Characterising the production of novel antimicrobials in *Streptomyces formicae*

Rebecca Devine

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy at the University of East Anglia

University of East Anglia

School of Biological Sciences

June 2019

Abstract

Antibiotic resistance poses a major risk to modern medicine, therefore finding new antimicrobial compounds is vital. Most currently used antibiotics originate from actinomycetes discovered more than half a century ago. Previous work from the Hutchings laboratory led to the isolation of a new *Streptomyces* species named *S. formicae* from the African fungus-farming plant-ant, *Tetraponera penzigi.* *S. formicae* produces novel pentacyclic polyketides, the formicamycins, that have potent antibacterial activity against drug-resistant pathogens including methicillin-resistant *Staphylococcus aureus* (MRSA) and vancomycin-resistant *Enterococci* (VRE). During this work, the genes responsible for formicamycin biosynthesis in the native producer were identified and characterised in detail using CRISPR/Cas9-mediated genome editing. In addition, we used cappable RNA- and ChIP-sequencing to determine the transcriptional organisation of the pathway and study the regulatory cascade controlling the production of- and host resistance to- these potent antimicrobials. We exploited this information to generate multiple mutants of *S. formicae* that overproduce formicamycins as well as various biosynthetic intermediates and shunt metabolites, some of which also have bioactivity. Attempts to understand the evolutionary origins of the biosynthetic pathway and the mode of action of these novel compounds are also presented. Furthermore, the potential for novel chemistry from *S. formicae* is not limited to the formicamycin pathway; antiSMASH analysis shows this talented strain contains at least 45 secondary metabolite biosynthetic gene clusters (BGCs). Under standard laboratory conditions, wild-type *S. formicae* also exhibits antifungal activity against the drug-resistant *Lomentospora prolificans,* and when the formicamycin BGC is deleted, the strain produces additional compounds with potent antibacterial activity against MRSA. Overall, this work demonstrates that searching under-explored environments for new species combined with genome editing is a promising route towards finding new anti-infectives.

This thesis is 305 pages and 79 298 words in length

Publications arising from the work in this thesis

Devine R, McDonald H, Noble K, Holmes NA, Qin Z, Chandra G, Wilkinson B and Hutchings MI (2019) Regulation and refactoring of the formicamycin biosynthetic gene cluster in *Streptomyces* species. In preparation, journal to be determined.

Qin Z*, **Devine R***, Booth T, Hutchings MI, Wilkinson B (2019) A two-step ring expansion-contraction mechanism for polyketide formicamycin biosynthesis. In preparation for the Journal of the American Chemical Society.

Qin Z*, **Devine R***, Hutchings MI, Wilkinson B (2019) A role for antibiotic biosynthesis monooxygenase domain proteins in fidelity control during aromatic polyketide biosynthesis. Nature Communications **10:**3611
*These authors contributed equally to this work

McLean T, Hutchings MI, Wilkinson B and **Devine R** (2019) Dissolution of the disparate: A review of co-ordinate regulation in antibiotic biosynthesis. Antibiotics 8:82-100

Holmes NA, **Devine R**, Qin Z, Seipke RF, Wilkinson B, Hutchings MI (2018) Complete genome sequence of *Streptomyces formicae* KY5, the formicamycin producer. Journal of Biotechnology 265:116–118

Devine R, Hutchings MI, Holmes NA (2017) Future directions for the discovery of antibiotics from actinomycete bacteria. Emerging Topics in Life Science. 1:1-12

Qin Z, Munnoch JT, **Devine R**, Holmes NA, Seipke RF, Wilkinson KA, Wilkinson B, Hutchings MI (2017) Formicamycins, antibacterial polyketides produced by *Streptomyces formicae* isolated from African Tetraponera plant-ants. Chemical Science 8(4):3218-3227

Contents

1 Introduction 9

1.1 Antibiotics 9

1.2 Antimicrobial resistance (AMR) 11

1.3 Natural products 14

1.3.1 Polyketides 14

1.3.2 Non-ribosomal peptides 20

1.3.3 Ribosomally synthesised natural products 24

1.3.4 Discovering the natural product antibiotics of the future 26

1.4 The genus *Streptomyces* 27

1.4.1 *Streptomyces* biology 28

1.4.2 Regulation of antibiotic production in *Streptomyces* 30

1.5 Tools for genome mining and natural product discovery in *Streptomyces* 35

1.5.1 Pleiotropic methods of inducing secondary metabolism 35

1.5.2 Genome editing of BGCs for discovery of novel natural products 37

1.6 Exploring novel environments to find new antibiotic producers 41

1.7 *Streptomyces formicae* 45

1.8 Novel antibiotics from *Streptomyces formicae* 48

1.9 Aims and objectives of this thesis 53

2 Materials and Methods 54

2.1 Chemicals and Reagents 54

2.2 Bacterial strains 54

2.3 Preparation of *Streptomyces* spores 56

2.4 Glycerol Stocks 56

2.5 Microscopy 56

2.6 DNA Extraction 57

2.7 DNA Quantification 57

2.8 Primers 57

2.9 Polymerase Chain Reaction (PCR) 58

2.9.1 PCRBIO® Taq DNA Polymerase 58

2.9.2 Q5® High-Fidelity DNA polymerase 59

2.10 Agarose gel electrophoresis 59

2.11 Gene Synthesis 60

2.12 Plasmid preparation 60

2.13 Restriction digest 60

2.14 Gel extraction 60

2.15 Ligation 60
2.16 Golden Gate 61
2.17 Gibson Assembly 61
2.18 Preparation and transformation of electrocompetent *E. coli* 61
2.19 Preparation and transformation of chemically competent *E. coli* 62
2.20 Tri-parental mating 62
2.21 Colony PCR in *E. coli* 63
2.22 Constructing gene deletions using Lambda λ RED methodology (ReDirect) 63
2.23 Constructing gene knockouts using CRISPR/Cas9 genome editing 63
2.24 Sequencing of small DNA fragments and plasmids 66
2.25 Conjugation 66
2.26 Colony PCR in *Streptomyces* species 66
2.27 Genetic complementation 67
2.28 Chemical Extraction of Secondary Metabolites from agar plates 67
2.29 High-performance Liquid Chromatography (HPLC) and Mass Spectrometry (LCMS) 67
2.30 Protein overexpression and purification 68
2.31 Protein analysis by SDS-PAGE 69
2.32 Quantification of protein from whole cell lysates 70
2.33 Western blot 71
2.34 Analysis of compounds by UV-Vis 72
2.35 RNA Extraction 72
2.36 RNA Quantification 73
2.37 RT-PCR 74
2.38 Cappable RNA Sequencing 75
2.39 Chromatin Immuno-precipitation (ChIP) Sequencing 75
2.40 Analysis of ChIP-Sequencing data 77
2.41 Solid culture colony bioassay 77
3 Identifying the formicamycin biosynthetic gene cluster and the mode of action of the formicamycins 78
3.1 Identifying the formicamycin biosynthetic gene cluster 79
3.2 Defining the edges of the formicamycin BGC 88
3.3 Identifying the enzyme responsible for halogenation of the formicamycins 91

