mol Biol Rev, 2010. **74**(3): p. 417-33.
5. Wright, G.D., *Bacterial resistance to antibiotics: enzymatic degradation and modification.* Adv Drug Deliv Rev, 2005. **57**(10): p. 1451-70.
6. Fishovitz, J., et al., *Penicillin-binding protein 2a of methicillin-resistant Staphylococcus aureus.* IUBMB Life, 2014. **66**(8): p. 572-7.
7. Jarlier, V. and H. Nikaido, *Mycobacterial cell wall: Structure and role in natural resistance to antibiotics.* FEMS Microbiology Letters, 1994. **123**: p. 11 - 18.
8. Yoneyama, H. and N. Taiji, *Mechanism of Efficient Elimination of Protein D2 in Outer Membrane of Imipenem-Resistant Pseudomonas aeruginosa.* Antimicrob Agents Chemother, 1993. **37**(11): p. 2385-2390.
9. Ochs, M.M., et al., *Negative Regulation of the Pseudomonas aeruginosa Outer Membrane Porin OprD Selective for Imipenem and Basic Amino Acids.* Antimicrob Agents Chemother, 1999. **43**: p. 1085-1090.
10. Li, H., et al., *Structure and function of OprD protein in Pseudomonas aeruginosa: from antibiotic resistance to novel therapies.* Int J Med Microbiol, 2012. **302**(2): p. 63-8.
11. Gill, M.J., et al., *Gonococcal Resistance to b-Lactams and Tetracycline Involves Mutation in Loop 3 of the Porin Encoded at the penB Locus.* Antimicrob Agents Chemother, 1998. **42**(11): p. 2799-2803.
12. Rabin, N., et al., *Biofilm formation mechanisms and targets for developing antibiofilm agents.* Future Medicinal Chemistry, 2015. **4**(7): p. 493-512.
13. Lebeaux, D., J.M. Ghigo, and C. Beloin, *Biofilm-related infections: bridging the gap between clinical management and fundamental aspects of recalcitrance toward antibiotics.* Microbiol Mol Biol Rev, 2014. **78**(3): p. 510-43.
14. Jefferson, K.K., D.A. Goldmann, and G.B. Pier, *Use of confocal microscopy to analyze the rate of vancomycin penetration through Staphylococcus aureus biofilms.* Antimicrob Agents Chemother, 2005. **49**(6): p. 2467-73.
15. Long, K.S., et al., *The Cfr rRNA methyltransferase confers resistance to Phenicols, Lincosamides, Oxazolidinones, Pleuromutilins, and Streptogramin A antibiotics.* Antimicrob Agents Chemother, 2006. **50**(7): p. 2500-5.
16. Tipper, D.J. and J.L. Strominger, *Mechanism of action of penicillins: a proposal based on their structural similarity to acyl-D-alanyl-D-alanine.* Proc Natl Acad Sci U S A., 1965. **54**(4): p. 1133-1141.
17. Bush, K., G.A. Jacoby, and A.A. Medeiros, *A Functional Classification Scheme for β-Lactamases and Its Correlation with Molecular Structure.* Antimicrobial Agents and Chemotherapy, 1995. **39**(6): p. 1211-1233.
18. Bush, K. and G.A. Jacoby, *Updated functional classification of beta-lactamases.* Antimicrob Agents Chemother, 2010. **54**(3): p. 969-76.

19. Peix, A., M.H. Ramírez-Bahena, and E. Velázquez, *The current status on the taxonomy of Pseudomonas revisited: An update.* Infect Genet Evol, 2018. **57**: p. 106-116.
20. Xin, X.F., B. Kvitko, and S.Y. He, *Pseudomonas syringae: what it takes to be a pathogen.* Nat Rev Microbiol, 2018. **16**(5): p. 316-328.
21. Gellatly, S.L. and R.E. Hancock, *Pseudomonas aeruginosa: new insights into pathogenesis and host defenses.* Pathog Dis, 2013. **67**(3): p. 159-73.
22. Lee, J. and L. Zhang, *The hierarchy quorum sensing network in Pseudomonas aeruginosa.* Protein Cell, 2015. **6**(1): p. 26-41.
23. Siryaporn, A., et al., *Surface attachment induces Pseudomonas aeruginosa virulence.* Proc Natl Acad Sci U S A, 2014. **111**(47): p. 16860-5.
24. Strempel, N., et al., *Human host defense peptide LL-37 stimulates virulence factor production and adaptive resistance in Pseudomonas aeruginosa.* PLoS One, 2013. **8**(12): p. e82240.
25. Galle, M., I. Carpentier, and R. Beyaert, *Structure and function of the Type III secretion system of Pseudomonas aeruginosa.* Curr Protein Pept Sci, 2012. **13**(8): p. 831-42.
26. Gallagher, L.A. and C. Manoil, *Pseudomonas aeruginosa PAO1 kills Caenorhabditis elegans by cyanide poisoning.* J Bacteriol, 2001. **183**(21): p. 6207-14.
27. Lenney, W. and F.J. Gilchrist, *Pseudomonas aeruginosa and cyanide production.* Eur Respir J, 2011. **37**(3): p. 482-3.
28. Kamath, S., V. Kapatral, and A.M. Chakrabarty, *Cellular function of elastase in Pseudomonas aeruginosa: role in the cleavage of nucleoside diphosphate kinase and in alginate synthesis.* Mol Microbiol, 1998. **30**(5): p. 933-41.
29. Cornelis, P. and J. Dingemans, *Pseudomonas aeruginosa adapts its iron uptake strategies in function of the type of infections.* Front Cell Infect Microbiol, 2013. **3**: p. 75.
30. Boucher, J.C., et al., *Mucoid Pseudomonas aeruginosa in cystic fibrosis: characterization of muc mutations in clinical isolates and analysis of clearance in a mouse model of respiratory infection.* Infect Immun, 1997. **65**(9): p. 3838-46.
31. Mann, E.E. and D.J. Wozniak, *Pseudomonas biofilm matrix composition and niche biology.* FEMS Microbiol Rev, 2012. **36**(4): p. 893-916.
32. Fuqua, W.C., S.C. Winans, and E.P. Greenberg, *Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators.* J Bacteriol, 1994. **176**(2): p. 269-75.
33. de Kievit, T.R., *Quorum sensing in Pseudomonas aeruginosa biofilms.* Environ Microbiol, 2009. **11**(2): p. 279-88.
34. D'Argenio, D.A., et al., *Growth phenotypes of Pseudomonas aeruginosa lasR mutants adapted to the airways of cystic fibrosis patients.* Mol Microbiol, 2007. **64**(2): p. 512-33.
35. Bjarnsholt, T., et al., *Quorum sensing and virulence of Pseudomonas aeruginosa during lung infection of cystic fibrosis patients.* PLoS One, 2010. **5**(4): p. e10115.
36. Subedi, D., et al., *Comparative genomics of clinical strains of Pseudomonas aeruginosa strains isolated from different geographic sites.* Sci Rep, 2018. **8**(1): p. 15668.