3.4 Predicted formicamycin biosynthetic pathway 102

3.5 Mode of action of the formicamycins..... 105

3.6 Discussion..... 117

4 Regulation of formicamycin production in *Streptomyces formicae* and mechanisms of host resistance 119

4.1 Determining the transcriptional organization of the formicamycin BGC 120

4.2 Characterising cluster situated regulators of formicamycin biosynthesis..... 126

4.2.1 The MarR-family regulator, ForJ 127

4.2.2 The two-component system, ForGF 132

4.2.3 The MarR-family regulator, ForZ..... 134

4.3 Identifying targets of ForJ, ForF and ForZ within the formicamycin BGC 136

4.3.1 ChIP Sequencing..... 136

4.3.2 Motif analysis of DNA binding sites identified by ChIP-seq 141

4.3.3 Mechanistic insights into regulation of the formicamycin BGC 144

4.4 Understanding the mechanisms of host resistance to formicamycins in *S. formicae*..... 149

4.5 Discussion..... 155

5 The biosynthetic and evolutionary link between fasamycins and formicamycins..... 161

5.1 Searching for BGCs related to the formicamycin BGC..... 162

5.2 The mechanism of conversion of a fasamycin to a formicamycin..... 175

5.3 Methylation of the fasamycins and formicamycins 182

5.4 Discussion..... 189

6 Genetic engineering of the formicamycin BGC to obtain novel, pre-defined products, by-products and biosynthetic intermediates 192

6.1 Identifying further congeners from the formicamycin biosynthetic pathway . 193

6.1.1 6-chlorogenistein 194

6.1.2 Formicapyridines; novel bi-products from the formicamycin biosynthesis pathway.. 195

6.2 The functions of the five putative cyclases in the biosynthesis of fasamycins, formicamycins and formicapyridines..... 202

6.3 Engineering highly efficient strains of *S. formicae* that overproduce specific pathway products..... 219

6.4 The proposed fasamycin, formicamycin and formicapyridine biosynthesis pathway in *S. formicae* 230

6.5 Discussion..... 233

7 Conclusions and further work 237

7.1 Characterising the biosynthesis of the novel antibiotics, the formicamycins, in *Streptomyces formicae* 237

7.2 The potential for the discovery of more novel natural products from *Streptomyces formicae* 242

7.3 Preliminary results of a genome mining project using CRISPR/Cas9 in *S. formicae* 249

7.4 Final conclusions 256

8 Appendix 258

8.1 Strains used and generated during this thesis 258

8.2 Plasmids used and generated during this thesis 265

8.3 Primers used and generated during this thesis 270

8.4 Phylogenetic analysis of ForS and similar ABM domain containing proteins 283

8.5 Bioassay images from CRISPR/Cas9 genome mining project 284

9 References 289

Acknowledgements

I would like to thank my supervisors, Professor Matt Hutchings and Professor Barrie Wilkinson, for pushing me to become a better scientist, encouraging me to question everything and think outside the box. I have been grateful for every opportunity that has come my way during my time at UEA and I have achieved things I could not have imagined when I started my PhD. This project also would not have been possible without the extensive collaboration with Dr Zhiwei Qin and Hannah McDonald. Thanks also to Govind Chandra for the bioinformatic analysis and to the BBSRC NRP DTP for funding.

I would like to thank all the members of the Hutchings lab, both past and present, for making it such a great environment to work in; especially Sarah, my best friend and PhD-twin from day one, John Munnoch for getting me started in the lab when I knew nothing, and everyone else for the insightful discussions and regular laughs. Thanks also to all those in the Pamela Salter office who have become my close friends.

I know that I would not be the person I am today without my amazing mother who gave up everything to give me a future, even when the worst happened. I feel so lucky to have this phenomenal woman as my mentor and my closest friend. Lastly, eternal thanks to my partner Edward, who has always been my biggest cheerleader, even, and perhaps most especially, when I did not believe in myself. Being able to share this journey with you has made the load so much more bearable, even though we were many miles apart for most of it. I am so grateful to you for all the help you have given me, PhD related and not, over the last five years. I can't wait to see what our future holds.

1 Introduction

1.1 Antibiotics

The use of antimicrobials in the treatment and prevention of disease was one of the most revolutionary interventions to be introduced into modern medicine. Antibiotics are compounds that either inhibit the growth of, or directly kill, bacterial cells. They act by targeting essential cell processes within microbial cells, such as inhibiting biosynthesis of the cell wall (e.g. the β-lactams), preventing protein synthesis usually by interacting with ribosomal subunits (e.g. chloramphenicol, the aminoglycosides), interacting with nucleic acids or their repair machinery (e.g. the fluoroquinolones, rifampicin), interfering with metabolic pathways such as fatty acid biosynthesis (e.g. isoniazid) or folic acid synthesis (e.g. the sulphonamides), or by disrupting the structure of the membrane (e.g. daptomycin) (Sultan *et al.*, 2018). By selectively targeting biochemical differences between prokaryotic and eukaryotic cells, antibiotics are less harmful to the host and cause relatively fewer side effects compared to other medicinal compounds. In fact, antibiotics are unique amongst pharmaceutical agents as they can cure disease rather than simply alleviating the symptoms (Demain, 2009). Most of the antibiotics that we currently use in the clinic were discovered more than half a century ago. Since the beginning of this 'antibiotic era', antibiotics have been used to rapidly treat infectious diseases that were once commonplace, such as cholera and typhoid fever, as well as wound infections that might once have proved fatal. In addition, the use of antibiotics has enabled more complex interventions such as organ transplant surgery and chemotherapy to be developed, as these immunocompromised patients can be protected from the potentially serious complications that could arise from infection (Livermore, 2004).

Approximately two thirds of all the antibiotics and chemotherapeutic agents currently used in the clinic are derived from the natural products of soil dwelling actinomycetes, most notably *Streptomyces* bacteria (Manteca and Yagüe, 2018). These microorganisms are constantly having to adapt to rapidly changing

environments to compete for nutrients and other resources and they produce these molecules as part of their secondary metabolism. Though costly, the production of antimicrobial secondary metabolites can provide a significant survival advantage over neighbouring bacteria in the competition for resources (Barke *et al.*, 2010). One of the earliest and most well publicised descriptions of an antibiotic produced naturally by a microorganism was in 1929 when Alexander Fleming observed an inhibitory substance originating from a contaminating mould that was lysing the bacterial cells it encountered (Fleming, 1929). Now known as penicillin, this finding was the primary initiator of the 'Golden Age' of antibiotic discovery. During the 1940s and 50s, almost all currently known classes of antibacterials were discovered, with between 70 and 80% being isolated from *Streptomyces* species (Bérdy, 2005). The first well characterised *Streptomyces* natural product was streptomycin, described in 1943, isolated from *Streptomyces griseus*. This became the first commercially available treatment for tuberculosis and is still used in the treatment of tuberculosis patients today (Ohnishi *et al.*, 2008). Chloramphenicol, now commonly used as a treatment for eye infections, was subsequently discovered from *Streptomyces venezuelae* in 1947 and is now made synthetically. Over 9000 bioactive molecules have been isolated from actinomycete bacteria since the 1940s and around 60 are still in clinical use today (Demain, 2009).

Despite this success, the discovery of new strains and bioactive molecules peaked in the 1950s due to high rates of rediscovery of already characterised compounds (Hover *et al.*, 2018). Instead, the focus turned to rational target-based design of synthetic compounds by pharmaceutical companies, an approach which was also largely unsuccessful (Peláez, 2006). Companies came to the realisation that it was more profitable to invest in the development of drugs that are administered longer term, like statins or chemotherapeutic agents, rather than anti-infectives that are often only required for a few days at a time. Furthermore, due to increasing concerns about antibiotic resistance, governments began restricting the use of new antibiotics for extreme cases of infection only, reducing potential profits from any developments even further (Demain, 2009). As a result, interest in antibiotic development by large pharmaceutical companies has plummeted, with just four

companies still actively engaged in antibiotic development today (Nature Biotechnology, 2018). The numbers of antibiotics approved by the Food and Drug Administration (FDA) in the US over the last 20 years has fallen by 56% and only three new antibacterial classes, including lipopeptides like daptomycin, have been introduced into the clinic since the 1970s (Butler and Buss, 2006). With no new compounds available to treat infectious disease and pathogens rapidly developing resistance to the antibiotics we have available, the treatment of bacterial infections is becoming increasingly challenging once again.

1.2 Antimicrobial resistance (AMR)

The huge success of antimicrobial therapy in the late 1900s led to the widespread distribution and use of these compounds, not just in a clinical setting, but also in veterinary medicine and agriculture. This extensive use means antibiotics have accumulated in the environment, resulting in extreme selection pressure for resistance to develop (Sultan *et al.*, 2018). Bacteria and other microorganisms can acquire or evolve mechanisms to resist antimicrobial toxicity by increasing the expression of efflux pumps to prevent accumulation of the compound in the cell, modifying the target so that the compound is no longer functional, or inactivating the antibiotic enzymatically, as in the case of the beta-lactamases (Kapoor, Saigal and Elongavan, 2017). These adaptations can occur spontaneously through mutation and are frequently shared between bacterial species on mobile genetic elements such as plasmids. This ability to undergo horizontal gene transfer, together with the relatively short generation times, means bacteria and other microorganisms are capable of evolving resistance much faster than we can generate new drugs (Ventola, 2015). Furthermore, antimicrobials are produced by bacteria and fungi as a part of their normal secondary metabolism, meaning their associated resistance genes are a naturally occurring phenomenon and exist in the environment all the time (Davies and Davies, 2010). Resistance genes are now alarmingly widespread in the environment. A global metagenomics survey found antibiotic resistance gene determinants (ARGDs) in all of the 71 different environments selected for testing,

including a secluded cave and frozen rock sediments that had likely not been exposed to human activity for thousands of years (Nesme *et al.*, 2014).

The ability of pathogenic bacteria to acquire and disseminate resistance genes significantly compromises the therapeutic potential of our available antibiotics. As a result, we are increasingly unable to treat diseases caused by drug-resistant strains of bacteria such as methicillin-resistant *Staphylococcus aureus* (MRSA) and *Mycobacterium tuberculosis,* the causative agent of tuberculosis. *M. tuberculosis* (also known as the tubercle bacillus or TB) often becomes multidrug-resistant (MDR-TB) because a combination of antibiotics are usually required for substantial lengths of time, from six months to up to two years, and strains of extensively drug-resistant (XDR) TB are becoming increasingly common (Davies and Davies, 2010). As well as Gram-positive bacterial pathogens, high rates of resistance to third generation antibiotics have been reported in *Escherichia coli* and *Klebsiella pneumoniae* isolates. The recent emergence of MCR-1 mediated colistin resistance in *E. coli* was especially shocking as these compounds were the last remaining antibiotics that did not appear to select for resistance and were therefore seen as the last resort treatment for persistent drug-resistant infection by Gram-negative pathogens (Liu *et al.*, 2016). In addition, it is not just bacterial pathogens that develop resistance; systemic infections with drug-resistant fungi also represent a major clinical challenge. Although mostly documented in *Candida* isolates, multidrug-resistance is also becoming a problem in many *Aspergillus* strains as well as in emerging human pathogens such as *Lomentospora* (formerly *Scedosporium*) (Wiederhold, 2017).

One of the main drivers of AMR is the overuse of antibiotics. Treatment by antibiotics is effective at removing antibiotic-sensitive bacteria from a community involved in infection, but resistant isolates will be left behind, meaning that during subsequent infections, patients are likely to be primarily colonized by bacteria that are already resistant to the frontline antibiotic (Ventola, 2015). The dangers of overusing and misusing antibiotics have been well-known for many years; in his Nobel Prize acceptance speech in 1945, Fleming warned that the inappropriate use of penicillin would lead to more serious, drug-resistant infections that could be passed on through the community (Alanis, 2005). Nevertheless, antibiotics continue to be

overprescribed in many countries and elsewhere remain unregulated and easily available both over the counter and online (Ventola, 2015). The overuse of antibiotics by clinicians is reflective of the lack of available diagnostic tests at the point of care. Without any information about the causative agent and the antimicrobial susceptibility profile, clinicians often have to use multiple antibiotics in the hope that one will be effective at treating the disease. Whilst assuring better patient outcomes, this approach drives the development of resistance in the clinic (Michael, Dominey-Howes and Labbate, 2014). Furthermore, when antibiotics are consumed, a large portion of the required dose is excreted in the active form and therefore enters waste water as sewage. Wastewater treatment plants are effective at removing living microorganisms from sewage but undegraded antimicrobials and the genetic material from treated pathogens remain. Similarly, when farm animals are treated with antimicrobials, the AMR genes generated during this selection will enter sludge, some of which is later spread on the land and used as crop fertilizer. All these practices have the potential to spread ever greater quantities of AMR genes in the environment, contributing to the rapid spread of drug-resistant pathogens (Singer *et al.*, 2016).

AMR is generally accepted as one of the greatest threats to modern healthcare. Without the means to treat infections we risk entering a post-antibiotic era, where minor infections become fatal once again and surgical interventions cannot take place due to the lack of available antibiotic prophylaxis. The 2014 report led by economist Lord Jim O'Neil predicted that if current rates of resistance continue, more than 10 million people could die from an antibiotic-resistant infection every year by 2050, making it a bigger killer than cancer. If AMR is not effectively controlled, it could also have significant economic impacts. It is estimated that $100 trillion USD of economic output will be lost every year and each person in the world will be $10,000 USD worse off by 2050. In order to combat this, 10 key areas were highlighted in the report as research priorities for the future, including better surveillance of resistant infections, developing rapid and accurate diagnostic tests for clinical use and finding new drugs with activity against resistant pathogens (O'Neil, 2016).

1.3 Natural products

Research has returned to natural products in an effort to discover new compounds with novel mechanisms of action that may be effective at treating drug-resistant infections. Natural products are predicted to be better candidates for the development of new antimicrobials than synthetic molecules as millions of years of evolutionary adaptation ensures that secondary metabolites provide the greatest selective advantage to the producing organism. As such they are often more effective at crossing the cell membranes of other bacteria and interacting with specific intracellular targets (Butler, Blaskovich and Cooper, 2013). Natural products are often large, highly decorated molecules generated by multi-enzyme complexes, such as polyketide synthases (PKS) and non-ribosomal peptide synthetases (NRPS), although some small molecules like terpenes and alkaloids also poses antimicrobial activity (Demain, 2009). The genes required for making natural products are often clustered together in the genome of the producing organism, with the core enzyme machinery being adjacent to accessory enzymes, transporters and regulatory elements. These biosynthetic gene clusters (BGCs) vary in size from a few thousand base pairs (bp) to over 100 kb (Bibb, 2005).

1.3.1 Polyketides

PKS gene clusters give rise to a large class of structurally diverse molecules, such as macrolides, aromatic compounds and polyenes (Hofeditz *et al.*, 2018). Many products of PKS BGCs are used as commercial antibiotics, such as erythromycin, or are used as a basis for the development of semi-synthetic molecules like doxycycline (Ward and Allenby, 2018). Polyketides are derived from the successive decarboxylative condensation of multiple extender units, usually malonyl-CoenzymeA (CoA) or methylmalonyl-CoA, to generate an elongated carbon chain of pre-determined length upon the incorporation of an initiating starter unit (Fischbach and Walsh, 2006; Rutledge and Challis, 2015). The chain length, as well as the extensive modifications to the oxidation state and stereochemistry of the resulting

keto groups on these compounds, contribute to the high degree of structural and chemical diversity of polyketides (Ray and Moore, 2002).