37. Folkesson, A., et al., *Adaptation of Pseudomonas aeruginosa to the cystic fibrosis airway: an evolutionary perspective.* Nat Rev Microbiol, 2012. **10**(12): p. 841-51.
38. Young, R.L., et al., *Neutrophil extracellular trap (NET)-mediated killing of Pseudomonas aeruginosa: evidence of acquired resistance within the CF airway, independent of CFTR.* PLoS One, 2011. **6**(9): p. e23637.

39. Bragonzi, A., et al., *Pseudomonas aeruginosa microevolution during cystic fibrosis lung infection establishes clones with adapted virulence.* Am J Respir Crit Care Med, 2009. **180**(2): p. 138-45.
40. Candido Caçador, N., et al., *Adaptation of Pseudomonas aeruginosa to the chronic phenotype by mutations in the algTmucABD operon in isolates from Brazilian cystic fibrosis patients.* PLoS One, 2018. **13**(11): p. e0208013.
41. Lister, P.D., D.J. Wolter, and N.D. Hanson, *Antibacterial-resistant Pseudomonas aeruginosa: clinical impact and complex regulation of chromosomally encoded resistance mechanisms.* Clin Microbiol Rev, 2009. **22**(4): p. 582-610.
42. Bassetti, M., et al., *How to manage Pseudomonas infections.* Drugs Context, 2018. **7**: p. 212527.
43. Fruci, M. and K. Poole, *Aminoglycoside-inducible expression of the mexAB-oprM multidrug efflux operon in Pseudomonas aeruginosa: Involvement of the envelope stress-responsive AmgRS two-component system.* PLoS One, 2018. **13**(10): p. e0205036.
44. Li, X.Z., L. Zhang, and K. Poole, *Role of the multidrug efflux systems of Pseudomonas aeruginosa in organic solvent tolerance.* J Bacteriol, 1998. **180**(11): p. 2987-91.
45. *Global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics*. 2017.
46. Ganesan, A., *The impact of natural products upon modern drug discovery.* Curr Opin Chem Biol, 2008. **12**(3): p. 306-17.
47. Watve, M.G., et al., *How many antibiotics are produced by the genus Streptomyces?* Arch Microbiol, 2001. **176**(5): p. 386-90.
48. Dyson, P., *Streptomyces : molecular biology and biotechnology*. 2011, Norfolk, UK: Caister Academic Press. xii, 257 p.
49. Barka, E.A., et al., *Taxonomy, Physiology, and Natural Products of Actinobacteria.* Microbiol Mol Biol Rev, 2016. **80**(1): p. 1-43.
50. Lewin, G.R., et al., *Evolution and Ecology of Actinobacteria and Their Bioenergy Applications.* Annu Rev Microbiol, 2016. **70**: p. 235-54.
51. Nouioui, I., et al., *Genome-Based Taxonomic Classification of the Phylum Actinobacteria.* Front Microbiol, 2018. **9**: p. 2007.
52. Carro, L., et al., *Genome-based classification of micromonosporae with a focus on their biotechnological and ecological potential.* Sci Rep, 2018. **8**(1): p. 525.
53. Redenbach, M., J. Scheel, and U. Schmidt, *Chromosome topology and genome size of selected actinomycetes species.* Antonie Van Leeuwenhoek, 2000. **78**(3-4): p. 227-35.
54. Cui, Y., et al., *Genome Sequence of Micromonospora terminaliae TMS7T, a New Endophytic Actinobacterium Isolated from the Medicinal Plant Terminalia mucronata.* Mol Plant Microbe Interact, 2020. **33**(5): p. 721-723.
55. Trujillo, M.E., et al., *Genome features of the endophytic actinobacterium Micromonospora lupini strain Lupac 08: on the process of adaptation to an endophytic life style?* PLoS One, 2014. **9**(9): p. e108522.
56. Oshida, T., et al., *Isolation and characterization of plasmids from Micromonospora zionensis and Micromonospora rosaria.* Plasmid, 1986. **16**(1): p. 74-6.
57. Li, X., X. Zhou, and Z. Deng, *Isolation and characterization of Micromonospora phage PhiHAU8 and development into a phasmid.* Appl Environ Microbiol, 2004. **70**(7): p. 3893-7.
58. Carro, L., et al., *Uncovering the potential of novel micromonosporae isolated from an extreme hyper-arid Atacama Desert soil.* Sci Rep, 2019. **9**(1): p. 4678.

59. Abdel-Mageed, W.M., et al., *Biotechnological and Ecological Potential of Micromonospora provocatoris sp. nov., a Gifted Strain Isolated from the Challenger Deep of the Mariana Trench.* Marine Drugs, 2021. **19**(5).
60. S Hifnawy, M., et al., *Induction of Antibacterial Metabolites by Co-Cultivation of Two Red-Sea-Sponge-Associated Actinomycetes.* Mar Drugs, 2020. **18**(5).
61. Fu, G., et al., *Micromonospora zhangzhouensis sp. nov., a Novel Actinobacterium Isolated from Mangrove Soil, Exerts a Cytotoxic Activity in vitro.* Sci Rep, 2020. **10**(1): p. 3889.
62. Trujillo, M.E., et al., *Micromonospora lupini sp. nov. and Micromonospora saelicesensis sp. nov., isolated from root nodules of Lupinus angustifolius.* Int J Syst Evol Microbiol, 2007. **57**(Pt 12): p. 2799-804.
63. Ni, X., et al., *Genetic engineering combined with random mutagenesis to enhance G418 production in Micromonospora echinospora.* J Ind Microbiol Biotechnol, 2014. **41**(9): p. 1383-90.
64. Van Lanen, S.G. and B. Shen, *Biosynthesis of enediyne antitumor antibiotics.* Curr Top Med Chem, 2008. **8**(6): p. 448-59.
65. Hirsch, A.M. and M. Valdés, *Micromonospora: An important microbe for biomedicine and potentially for biocontrol and biofuels.* Soil Biology and Biochemistry, 2010. **42**(4): p. 536-542.
66. Wang, B., et al., *Detection of functional microorganisms in benzene [a] pyrene-contaminated soils using DNA-SIP technology.* Journal of Hazardous Materials, 2021. **407**: p. 124788.
67. Ortúzar, M., et al., *Micromonospora metallophores: A plant growth promotion trait useful for bacterial-assisted phytoremediation?* Sci Total Environ, 2020. **739**: p. 139850.
68. Suarez, J.E., C. Barbes, and C. Hardisson, *Germination of spores of Micromonospora chalcea: physiological and biochemical changes.* J Gen Microbiol, 1980. **121**(1): p. 159-67.
69. Suarez, J.E. and C. Hardisson, *Morphological characteristics of colony development in Micromonospora chalcea.* J Bacteriol, 1985. **162**(3): p. 1342-4.