PKS enzymes can be classified into three distinct groups; type 1, type 2 and type 3. Type 1 PKS enzymes are large enzymes with distinct catalytic domains, known as modules, each responsible for the addition of a single extender unit and any modification of the associated β-keto group. Each module therefore elongates the product by 2 carbons using an acyltransferase domain (AT) that loads the acyl carrier protein (ACP) with a single extender unit, on which the ketosynthase (KS) can then act to form a carbon-carbon bond between the extender unit and the growing product (Wang, Yuan and Zheng, 2019). This modular architecture of T1PKSs means that the chain length of the final product often corresponds with the number of modules present in the PKS (Rutledge and Challis, 2015). Modules may also contain domains that modify the β-keto group to either a hydroxyl group by the action of a ketoreductase (KR), to a carbon-carbon double bond by the action of a dehydratase (DH) or to a single carbon-carbon bond by an enoylreductase (ER) (Dutta *et al.*, 2014). Analysis of the catalytic domains can predict the β-keto group modifications that will be incorporated, making it possible to estimate the structure of a natural product from the architecture of the predicted T1PKS encoded in a genome sequence (Rutledge and Challis, 2015). Once the chain is extended to the pre-determined length (i.e. once it reaches the final module in the PKS), the ACP transfers the polyketide to the thioesterase (TE) unit for release from the PKS by hydrolysis before further modifications by accessory enzymes take place (Dutta *et al.*, 2014).

A good example of this modular system is the biosynthesis of erythromycin A. The PKS for erythromycin A biosynthesis consists of a loading module and six extender modules. The loading module is responsible for loading one propionyl-CoA starter unit onto the ACP. The six extender modules then each add a single methylmalonyl-CoA extender unit using the combined actions of the AT and the KS that catalyses the condensation of neighbouring extender units to form a growing polyketide chain. Some of the modules also encode for KR, DH and ER subunits to modify the extender unit that is added by modifying the oxidative state of the β-keto group as described above. Once the poly-β-keto intermediate is released from the ACP by the TE, it is

cyclised to form 6-deoxyerythronolide B before hydroxylase, O-methyltransferase and glycosyltransferase enzymes form the final erythromycin A structure (**Figure 1.1).**

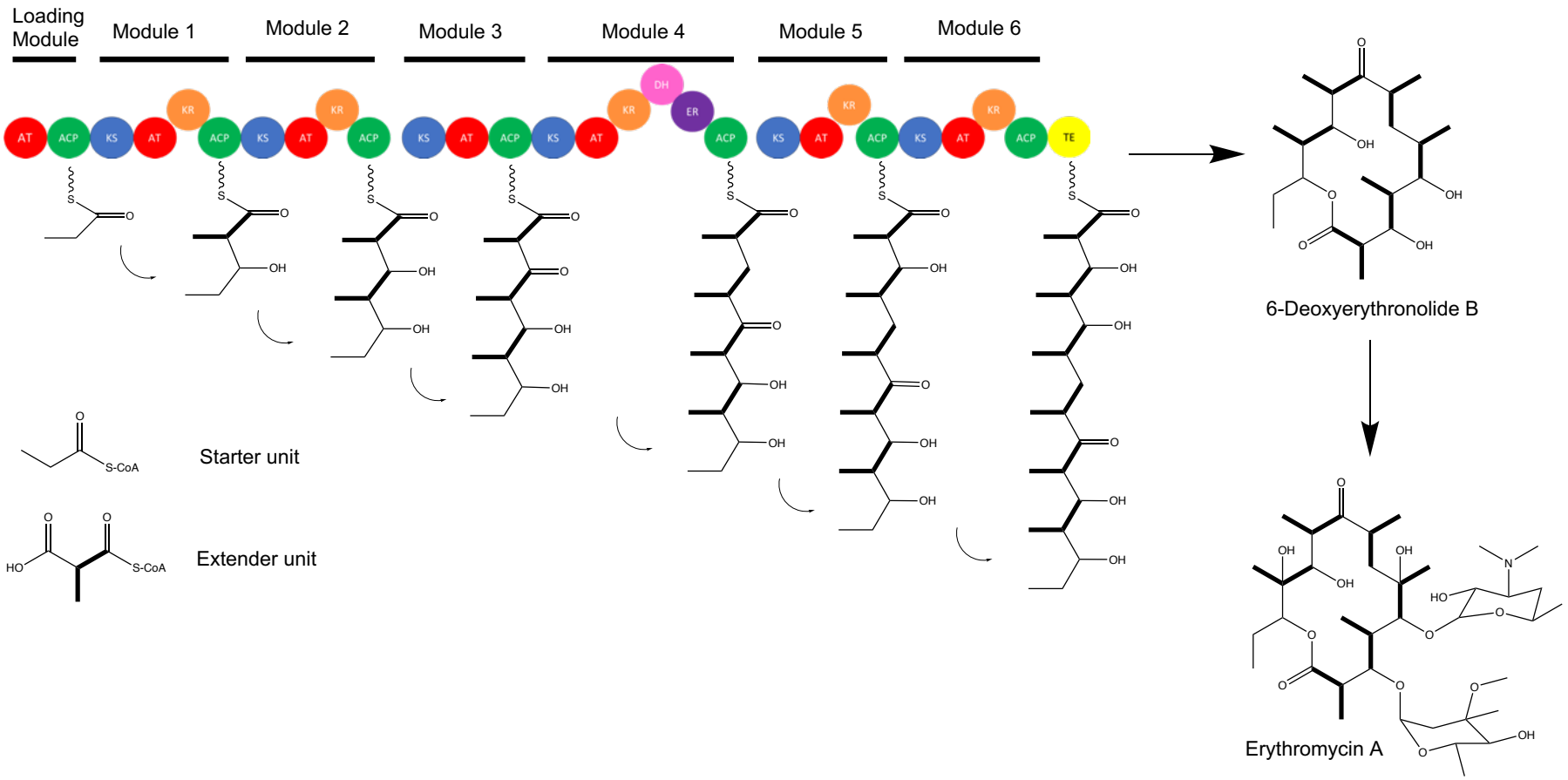


Figure 1.1: Erythromycin A is synthesised by a type I PKS consisting of six modules that each incorporate a methylmalonyl-CoA extender unit onto the growing polyketide chain. Each extender unit is loaded onto the acyl carrier protein (ACP) by an acyltransferase (AT). A ketosynthase (KS) domain then forms the carbon-carbon bond between neighbouring extender units which can be further modified by ketoreductases (KR), dehydratases (DH) and enoylreductases (ER) to change the oxidative state of the keto group. Carbons in bold represent the extender unit incorporation. Figure adapted from (Chan *et al.*, 2009).

In contrast, type 2 PKS gene clusters usually consist of a separately encoded ACP enzyme and the two subunits of the heterodimeric KS, KS_α and KS_β, that are used iteratively to generate a polyketide chain (Fischbach and Walsh, 2006). KS_α is responsible for the Claisen condensation between extender units while KS_β controls the number of units incorporated, therefore it is often referred to as the chain length factor (Hertweck *et al.*, 2007; Chen, Re and Burkart, 2018). Additional ketoreductases, cyclases and aromatases can then act on the resulting poly-β-keto intermediate to generate a range of polyphenolics that can be tailored by oxygenases, methyltransferases, glycosyltransferases and halogenases that are also encoded within the BGC, much like in type 1 polyketide biosynthesis (Hertweck *et al.*, 2007; Rutledge and Challis, 2015). Assembly by these pathways is considered more complex than Type 1 PKSs as the process relies on multiple enzymes instead of a single modular protein, making structural predictions difficult. Furthermore, the highly unstable poly-β-keto intermediates that are generated are often spontaneously cyclized and therefore cannot be isolated for characterisation (Hertweck *et al.*, 2007). Type 3 PKS gene clusters use free CoA-linked thioester substrates without the need for an ACP (Ray and Moore, 2002).

Type 2 PKS BGCs are employed by actinomycete bacteria to create a diverse family of polyphenolic compounds with a wide range of bioactivities (**Figure 1.2**). Perhaps the most intensively studied T2PKS gene cluster is the actinorhodin BGC. Actinorhodin is a blue-pigmented benzoisochromanequinone antibiotic produced by the model actinomycete *Streptomyces coelicolor* (Okamoto *et al.*, 2009; Chen, Re and Burkart, 2018). The actinorhodin BGC was the first cluster of genes to be cloned and heterologously expressed in another organism. It was this work that laid the groundwork for all natural products biochemistry being conducted today; the fact that a non-producing strain of *Streptomyce*s could produce the compound without lethality on introduction of the 35 kb chromosomal fragment meant that all the genes for biosynthesis, regulation and host-resistance were all clustered together (Malpartida and Hopwood, 1984). Actinorhodin is often studied today as the model secondary metabolite from *Streptomyces* because its distinct blue colour means that production levels can easily be quantified. The tetracyclines also originate from

T2PKS pathways in multiple *Streptomyces* species and are one of the most important antibiotic classes to modern medicine. These broad spectrum antibiotics inhibit both Gram-positive and Gram-negative pathogens by blocking the attachment of tRNA to the ribosomal acceptor (A) site, thereby preventing protein synthesis (Chopra and Roberts, 2001). Another example is doxorubicin, an anthracycline made by a T2PKS pathway in *Streptomyces peucetius,* that has potent anticancer activity due to its ability to intercalate DNA. Compounds of this class can also inhibit the activity of DNA topoisomerase enzymes, giving them antibacterial activity as well (Marinello *et al.*, 2018).

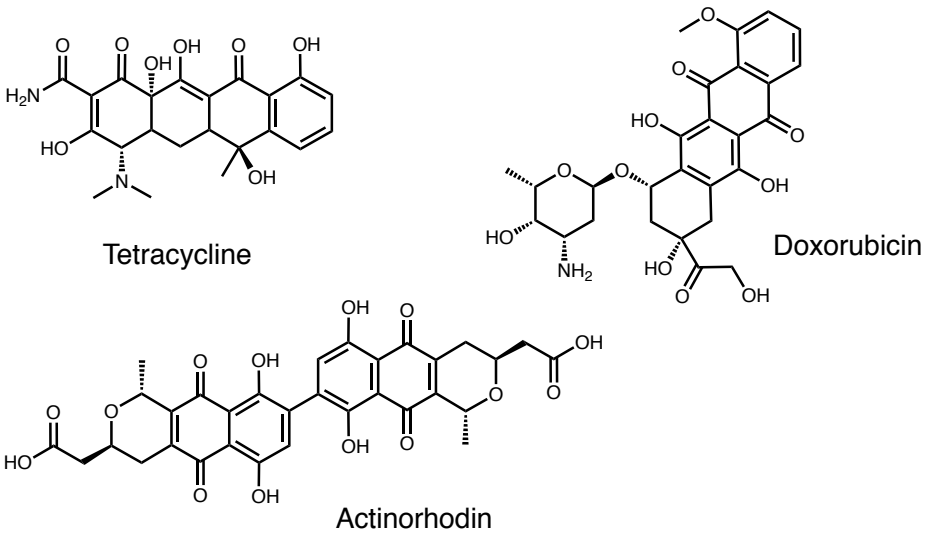


Figure 1.2: Example structures of natural products with antibiotic activity that originate from T2PKS pathways.

1.3.2 Non-ribosomal peptides

NRPSs are large, multifunctional mega-proteins responsible for the biosynthesis of a significant subclass of peptide natural products. Like polyketide synthases, NRPS complexes have a highly organised architecture where individual catalytic domains are responsible for the incorporation and modification of a single monomer unit onto the growing peptide chain. The major difference here is that the monomers are amino acids instead of carboxyl groups (Hur, Vickery and Burkart, 2012; Singh, Chaudhary and Sareen, 2017). Much like polyketide synthases, NRPSs can be classified into three groups. Type A NRPS biosynthesis occurs in a linear manner where the number of modules present in the synthetase determines the number and order of the amino acids in the final peptide product (Hur, Vickery and Burkart, 2012). A minimal module consists of an adenylation (A) domain responsible for selecting the amino acid monomer, a condensation (C) domain which forms the peptide bond between the amino acids and a thiolation (T) domain, also known as the peptidyl carrier protein (PCP) which acts as a carrier and holds the growing peptide chain in place while it is modified by downstream enzymes (Fischbach and Walsh, 2006; Hur, Vickery and Burkart, 2012; Singh, Chaudhary and Sareen, 2017). The process of activation and condensation of amino acids is repeated until the final module containing the thioesterase (TE) domain is reached, where the peptide chain will be released by hydrolysis (**Figure 1.3**) (Singh, Chaudhary and Sareen, 2017). During type B biosynthesis, modules or domains of the synthetase are used iteratively to synthesise the final product, which will consist of multiple repeated peptide sequences. Finally, type C NRPS clusters are non-linear and generate peptide sequences that are independent of the modules present in the synthetase (Hur, Vickery and Burkart, 2012).