70. Nodwell, J.R.Y.R., *Microbe Profile: Streptomyces coelicolor: a burlesque of pigments and phenotypes.* Microbiology. **165**(9): p. 953.
71. Glazebrook, M.A., et al., *Sporulation of Streptomyces venezuelae in submerged cultures.* J Gen Microbiol, 1990. **136**(3): p. 581-8.
72. Bush, M.J., et al., *c-di-GMP signalling and the regulation of developmental transitions in streptomycetes.* Nat Rev Microbiol, 2015. **13**(12): p. 749-60.
73. Fuchino, K., et al., *Dynamic gradients of an intermediate filament-like cytoskeleton are recruited by a polarity landmark during apical growth.* Proc Natl Acad Sci U S A, 2013. **110**(21): p. E1889-97.
74. Ditkowski, B., et al., *Dynamic interplay of ParA with the polarity protein, Scy, coordinates the growth with chromosome segregation in Streptomyces coelicolor.* Open Biol, 2013. **3**(3): p. 130006.
75. Bush, M.J., et al., *Hyphal compartmentalization and sporulation in Streptomyces require the conserved cell division protein SepX.* Nat Commun, 2022. **13**(1): p. 71.
76. Ramos-León, F., et al., *A conserved cell division protein directly regulates FtsZ dynamics in filamentous and unicellular actinobacteria.* Elife, 2021. **10**.
77. Schumacher, M.A., et al., *The Streptomyces master regulator BldD binds c-di-GMP sequentially to create a functional BldD2-(c-di-GMP)4 complex.* Nucleic Acids Res, 2017. **45**(11): p. 6923-6933.

78. Elliot, M.A., et al., *BldD from Streptomyces coelicolor is a non-essential global regulator that binds its own promoter as a dimer.* FEMS Microbiol Lett, 2003. **225**(1): p. 35-40.
79. Sosio, M., et al., *Organization of the teicoplanin gene cluster in Actinoplanes teichomyceticus.* Microbiology (Reading), 2004. **150**(Pt 1): p. 95-102.
80. Buttner, M.J., *Actinoplanes Swims into the Molecular Age.* J Bacteriol, 2017. **199**(12).
81. Mouri, Y., et al., *Regulation of Sporangium Formation by BldD in the Rare Actinomycete Actinoplanes missouriensis.* J Bacteriol, 2017. **199**(12).
82. Hashiguchi, Y., et al., *Regulation of Sporangium Formation, Spore Dormancy, and Sporangium Dehiscence by a Hybrid Sensor Histidine Kinase in Actinoplanes missouriensis: Relationship with the Global Transcriptional Regulator TcrA.* J Bacteriol, 2020. **202**(21).
83. Tatusova, T., et al., *NCBI prokaryotic genome annotation pipeline.* Nucleic Acids Res, 2016. **44**(14): p. 6614-24.
84. Costa, S.S., et al., *First Steps in the Analysis of Prokaryotic Pan-Genomes.* Bioinform Biol Insights, 2020. **14**: p. 1177932220938064.
85. Riesco, R., et al., *Defining the Species Micromonospora saelicesensis and Micromonospora noduli Under the Framework of Genomics.* Frontiers in Microbiology, 2018. **9**.
86. Zin, N.M., et al., *Adaptation to Endophytic Lifestyle Through Genome Reduction by Kitasatospora sp. SUK 42.* Front Bioeng Biotechnol, 2021. **9**: p. 740722.
87. Belknap, K.C., et al., *Genome mining of biosynthetic and chemotherapeutic gene clusters in Streptomyces bacteria.* Sci Rep, 2020. **10**(1): p. 2003.
88. Bentley, S.D., et al., *Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2).* Nature, 2002. **417**(6885): p. 141-7.
89. van Bergeijk, D.A., et al., *Ecology and genomics of Actinobacteria: new concepts for natural product discovery.* Nature Reviews Microbiology, 2020. **18**(10): p. 546-558.
90. Algora-Gallardo, L., et al., *Bilateral symmetry of linear streptomycete chromosomes.* Microb Genom, 2021. **7**(11).
91. Adamek, M., et al., *Comparative genomics reveals phylogenetic distribution patterns of secondary metabolites in Amycolatopsis species.* BMC Genomics, 2018. **19**(1): p. 426.
92. Ziemert, N., et al., *Diversity and evolution of secondary metabolism in the marine actinomycete genus Salinispora.* Proc Natl Acad Sci U S A, 2014. **111**(12): p. E1130-9.
93. Gomez-Escribano, J.P., S. Alt, and M.J. Bibb, *Next Generation Sequencing of Actinobacteria for the Discovery of Novel Natural Products.* Mar Drugs, 2016. **14**(4).
94. Bigot, S., et al., *KOPS: DNA motifs that control E. coli chromosome segregation by orienting the FtsK translocase.* EMBO J, 2005. **24**(21): p. 3770-80.
95. Hendrickson, H.L., et al., *Chromosome architecture constrains horizontal gene transfer in bacteria.* PLoS Genet, 2018. **14**(5): p. e1007421.
96. Villalobos, A.S., et al., *Micromonospora tarapacensis sp. nov., a bacterium isolated from a hypersaline lake.* International Journal of Systematic and Evolutionary Microbiology. **71**(11): p. 005109.
97. Saygin, H., et al., *Micromonospora deserti sp. nov., isolated from the Karakum Desert.* International Journal of Systematic and Evolutionary Microbiology. **70**(1): p. 282.

98. Carro, L., et al., *Micromonospora acroterricola sp. nov., a novel actinobacterium isolated from a high altitude Atacama Desert soil.* International Journal of Systematic and Evolutionary Microbiology, 2021. **69**(11): p. 3426.
99. Contador, C.A., et al., *Atacama Database: a platform of the microbiome of the Atacama Desert.* Antonie van Leeuwenhoek, 2020. **113**(2): p. 185-195.
100. Qi, S., et al., *Secondary Metabolites from Marine Micromonospora: Chemistry and Bioactivities.* Chemistry & Biodiversity, 2020. **17**(4): p. e2000024.
101. WEINSTEIN, M.J., et al., *GENTAMICIN, A NEW ANTIBIOTIC COMPLEX FROM MICROMONOSPORA.* J Med Chem, 1963. **6**: p. 463-4.
102. Ban, Y.H., et al., *Complete reconstitution of the diverse pathways of gentamicin B biosynthesis.* Nat Chem Biol, 2019. **15**(3): p. 295-303.
103. Unwin, J., et al., *Gene cluster in Micromonospora echinospora ATCC15835 for the biosynthesis of the gentamicin C complex.* J Antibiot (Tokyo), 2004. **57**(7): p. 436-45.
104. Cohen, D.R. and C.A. Townsend, *C-N-Coupled Metabolites Yield Insights into Dynemicin A Biosynthesis.* Chembiochem, 2020. **21**(15): p. 2137-2142.
105. Zein, N., et al., *Kedarcidin chromophore: an enediyne that cleaves DNA in a sequence-specific manner.* Proc Natl Acad Sci U S A, 1993. **90**(7): p. 2822-6.