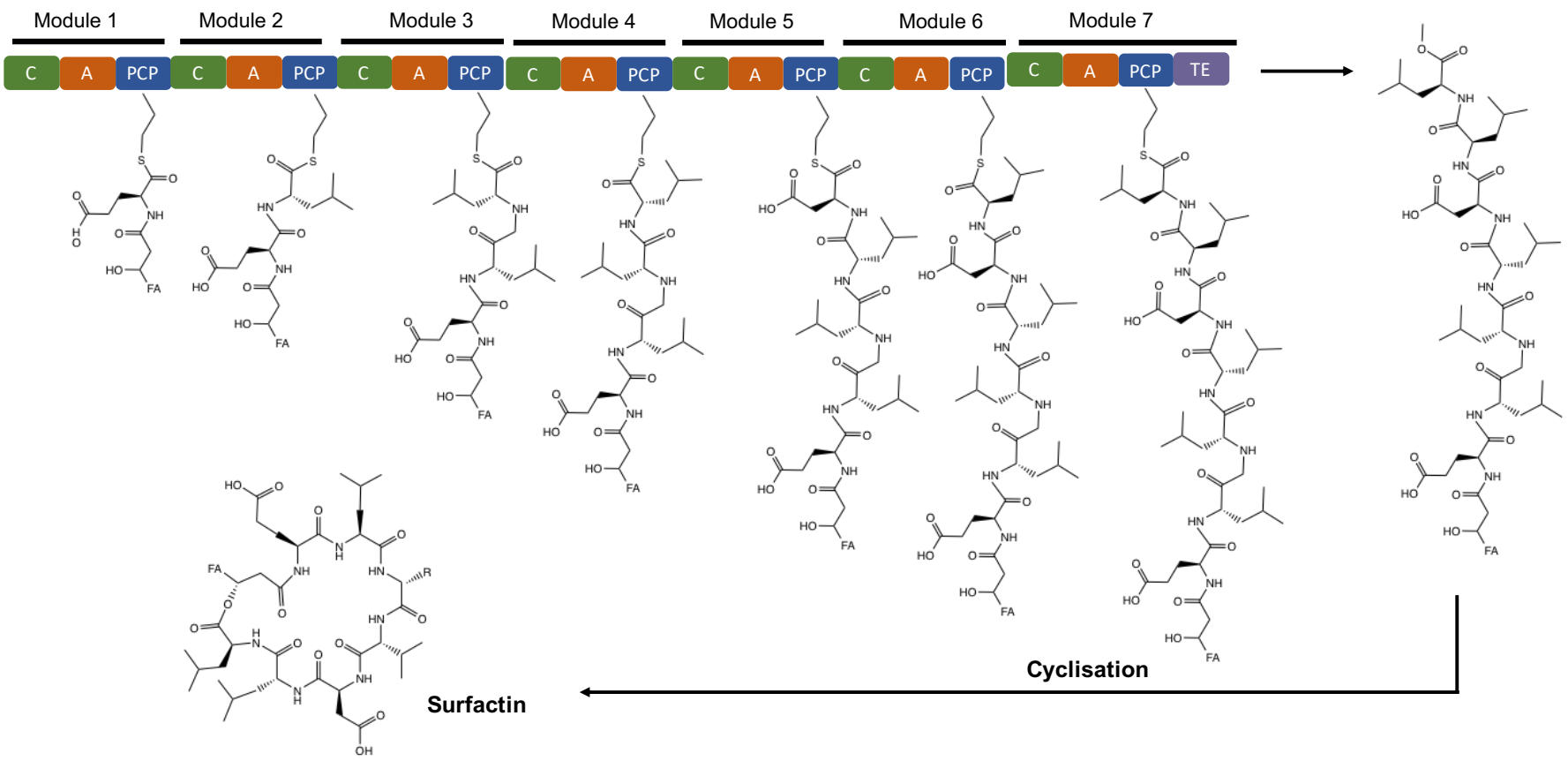


Figure 1.3 NRPSs consist of multiple modules encoding an adenylation (A) domain for selecting the amino acid monomer, a condensation (C) domain that forms peptide bonds between amino acid monomers and a peptidyl carrier protein (PCP) domain to hold the growing peptide product. During this modular peptide biosynthesis, each module adds a single amino acid to the growing peptide chain until the thioesterase (TE) domain is reached. In this example, the cyclic lipopeptide surfactin is synthesised by the modular NRPS encoded by *Bacillus subtilis*. Once released from the NRPS, the peptide chain is cyclised to form the final product. Figure adapted from (Martínez-Núñez and López, 2016).

NRPS natural products also represent a hugely diverse class of chemical structures. In addition to the variety of available A domains, NRPSs can incorporate both L-and D-amino acids into the peptide using the action of epimerization enzymes (Fischbach and Walsh, 2006; Singh, Chaudhary and Sareen, 2017). The presence of D-amino acids provides the stereochemical constraints for proper modification of the peptide chain by downstream tailoring domains. As more amino acids are incorporated into polypeptide chain, cysteine, serine and threonines can be modified by cyclisation domains to form heterocyclic rings which can be further reduced or oxidised to generate additional structural diversity. Tailoring enzymes can also incorporate sugars and fatty acids to the peptide chain, further altering both the structure and the biological activity of the resulting natural product (Hur, Vickery and Burkart, 2012). Some NRPS BGCs have also been shown to diverge from the typical 'rules' described above and instead of following a linear biosynthetic pathway, modules can be skipped or used iteratively to create further diversity between structures. An example is in the biosynthesis of the siderophore coelichelin, from the tri-modular *cch* gene cluster of *S. coelicolor*, where the first module is used iteratively to incorporate two copies of the same residue into the peptide, and the second module is skipped in one of these iterations to form a direct linkage between the amino acid substrates from the first and third modules (Corre and Challis, 2009).

Glycopeptide antibiotics like vancomycin, that are invaluable for the control of Gram-positive infections in the clinic, are synthesised by Type A NRPS BGCs (Felnagle *et al.*, 2008). This diverse group of compounds is characterised by high levels of crosslinking between the amino acids present in the peptide. Glycopeptides inhibit biosynthesis of the bacterial cell wall by interacting with both peptidoglycan and Lipid II. This distinct mechanism of action means that resistance to glycopeptides usually evolves more slowly than other classes of antibiotics as pathogens are unable to respond by upregulating transporters or mutating target enzymes (Yim *et al.*, 2014). Nevertheless, strains of vancomycin-resistant *Enterococci* now present a major clinical challenge, so the discovery of new glycopeptide antibiotics is considered of high importance.

Recently, several novel NRPS natural products have been discovered that represent promising advances in the search for new antibiotics with activity against drug-resistant infections (**Figure 1.4**). For example, daptomycin is a natural product from an NRPS BGC in *Streptomyces roseosporus* initially discovered in the early 1980s (Hur, Vickery and Burkart, 2012; Singh, Chaudhary and Sareen, 2017). The biosynthesis of daptomycin is interesting as it is initiated by the condensation of the N-terminal amino acid, tryptophan, with a long-chain β-OH fatty acid (Fischbach and Walsh, 2006; Wittmann *et al.*, 2008). The presence of the fatty acid side chain means that when calcium ions bind to the cyclic peptide, daptomycin can insert itself into the bacterial cell membrane, perforating it and causing cell death. This mechanism of antibacterial activity is unique and makes daptomycin a valuable treatment against drug-resistant infections, particularly vancomycin-resistant MRSA (Hur, Vickery and Burkart, 2012). The recent discovery of teixobactin also gained much attention from both scientists and the public as it represented a new class of antibiotics. Teixobactin was discovered using the iChip, a new tool used to screen for microorganisms that cannot be grown using standard microbiological approaches, a group that is thought to represent over 99% of all microorganisms on Earth (Nichols *et al.*, 2010; Ling *et al.*, 2015; Piddock, 2015). Teixobactin is a peptide antibiotic consisting of eleven amino acid residues, including seven L-amino acids and four D-amino acids, synthesised by two large NRPS genes in the previously unknown β-proteobacterium *Eleftheria terrae*. Teixobactin shows potent antimicrobial activity against MRSA, vancomycin-resistant *Enterococcus* (VRE), penicillin-resistant *Streptococcus pneumonia*, *Clostridium difficile* and *M. tuberculosis* and does not produce resistant mutants under laboratory conditions. Teixobactin is predicted to inhibit peptidoglycan biosynthesis by binding lipid II, however the mode of action and the potential resistance mechanisms are yet to be extensively investigated (Guo *et al.*, 2018).