106. Suarez, C., et al., *Isolation of bacteria at different points of Pleurotus ostreatus cultivation and their influence in mycelial growth.* Microbiol Res, 2019. **234**: p. 126393.
107. Kronheim, S., et al., *A chemical defence against phage infection.* Nature, 2018.
108. Becher, P.G., et al., *Developmentally regulated volatiles geosmin and 2-methylisoborneol attract a soil arthropod to Streptomyces bacteria promoting spore dispersal.* Nat Microbiol, 2020. **5**(6): p. 821-829.
109. Sagot, B., et al., *Osmotically induced synthesis of the dipeptide N-acetylglutaminylglutamine amide is mediated by a new pathway conserved among bacteria.* Proc Natl Acad Sci U S A, 2010. **107**(28): p. 12652-7.
110. Richter, A.A., et al., *Biosynthesis of the Stress-Protectant and Chemical Chaperon Ectoine: Biochemistry of the Transaminase EctB.* Frontiers in Microbiology, 2019. **10**: p. 2811.
111. Chen, H. and L. Du, *Iterative polyketide biosynthesis by modular polyketide synthases in bacteria.* Appl Microbiol Biotechnol, 2016. **100**(2): p. 541-57.
112. Agrawal, P., et al., *RiPPMiner: a bioinformatics resource for deciphering chemical structures of RiPPs based on prediction of cleavage and cross-links.* Nucleic Acids Res, 2017. **45**(W1): p. W80-W88.
113. Baunach, M., et al., *The Landscape of Recombination Events That Create Nonribosomal Peptide Diversity.* Molecular Biology and Evolution, 2021.
114. Kharel, M.K., et al., *Isolation and characterization of the tobramycin biosynthetic gene cluster from Streptomyces tenebrarius.* FEMS Microbiol Lett, 2004. **230**(2): p. 185-90.
115. Shentu, X.P., et al., *Substantial improvement of toyocamycin production in Streptomyces diastatochromogenes by cumulative drug-resistance mutations.* PLoS One, 2018. **13**(8): p. e0203006.
116. Chen, S., W.A. Kinney, and S. Van Lanen, *Nature's combinatorial biosynthesis and recently engineered production of nucleoside antibiotics in Streptomyces.* World Journal of Microbiology and Biotechnology, 2017. **33**(4): p. 66.
117. Takano, H., *The regulatory mechanism underlying light-inducible production of carotenoids in nonphototrophic bacteria.* Biosci Biotechnol Biochem, 2016. **80**(7): p. 1264-73.

118. Takano, H., et al., *Genetic control for light-induced carotenoid production in non-phototrophic bacteria.* J Ind Microbiol Biotechnol, 2006. **33**(2): p. 88-93.
119. Liu, G., et al., *Molecular regulation of antibiotic biosynthesis in streptomyces.* Microbiol Mol Biol Rev, 2013. **77**(1): p. 112-43.
120. Weber, T., et al., *antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters.* Nucleic Acids Res, 2015. **43**(W1): p. W237-43.
121. Udwary, D.W., et al., *Genome sequencing reveals complex secondary metabolome in the marine actinomycete Salinispora tropica.* Proc Natl Acad Sci U S A, 2007. **104**(25): p. 10376-81.
122. Hoskisson Paul, A., F. Seipke Ryan, and D. Wright Gerard, *Cryptic or Silent? The Known Unknowns, Unknown Knowns, and Unknown Unknowns of Secondary Metabolism.* mBio. **11**(5): p. e02642-20.
123. Kautsar, S.A., et al., *MIBiG 2.0: a repository for biosynthetic gene clusters of known function.* Nucleic Acids Res, 2020. **48**(D1): p. D454-D458.
124. Gilchrist, C.L.M. and Y.H. Chooi, *Clinker & clustermap.js: Automatic generation of gene cluster comparison figures.* Bioinformatics, 2021.
125. Pan, R., et al., *Exploring Structural Diversity of Microbe Secondary Metabolites Using OSMAC Strategy: A Literature Review.* Front Microbiol, 2019. **10**: p. 294.
126. Shirling, E.B. and D. Gottlieb, *Methods for Characterization of Streptomyces Species.* International Journal of Systemic Bacteriology, 1966. **16**(3): p. 313-340.
127. Monod, J., *THE GROWTH OF BACTERIAL CULTURES.* Annual Review of Microbiology, 1949. **3**(1): p. 371-394.
128. Romero-Rodríguez, A., et al., *Transcriptomic analysis of a classical model of carbon catabolite regulation in Streptomyces coelicolor.* BMC Microbiol, 2016. **16**: p. 77.
129. Gubbens, J., et al., *Identification of glucose kinase-dependent and -independent pathways for carbon control of primary metabolism, development and antibiotic production in Streptomyces coelicolor by quantitative proteomics.* Mol Microbiol, 2012. **86**(6): p. 1490-507.
130. Liefke, E., D. Kaiser, and U. Onken, *Growth and product formation of actinomycetes cultivated at increased total pressure and oxygen partial pressure.* Appl Microbiol Biotechnol, 1990. **32**(6): p. 674-9.
131. Kuhl, M., et al., *Microparticles enhance the formation of seven major classes of natural products in native and metabolically engineered actinobacteria through accelerated morphological development.* Biotechnol Bioeng, 2021. **118**(8): p. 3076-3093.
132. Rigali, S., et al., *Feast or famine: the global regulator DasR links nutrient stress to antibiotic production by Streptomyces.* EMBO reports, 2008. **9**(7): p. 670-675.
133. Nicault, M., et al., *Elicitation of Antimicrobial Active Compounds by.* Microorganisms, 2021. **9**(1).
134. Kato, J.Y., et al., *Biosynthesis of gamma-butyrolactone autoregulators that switch on secondary metabolism and morphological development in Streptomyces.* Proc Natl Acad Sci U S A, 2007. **104**(7): p. 2378-83.
135. Ohnishi, Y., et al., *AdpA, a central transcriptional regulator in the A-factor regulatory cascade that leads to morphological development and secondary metabolism in Streptomyces griseus.* Biosci Biotechnol Biochem, 2005. **69**(3): p. 431-9.
136. Yushchuk, O., et al., *Eliciting the silent lucensomycin biosynthetic pathway in Streptomyces cyanogenus S136 via manipulation of the global regulatory gene adpA.* Sci Rep, 2021. **11**(1): p. 3507.

137. Qi, Y., K.K. Nepal, and J.A.V. Blodgett, *A comparative metabologenomic approach reveals mechanistic insights into.* Proc Natl Acad Sci U S A, 2021. **118**(31).
138. Gehrke, E.J., et al., *Silencing cryptic specialized metabolism in.* Elife, 2019. **8**.
139. Nah, H.J., et al., *Cloning and Heterologous Expression of a Large-sized Natural Product Biosynthetic Gene Cluster in.* Front Microbiol, 2017. **8**: p. 394.
140. Donadio, S., et al., *Modular organization of genes required for complex polyketide biosynthesis.* Science, 1991. **252**(5006): p. 675-9.
141. Gomez-Escribano, J.P. and M.J. Bibb, *Engineering Streptomyces coelicolor for heterologous expression of secondary metabolite gene clusters.* Microb Biotechnol, 2011. **4**(2): p. 207-15.
142. Spagnolo, F., M. Trujillo, and J.J. Dennehy, *Why Do Antibiotics Exist?* mBio, 2021. **12**(6): p. e0196621.
143. Chase, A.B., et al., *Vertical Inheritance Facilitates Interspecies Diversification in Biosynthetic Gene Clusters and Specialized Metabolites.* mBio, 2021. **12**(6): p. e0270021.
144. Waglechner, N., A.G. McArthur, and G.D. Wright, *Phylogenetic reconciliation reveals the natural history of glycopeptide antibiotic biosynthesis and resistance.* Nat Microbiol, 2019. **4**(11): p. 1862-1871.
145. Popin, R.V., et al., *Mining of Cyanobacterial Genomes Indicates Natural Product Biosynthetic Gene Clusters Located in Conjugative Plasmids.* Front Microbiol, 2021. **12**: p. 684565.
146. Liu, M., et al., *ICEberg 2.0: an updated database of bacterial integrative and conjugative elements.* Nucleic Acids Res, 2019. **47**(D1): p. D660-D665.
147. Dragoš, A., et al., *Phages carry interbacterial weapons encoded by biosynthetic gene clusters.* Curr Biol, 2021. **31**(16): p. 3479-3489.e5.
148. Fleischmann, R.D., et al., *Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.* Science, 1995. **269**(5223): p. 496-512.
149. Bachmann, B.O., S.G. Van Lanen, and R.H. Baltz, *Microbial genome mining for accelerated natural products discovery: is a renaissance in the making?* J Ind Microbiol Biotechnol, 2014. **41**(2): p. 175-84.
150. Iglesias, A., et al., *Out of the Abyss: Genome and Metagenome Mining Reveals Unexpected Environmental Distribution of Abyssomicins.* Front Microbiol, 2020. **11**: p. 645.
151. Yu, Y., et al., *Identification of the streptothricin and tunicamycin biosynthetic gene clusters by genome mining in Streptomyces sp. strain fd1-xmd.* Appl Microbiol Biotechnol, 2018. **102**(6): p. 2621-2633.
152. Ju, K.S., et al., *Discovery of phosphonic acid natural products by mining the genomes of 10,000 actinomycetes.* Proc Natl Acad Sci U S A, 2015. **112**(39): p. 12175-80.
153. Heather, J.M. and B. Chain, *The sequence of sequencers: The history of sequencing DNA.* Genomics, 2016. **107**(1): p. 1-8.
154. Levy, S.E. and B.E. Boone, *Next-Generation Sequencing Strategies.* Cold Spring Harb Perspect Med, 2019. **9**(7).
155. Reuter, J.A., D.V. Spacek, and M.P. Snyder, *High-throughput sequencing technologies.* Mol Cell, 2015. **58**(4): p. 586-97.
156. *Illumina Sequencing Technology*. 2010.
157. Gaspar, J.M., *NGmerge: merging paired-end reads via novel empirically-derived models of sequencing errors.* BMC Bioinformatics, 2018. **19**(1): p. 536.
158. Yoshinaga, Y., et al., *Genome Sequencing*, in *Fungal Genomics: Methods and Protocols*, R.P. de Vries, A. Tsang, and I.V. Grigoriev, Editors. 2018, Springer New York: New York, NY. p. 37-52.

159. Lu, H., F. Giordano, and Z. Ning, *Oxford Nanopore MinION Sequencing and Genome Assembly.* Genomics Proteomics Bioinformatics, 2016. **14**(5): p. 265-279.
160. Quick, J., A.R. Quinlan, and N.J. Loman, *A reference bacterial genome dataset generated on the MinION™ portable single-molecule nanopore sequencer.* GigaScience, 2014. **3**(1): p. 2047-217X-3-22.
161. Rang, F.J., W.P. Kloosterman, and J. de Ridder, *From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy.* Genome Biol, 2018. **19**(1): p. 90.
162. Loman, N.J., J. Quick, and J.T. Simpson, *A complete bacterial genome assembled de novo using only nanopore sequencing data.* Nat Methods, 2015. **12**(8): p. 733-5.
163. Payne, A., et al., *BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files.* Bioinformatics, 2019. **35**(13): p. 2193-2198.
164. Li, H.a.D.R., *Fast and accurate long-read alignment with Burrows–Wheeler transform.* Bioinformatics, 2010. **26**(5): p. 589--595.
165. Li, H.a.D.R., *Fast and accurate short read alignment with Burrows-Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754--1760.
166. Olson, N.D., et al., *Best practices for evaluating single nucleotide variant calling methods for microbial genomics.* Front Genet, 2015. **6**: p. 235.
167. Wick, R.R., et al., *Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads.* PLoS Comput Biol, 2017. **13**(6): p. e1005595.
168. Bankevich, A., et al., *SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing.* J Comput Biol, 2012. **19**(5): p. 455-77.
169. Koren, S., et al., *Canu: scalable and accurate long-read assembly via adaptive.* Genome Res, 2017. **27**(5): p. 722-736.
170. Kolmogorov, M., et al., *Assembly of long, error-prone reads using repeat graphs.* Nat Biotechnol, 2019. **37**(5): p. 540-546.
171. NCBI. *NCBI Assembly Database Help Page*. Available from: https://www.ncbi.nlm.nih.gov/assembly/help/.
172. Marçais, G., et al., *MUMmer4: A fast and versatile genome alignment system.* PLoS Comput Biol, 2018. **14**(1): p. e1005944.
173. Darling, A.E., B. Mau, and N.T. Perna, *progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement.* PLoS One, 2010. **5**(6): p. e11147.
174. Gurevich, A.a.S.V.a.V.N.a.T.G., *QUAST : quality assessment tool for genome assemblies.* Bioinformatics, 2013. **29**(8): p. 1072--1075.
175. Gomez-Escribano, J.P., et al., *The Streptomyces leeuwenhoekii genome: de novo sequencing and assembly in single contigs of the chromosome, circular plasmid pSLE1 and linear plasmid pSLE2.* BMC Genomics, 2015. **16**: p. 485.
176. Schmid, M., et al., *Pushing the limits of de novo genome assembly for complex prokaryotic genomes harboring very long, near identical repeats.* Nucleic Acids Res, 2018. **46**(17): p. 8953-8965.
177. Goldstein, S., et al., *Evaluation of strategies for the assembly of diverse bacterial genomes using MinION long-read sequencing.* BMC Genomics, 2019. **20**(1): p. 23.