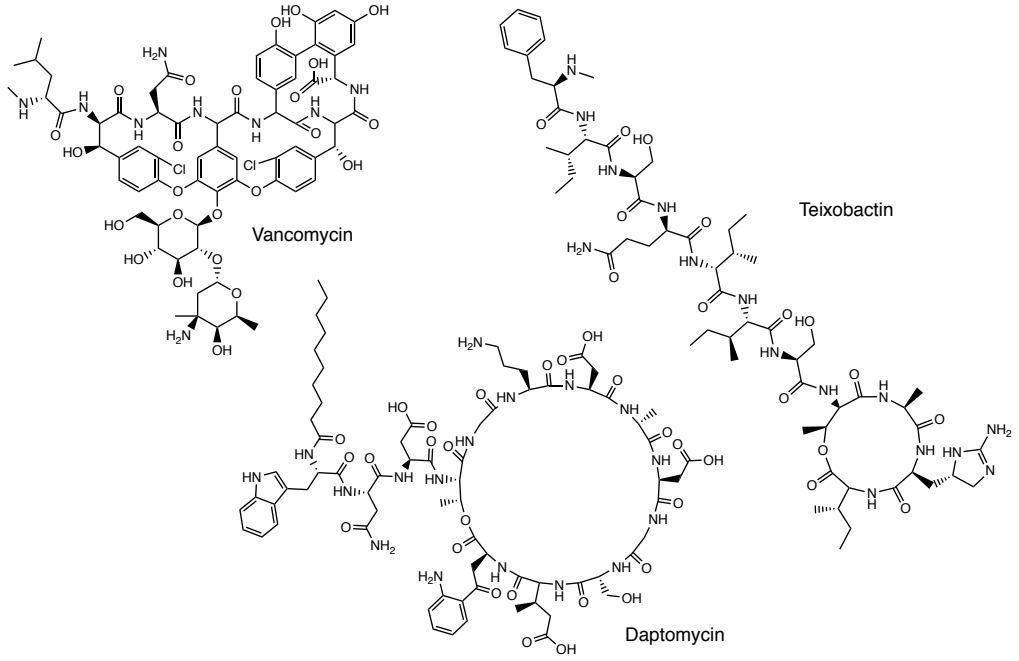


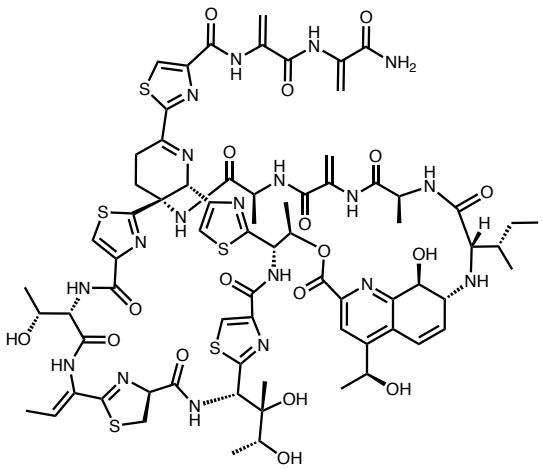
Figure 1.4: Example structures of natural products with antibiotic activity that originate from NRPS pathways.

1.3.3 Ribosomally synthesised natural products

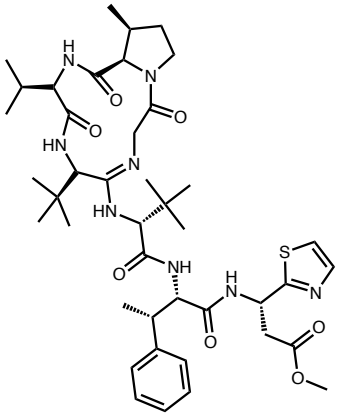
PKS and NRPS BGCs are relatively easy to identify in actinomycete genomes due to their characteristic modular organisation. The increased availability of genome sequence data in recent years has also revealed the range of unexplored biochemistry available from Ribosomally synthesised and Post-translationally modified Peptides, known as RiPPs. The first RiPPs to be identified were the bacteriocins, a diverse family of small peptides produced by more than 99% of bacteria in order to kill or inhibit the growth of other microorganisms during normal environmental competition (Devine, Hutchings and Holmes, 2017). Several key properties of bacteriocins, such as their heat-stability and the fact that they are sensitive to proteases and therefore generally harmless to humans and the environment, mean that bacteriocins have been widely used to treat bacterial infections, aid cancer treatments and extend the shelf-life of food products (Yang *et al.*, 2014). RiPPs are now known to be a structurally diverse group of natural products

and have been the subject of many natural product discovery efforts in recent years. This structural diversity originates from the variety of post-translational modifications that can occur on these peptides and contributes to the wide-ranging activity displayed by these compounds (Ortega and Van Der Donk, 2016).

Most RiPPs are encoded as relatively simple precursor peptides usually consisting of an N-terminal leader region, which contains a recognition site for downstream enzymes, and a C-terminal core which becomes the mature compound after post-translational modification. The leader peptide is removed by proteases, so does not become part of the final product, but the physical separation of the recognition sequence and the core peptide means that a variety of different core peptides can be modified by individual enzymes in a biosynthetic pathway (Ortega and Van Der Donk, 2016; Hudson and Mitchell, 2018). A wide variety of remarkable structures have been described that originate from these ribosomally synthesised peptides. For example, the lasso-peptides are bioactive compounds with a characteristic 'threaded-loop' structure where the C-terminal peptide tail is threaded through a macrolactam ring formed at the N-terminal. Thiopeptides like thiostrepton from *Streptomyces azureus,* are macrocyclic peptides that all contain a central six-membered nitrogen-containing ring and lanthipeptides such as nisin contain the thioether amino acids lanthionine of 3-methyllanthionine formed by the dehydration and subsequent thiolation of selected serine and threonine residues (Ortega and Van Der Donk, 2016; Hudson and Mitchell, 2018). Perhaps the most complex of the RiPPs are the bottromycins, that contain a macrocyclic amidine, several non-proteinogenic amino acids, three β-methylated amino acids and a thiazole. The biosynthesis of the bottromycins is unusual amongst RiPPs as they use a C-terminal follower peptide instead of the N-terminal leader peptide to encode the recognition sequence for post-translational processing. These complex molecules represent a promising new class of antibiotics with broad ranging activity against drug resistant pathogens (**Figure 1.5**) (Crone, Leeper and Truman, 2012; Gomez-Escribano *et al.*, 2012).



Thiostrepton



Bottromycin A2

Figure 1.5: Example structures of ribosomally synthesised natural products with antibiotic activity.

1.3.4 Discovering the natural product antibiotics of the future

Genes encoding the biosynthetic machinery for antimicrobial secondary metabolites are found in many microorganisms, but the actinomycetes are best known for the diversity of molecules they are able to produce. Recent advances in genome sequencing technologies has revealed that most actinomycetes encode many more natural products than previously identified, but only a fraction of these BGCs are expressed under standard laboratory conditions. Even though microbiologists and natural products chemists have been studying the model actinomycete, *Streptomyces coelicolor*, since the 1960s, only five antibiotic gene clusters were characterized in any detail before the early 2000s. When the full genome sequence of the strain was published in 2002, it was revealed the strain actually contained the genetic information for around 30 secondary metabolites (Bentley *et al.*, 2002; Challis, 2014; van Keulen and Dyson, 2014; Ward and Allenby, 2018). This discovery regenerated the process of natural product discovery by introducing an era of 'genome mining'. This term is now used to cover a breadth of secondary metabolite analysis including the identification and structural elucidation of new compounds from a producing organism, the bioinformatic prediction of gene products and pathways and the study of the regulatory control of biosynthetic pathways (O

Bachmann, Van Lanen and Baltz, 2014). The relative explosion of bacterial genome sequencing in recent years has made identifying BGCs encoding for novel natural products easier than ever before. Upon completion of a new genome sequence, bioinformatic tools like AntiSMASH can be used to search for genes that may be responsible for the synthesis of these complex molecules by searching for certain known characteristics of these BGCs (Weber *et al.*, 2015). The focus now is on identifying novel compounds from both old and new bacterial strains with an emphasis on developing more sophisticated screening methods in order to prevent rediscovery (Manteca and Yagüe, 2018).

1.4 The genus *Streptomyces*

Streptomyces are filamentous, Gram-positive bacteria of the order Actinomycetales, whose members are commonly known as actinomycetes. These saprophytic organisms are obligate aerobes and are among the most numerous bacteria found in the soil (Bentley *et al.*, 2002). As a genus, *Streptomyces* dominate the bacterial world in the large variety of secondary metabolites they produce, from signalling molecules to antimicrobials. They are responsible for the production of about 80% of all known bioactive natural products (Demain, 2009). Although *Streptomyces* species have been intensively screened for novel natural products since the Golden Age of antibiotic discovery, efforts intensified after genome sequencing data showed most strains can produce many more antibiotics than previously thought. Sequencing of the relatively large genomes of *Streptomyces* bacteria has shown that each strain contains between 20 and 60 secondary metabolite gene clusters, even though most only produce a few of their encoded secondary metabolites under standard laboratory conditions. This means that the majority of the chemistry available from these organisms is yet to be discovered (Watve *et al.*, 2001; Kemung *et al.*, 2018). The challenge now is to identify novel BGCs and encourage production of the encoded compounds in the laboratory. By drawing connections between secondary metabolites and the gene sequences that encode them, we can gain more of an insight into how these compounds contribute to the basic biology of the producing organisms and the role these molecules play within their natural environments.