178. Alanjary, M., K. Steinke, and N. Ziemert, *AutoMLST: an automated web server for generating multi-locus species trees highlighting natural product potential.* Nucleic Acids Res, 2019. **47**(W1): p. W276-W282.
179. Medema, M.H., et al., *antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences.* Nucleic Acids Res, 2011. **39**(Web Server issue): p. W339-46.

180. Eddy, S.R., *Accelerated Profile HMM Searches.* PLoS Comput Biol, 2011. **7**(10): p. e1002195.
181. Blin, K., et al., *antiSMASH 6.0: improving cluster detection and comparison capabilities.* Nucleic Acids Res, 2021. **49**(W1): p. W29-W35.
182. Navarro-Muñoz, J.C., et al., *A computational framework to explore large-scale biosynthetic diversity.* Nature Chemical Biology, 2020. **16**(1): p. 60-68.
183. Kautsar, S.A., et al., *BiG-SLiCE: A highly scalable tool maps the diversity of 1.2 million biosynthetic gene clusters.* Gigascience, 2021. **10**(1).
184. Skinnider, M.A., et al., *Comprehensive prediction of secondary metabolite structure and biological activity from microbial genome sequences.* Nat Commun, 2020. **11**(1): p. 6058.
185. Russell, A.H. and A.W. Truman, *Genome mining strategies for ribosomally synthesised and post-translationally modified peptides.* Comput Struct Biotechnol J, 2020. **18**: p. 1838-1851.
186. Xu, L., et al., *Complete genome sequence and comparative genomic analyses of the vancomycin-producing Amycolatopsis orientalis.* BMC Genomics, 2014. **15**: p. 363.
187. Mungan, M.D., et al., *ARTS 2.0: feature updates and expansion of the Antibiotic Resistant Target Seeker for comparative genome mining.* Nucleic Acids Research, 2020. **48**(W1): p. W546-W552.
188. Jalili, V., et al., *The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2020 update.* Nucleic Acids Res, 2020. **48**(W1): p. W395-W402.
189. Schindelin, J., et al., *Fiji: an open-source platform for biological-image analysis.* Nature Methods, 2012. **9**(7): p. 676-682.
190. Schwedock, J., et al., *Assembly of the cell division protein FtsZ into ladder-like structures in the aerial hyphae of Streptomyces coelicolor.* Mol Microbiol, 1997. **25**(5): p. 847-58.
191. Jyothikumar, V., et al., *Cardiolipin synthase is required for Streptomyces coelicolor morphogenesis.* Molecular Microbiology, 2012. **84**(1): p. 181-197.
192. Bosi, E., et al., *MeDuSa: a multi-draft based scaffolder.* Bioinformatics, 2015. **31**(15): p. 2443-51.
193. Jain, C., et al., *High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries.* Nat Commun, 2018. **9**(1): p. 5114.
194. Yu, G., *Data Integration, Manipulation and Visualization of Phylogenetic Trees*. 2022: CRC Press.
195. Seemann, T., *Prokka: rapid prokaryotic genome annotation.* Bioinformatics, 2014. **30**(14): p. 2068-9.
196. Page, A.J., et al., *Roary: rapid large-scale prokaryote pan genome analysis.* Bioinformatics, 2015. **31**(22): p. 3691-3693.
197. Elek, A., M. Kuzman, and K. Vlahovicek, *coRdon: Codon usage analysis and prediction of gene expressivity.* R package version, 2019. **1**(3).
198. Hofacker, I.L., *Vienna RNA secondary structure server.* Nucleic acids research, 2003. **31**(13): p. 3429-3431.
199. Bodenhofer, U., et al., *msa: an R package for multiple sequence alignment.* Bioinformatics, 2015. **31**(24): p. 3997-3999.
200. Zhou, L., et al., *ggmsa: a visual exploration tool for multiple sequence alignment and associated data.* Briefings in Bioinformatics, 2022. **23**(4): p. bbac222.
201. Moreno-Hagelsieb, G. and K. Latimer, *Choosing BLAST options for better detection of orthologs as reciprocal best hits.* Bioinformatics, 2008. **24**(3): p. 319-324.

202. Kudo, N., et al., *Cloning and characterization of a gene involved in aerial mycelium formation in Streptomyces griseus.* J Bacteriol, 1995. **177**(22): p. 6401-10.
203. Al-Bassam, M.M., et al., *Response regulator heterodimer formation controls a key stage in Streptomyces development.* PLoS Genet, 2014. **10**(8): p. e1004554.
204. Bibb, M.J., V. Molle, and M.J. Buttner, *sigma(BldN), an extracytoplasmic function RNA polymerase sigma factor required for aerial mycelium formation in Streptomyces coelicolor A3(2).* J Bacteriol, 2000. **182**(16): p. 4606-16.
205. Bibb, M.J., et al., *Expression of the chaplin and rodlin hydrophobic sheath proteins in Streptomyces venezuelae is controlled by σ(BldN) and a cognate anti-sigma factor, RsbN.* Mol Microbiol, 2012. **84**(6): p. 1033-49.
206. Bush, M.J., et al., *Genes Required for Aerial Growth, Cell Division, and Chromosome Segregation Are Targets of WhiA before Sporulation in Streptomyces venezuelae.* mBio, 2013. **4**(5): p. e00684-13.
207. Davis, N.K. and K.F. Chater, *The Streptomyces coelicolor whiB gene encodes a small transcription factor-like protein dispensable for growth but essential for sporulation.* Mol Gen Genet, 1992. **232**(3): p. 351-8.
208. Kelemen, G.H., et al., *The positions of the sigma-factor genes, whiG and sigF, in the hierarchy controlling the development of spore chains in the aerial hyphae of Streptomyces coelicolor A3(2).* Mol Microbiol, 1996. **21**(3): p. 593-603.
209. Aínsa, J.A., H.D. Parry, and K.F. Chater, *A response regulator-like protein that functions at an intermediate stage of sporulation in Streptomyces coelicolor A3(2).* Molecular Microbiology, 1999. **34**(3): p. 607-619.
210. *SnapGene Software, from Insightful Software (available at snapgene.com)*.
211. Warnes, M.G.R., et al., *Package 'gplots'.* Various R programming tools for plotting data, 2016.
212. Dixon, P., *VEGAN, a package of R functions for community ecology.* Journal of Vegetation Science, 2003. **14**(6): p. 927-930.
213. Paradis, E., J. Claude, and K. Strimmer, *APE: analyses of phylogenetics and evolution in R language.* Bioinformatics, 2004. **20**(2): p. 289-290.
214. Lu, J. and S.L. Salzberg, *SkewIT: The Skew Index Test for large-scale GC Skew analysis of bacterial genomes.* PLoS Comput Biol, 2020. **16**(12): p. e1008439.
215. Durand, G.A., D. Raoult, and G. Dubourg, *Antibiotic discovery: History, methods and perspectives.* Int J Antimicrob Agents, 2018.
216. Kozdrój, J., K. Frączek, and D. Ropek, *Assessment of bioaerosols in indoor air of glasshouses located in a botanical garden.* Building and Environment, 2019. **166**: p. 106436.
217. Ratnakomala, S., et al., *DIVERSITY OF ACTINOMYCETES FROM EKA KARYA BOTANICAL GARDEN, BALI.* BIOTROPIA - The Southeast Asian Journal of Tropical Biology, 2016. **23**(1): p. 42-51.
218. De Oliveira, D.M.P., et al., *Antimicrobial Resistance in ESKAPE Pathogens.* Clin Microbiol Rev, 2020. **33**(3).
219. Wei, J., et al., *α-Glucosidase inhibitors and phytotoxins from Streptomyces xanthophaeus.* Natural Product Research, 2017. **31**(17): p. 2062-2066.
220. Yu, H., X. Xie, and S.-M. Li, *Coupling of cyclo-l-Trp-l-Trp with Hypoxanthine Increases the Structure Diversity of Guanitrypmycins.* Organic Letters, 2019. **21**(22): p. 9104-9108.
221. Zhang, S., et al., *Enhanced Rishirilide Biosynthesis by a Rare In-Cluster Phosphopantetheinyl Transferase in Streptomyces xanthophaeus.* Microbiology Spectrum, 2022. **0**(0): p. e03247-22.

222. Aoyagi, T., et al., *Benarthin: a new inhibitor of pyroglutamyl peptidase. I. Taxonomy, fermentation, isolation and biological activities.* J Antibiot (Tokyo), 1992. **45**(7): p. 1079-83.
223. Hatsu, M., et al., *Benarthin: a new inhibitor of pyroglutamyl peptidase. II. Physico-chemical properties and structure determination.* J Antibiot (Tokyo), 1992. **45**(7): p. 1084-7.
224. Hatsu, M., et al., *Benarthin: a new inhibitor of pyroglutamyl peptidase. III. Synthesis and structure-activity relationships.* J Antibiot (Tokyo), 1992. **45**(7): p. 1088-95.
225. Singh, V., et al., *Isolation, Screening, and Identification of Novel Isolates of Actinomycetes from India for Antimicrobial Applications.* Frontiers in Microbiology, 2016. **7**.
226. Willey, J.M., et al., *Morphogenetic surfactants and their role in the formation of aerial hyphae in Streptomyces coelicolor.* Mol Microbiol, 2006. **59**(3): p. 731-42.
227. Richter, T.K., C.C. Hughes, and B.S. Moore, *Sioxanthin, a novel glycosylated carotenoid, reveals an unusual subclustered biosynthetic pathway.* Environ Microbiol, 2015. **17**(6): p. 2158-71.
228. Carro, L., et al., *Micromonospora cremea sp. nov. and Micromonospora zamorensis sp. nov., isolated from the rhizosphere of Pisum sativum.* Int J Syst Evol Microbiol, 2012. **62**(Pt 12): p. 2971-7.
229. Trujillo, M.E., et al., *Micromonospora coriariae sp. nov., isolated from root nodules of Coriaria myrtifolia.* Int J Syst Evol Microbiol, 2006. **56**(Pt 10): p. 2381-5.
230. Huang, X., et al., *Environmental Sensing in Actinobacteria: a Comprehensive Survey on the Signaling Capacity of This Phylum.* Journal of Bacteriology, 2015. **197**(15): p. 2517-2535.
231. Hoskisson, P.A., G.P. Sharples, and G. Hobbs, *Differentiation and protease production in Micromonospora echinospora (ATCC 15837).* Antonie Van Leeuwenhoek, 2006. **89**(1): p. 191-5.
232. Yagüe, P., et al., *Subcompartmentalization by cross-membranes during early growth of Streptomyces hyphae.* Nat Commun, 2016. **7**: p. 12467.
233. Trujillo, M.E., K. Hong, and O. Genilloud, *The Family Micromonosporaceae*, in *The Prokaryotes: Actinobacteria*, E. Rosenberg, et al., Editors. 2014, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 499-569.
234. Dietz, A. and J. Mathews, *Classification of Streptomyces spore surfaces into five groups.* Appl Microbiol, 1971. **21**(3): p. 527-33.
235. Williams, S.T., et al., *Numerical classification of Streptomyces and related genera.* J Gen Microbiol, 1983. **129**(6): p. 1743-813.
236. Martín, J.F. and P. Liras, *The Balance Metabolism Safety Net: Integration of Stress Signals by Interacting Transcriptional Factors in Streptomyces and Related Actinobacteria.* Frontiers in Microbiology, 2020. **10**.
237. Corretto, E., et al., *Complete genome sequence of the heavy metal resistant bacterium.* Stand Genomic Sci, 2017. **12**: p. 2.
238. Kroppenstedt, R.M., et al., *Eight new species of the genus Micromonospora, Micromonospora citrea sp. nov., Micromonospora echinaurantiaca sp. nov., Micromonospora echinofusca sp. nov. Micromonospora fulviviridis sp. nov., Micromonospora inyonensis sp. nov., Micromonospora peucetia sp. nov., Micromonospora sagamiensis sp. nov., and Micromonospora viridifaciens sp. nov.* Syst Appl Microbiol, 2005. **28**(4): p. 328-39.
239. Leskiw, B.K., et al., *Accumulation of bldA-specified tRNA is temporally regulated in Streptomyces coelicolor A3(2).* Journal of Bacteriology, 1993. **175**(7): p. 1995-2005.

240. Ueda, Y., I. Kumagai, and K.-i. Miura, *The effect of a unique D-loop structure of a minor tRNA leuUUA from Streptomyces on its structural stability and amino acid accepting activity.* Nucleic Acids Research, 1992. **20**(15): p. 3911-3917.
241. Chen, X., et al., *A new bacterial tRNA enhances antibiotic production in Streptomyces by circumventing inefficient wobble base-pairing.* Nucleic Acids Research, 2022. **50**(12): p. 7084-7096.
242. Kalan, L., et al., *A Cryptic Polyene Biosynthetic Gene Cluster in Streptomyces calvus Is Expressed upon Complementation with a Functional bldA Gene.* Chemistry & Biology, 2013. **20**(10): p. 1214-1224.
243. Hutchings, M.I., A.W. Truman, and B. Wilkinson, *Antibiotics: past, present and future.* Curr Opin Microbiol, 2019. **51**: p. 72-80.
244. Ziemert, N., M. Alanjary, and T. Weber, *The evolution of genome mining in microbes - a review.* Nat Prod Rep, 2016. **33**(8): p. 988-1005.
245. Awakawa, T., et al., *Characterization of the biosynthesis gene cluster for alkyl-O-dihydrogeranyl-methoxyhydroquinones in Actinoplanes missouriensis.* Chembiochem, 2011. **12**(3): p. 439-48.
246. Grubbs, K.J., et al., *Large-Scale Bioinformatics Analysis of Bacillus Genomes Uncovers Conserved Roles of Natural Products in Bacterial Physiology.* mSystems, 2017. **2**(6): p. e00040-17.
247. Caicedo-Montoya, C., M. Manzo-Ruiz, and R. Ríos-Estepa, *Pan-Genome of the Genus Streptomyces and Prioritization of Biosynthetic Gene Clusters With Potential to Produce Antibiotic Compounds.* Frontiers in Microbiology, 2021. **12**.
248. Palmer, M., et al., *All ANIs are not created equal: implications for prokaryotic species boundaries and integration of ANIs into polyphasic taxonomy.* International Journal of Systematic and Evolutionary Microbiology, 2020. **70**(4): p. 2937-2948.
249. Park, C.J., J.T. Smith, and C.P. Andam, *Horizontal Gene Transfer and Genome Evolution in the Phylum Actinobacteria*, in *Horizontal Gene Transfer: Breaking Borders Between Living Kingdoms*, T.G. Villa and M. Viñas, Editors. 2019, Springer International Publishing: Cham. p. 155-174.
250. Ara, I. and T. Kudo, *Two new species of the genus Micromonospora: Micromonospora chokoriensis sp. nov. and Micromonospora coxensis sp. nov., isolated from sandy soil.* J Gen Appl Microbiol, 2007. **53**(1): p. 29-37.
251. Jongrungruangchok, S., S. Tanasupawat, and T. Kudo, *Micromonospora krabiensis sp. nov., isolated from marine soil in Thailand.* J Gen Appl Microbiol, 2008. **54**(2): p. 127-33.
252. Li, L., et al., *Micromonospora avicenniae sp. nov., isolated from a root of Avicennia marina.* Antonie Van Leeuwenhoek, 2013. **103**(5): p. 1089-96.
253. Ramírez-Durán, N., et al., *Taxogenomic and Comparative Genomic Analysis of the Genus Saccharomonospora Focused on the Identification of Biosynthetic Clusters PKS and NRPS.* Frontiers in Microbiology, 2021. **12**.
254. Schwientek, P., et al., *The complete genome sequence of the acarbose producer Actinoplanes sp. SE50/110.* BMC Genomics, 2012. **13**(1): p. 112.
255. Kumar, S., et al., *MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms.* Mol Biol Evol, 2018. **35**(6): p. 1547-1549.
256. Ciufo, S., et al., *Using average nucleotide identity to improve taxonomic assignments in prokaryotic genomes at the NCBI.* Int J Syst Evol Microbiol, 2018. **68**(7): p. 2386-2392.
257. Zheng, W.X., et al., *Essentiality drives the orientation bias of bacterial genes in a continuous manner.* Sci Rep, 2015. **5**: p. 16431.

258. Hendrickson, H. and J.G. Lawrence, *Selection for chromosome architecture in bacteria.* J Mol Evol, 2006. **62**(5): p. 615-29.
259. Jost, L., *Partitioning diversity into independent alpha and beta components.* Ecology, 2007. **88**(10): p. 2427-39.
260. Xia, L., et al., *Biosynthetic gene cluster profiling predicts the positive association between antagonism and phylogeny in Bacillus.* Nat Commun, 2022. **13**(1): p. 1023.
261. Bruns, H., et al., *Function-related replacement of bacterial siderophore pathways.* ISME J, 2018. **12**(2): p. 320-329.
262. Rocha, E.P. and A. Danchin, *Gene essentiality determines chromosome organisation in bacteria.* Nucleic Acids Res, 2003. **31**(22): p. 6570-7.
263. Wang, M., et al., *Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking.* Nat Biotechnol, 2016. **34**(8): p. 828-837.
264. Schorn, M.A., et al., *A community resource for paired genomic and metabolomic data mining.* Nat Chem Biol, 2021. **17**(4): p. 363-368.
265. Covington, B.C., J.A. McLean, and B.O. Bachmann, *Comparative mass spectrometry-based metabolomics strategies for the investigation of microbial secondary metabolites.* Nat Prod Rep, 2017. **34**(1): p. 6-24.
266. Amos, G.C.A., et al., *Comparative transcriptomics as a guide to natural product discovery and biosynthetic gene cluster functionality.* Proc Natl Acad Sci U S A, 2017. **114**(52): p. E11121-E11130.
267. Chang, Y., et al., *Overproduction of gentamicin B in industrial strain.* Metab Eng Commun, 2019. **9**: p. e00096.
268. Tong, Y., et al., *CRISPR–Cas9, CRISPRi and CRISPR-BEST-mediated genetic manipulation in streptomycetes.* Nature Protocols, 2020.
269. McHugh, R.E., et al., *Biosynthesis of Aurodox, a Type III Secretion System Inhibitor from Streptomyces goldiniensis.* Appl Environ Microbiol, 2022. **88**(15): p. e0069222.
270. Chandra, G. and K.F. Chater, *Developmental biology of Streptomyces from the perspective of 100 actinobacterial genome sequences.* FEMS Microbiol Rev, 2014. **38**(3): p. 345-79.
271. Yagüe, P., et al., *Transcriptomic analysis of Streptomyces coelicolor differentiation in solid sporulating cultures: first compartmentalized and second multinucleated mycelia have different and distinctive transcriptomes.* PLoS One, 2013. **8**(3): p. e60665.
272. Emms, D.M. and S. Kelly, *OrthoFinder: phylogenetic orthology inference for comparative genomics.* Genome Biol, 2019. **20**(1): p. 238.
273. Chandra, G. and K.F. Chater, *Evolutionary flux of potentially bldA-dependent Streptomyces genes containing the rare leucine codon TTA.* Antonie van Leeuwenhoek, 2008. **94**(1): p. 111-126.
274. Kataoka, M., S. Kosono, and G. Tsujimoto, *Spatial and temporal regulation of protein expression by bldA within a Streptomyces lividans colony.* FEBS Lett, 1999. **462**(3): p. 425-9.
275. Redondo-Salvo, S., et al., *Pathways for horizontal gene transfer in bacteria revealed by a global map of their plasmids.* Nat Commun, 2020. **11**(1): p. 3602.
276. Richter, M. and R. Rosselló-Móra, *Shifting the genomic gold standard for the prokaryotic species definition.* Proc Natl Acad Sci U S A, 2009. **106**(45): p. 19126-31.
277. Szafran, M.J., et al., *Spatial rearrangement of the Streptomyces venezuelae linear chromosome during sporogenic development.* Nat Commun, 2021. **12**(1): p. 5222.
278. Lioy, V.S., et al., *Dynamics of the compartmentalized Streptomyces chromosome during metabolic differentiation.* Nat Commun, 2021. **12**(1): p. 5221.

279. Blin, K., et al., *The antiSMASH database version 2: a comprehensive resource on secondary metabolite biosynthetic gene clusters.* Nucleic Acids Res, 2019. **47**(D1): p. D625-D630.
280. Kieser, T., et al., *Practical streptomyces genetics*. Vol. 291. 2000: John Innes Foundation Norwich.