Together, these approaches will allow the expression of novel natural products in the laboratory and lead to the discovery of new antimicrobials (O Bachmann, Van Lanen and Baltz, 2014).

1.4.1 *Streptomyces* biology

Most of what we know about the complex, multicellular developmental stages of *Streptomyces* growth comes from the extensive investigation of the model organisms, *S. coelicolor* and, more recently, *S. venezuelae* which sporulates rapidly even in liquid culture. *Streptomyces* species grow as multi-cellular branched filaments that form a mass of fungus-like mycelium. In fact, the name *Streptomyces* means 'twisted fungus' due to their appearance under the light microscope (Hopwood, 2007). *Streptomyces* hyphae originate from the germination of a single spore that produces vegetative mycelium comprised of branching filaments that grow outwards by polar tip extension in order to scavenge for nutrients (Bush *et al.*, 2015). Under nutrient-rich conditions, dense mycelial growth can be achieved as streptomycetes can break down the insoluble remains of other organisms in the soil environment, including lignocellulose and chitin (Bentley *et al.*, 2002). From these vegetative mycelium, reproductive aerial hyphae emerge that are coated in a hydrophobic sheath which allows the hyphae to escape the aqueous environment of the vegetative mycelium. When aerial hyphae are formed, chromosome replication and segregation occurs, followed by cell division and cell wall assembly. This highly synchronised and tightly regulated process differentiates the multi-genomic hyphae into a long chain of uni-genomic pre-spores, separated by multiple septa that form along the chain. On receipt of appropriate environmental signals, these pre-spore compartments will mature and be released as dormant, thick-walled spores, that will only germinate when conditions are suitable for the growth of new vegetative mycelium (**Figure 1.6**) (Bush *et al.*, 2015).

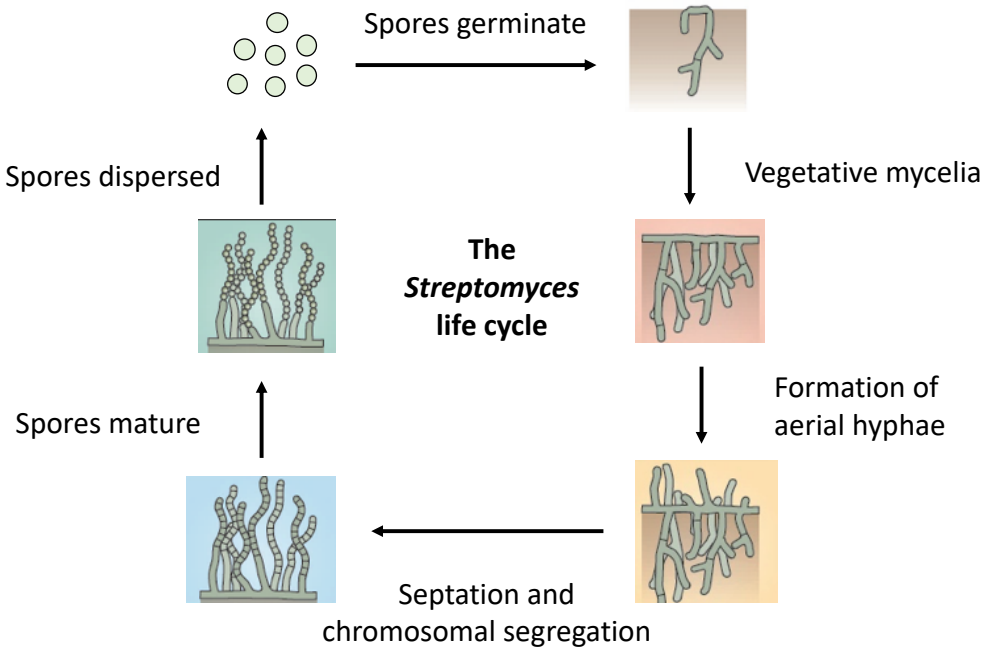


Figure 1.6: Life cycle of *Streptomyces* species. Under nutrient rich conditions, single spores germinate and form vegetative hyphae which grow out into the soil. Upon nutrient starvation, aerial hyphae are then erected, which differentiate into a long chain of pre-spores, each containing a single chromosome. Once the spores have matured, they are dispersed into the environment where they lay dormant until conditions favour germination. Figure adapted from (Bush *et al.*, 2015).

Streptomyces spores are ubiquitous in soils around the world. The soil is a particularly complex and variable environment, made up of highly mixed microbial communities that are all competing for nutrition and physical space. As well as allowing streptomycetes to survive these harsh environments, the formation of desiccation-resistant spores allows the otherwise non-motile mycelium to disperse to more favourable environments under competition and other environmental stressors (Bentley *et al.*, 2002; Bush *et al.*, 2015). The production of a wide array of secondary metabolites also gives *Streptomyces* species a distinct advantage in these harsh conditions. Secondary metabolites are non-essential for growth and are usually made after vegetative growth, as the organism is entering the sporulation stage. Their range of biological activities contributes to the inhibition of other microorganisms present in the surroundings that may be competing for the same nutrients, as well as signalling between microbes or triggering differentiation and transporting metal ions. The ability of streptomycetes to produce antimicrobial secondary metabolites also means these strains are often able to colonize plant roots and protect the host plant against pathogens. In return, the *Streptomyces* bacteria

acquire rich nutrients from the plant, again contributing to a survival advantage (Challis and Hopwood, 2003).

Streptomyces DNA has an unusually high GC content, often over 70%, and is structured in large, linear chromosomes. Their genomes range in size between 7 and 11 Mb and the chromosome contains a large number of genes enabling these bacteria to compete in the dynamic soil environment (Hopwood, 2006). Most of the essential genetic material responsible for cell growth and development is located at the centre of the chromosome, with a remarkable number of genes in this region being highly conserved between *Streptomyces* species and other actinomycetes. In contrast, around 1-2 Mb at either end of the chromosome consists of highly variable DNA sequences, often containing the BGCs for secondary metabolite production. High rates of lateral DNA transfer between *Streptomyces* and other species can take place at these chromosome arms, allowing for the exchange of secondary metabolite biosynthesis genes, resistance genes and regulatory elements and ultimately, the evolution of new BGCs. Horizontally acquired genes that provide a sufficient selective advantage can make their way towards the more stable, central region of the chromosome over time to become a more permanent part of the genome (Chater and Chandra, 2006).

1.4.2 Regulation of antibiotic production in *Streptomyces*

1.4.2.1 Linking growth and development to secondary metabolism

As described above, the production of secondary metabolites by *Streptomycetes* generally begins during the development of aerial hyphae and continues during sporulation. The links between these processes are complex and tightly regulated by multiple gene families. The *bld* gene family of regulators are involved in the precise control of multiple developmental processes, in particular the erection of aerial hyphae, but they have also been shown to play a role in the control of secondary metabolite production (Bibb, 2005). For example, the DNA binding protein BldD interacts with cyclic di-GMP to repress around 170 genes involved in sporulation during vegetative growth including genes encoding for vital components of cell

division, chromosome segregation and septation formation (Bush *et al.*, 2015). The BldD-cyclic di-GMP complex also indirectly affects antibiotic biosynthesis in *Streptomyces* by interacting with other *bld* regulators such as BldC, a small DNA-binding protein that controls transcription of genes within multiple secondary metabolite gene clusters. BldD-cyclic di-GMP can also control the expression of biosynthetic genes directly by binding to promoter regions within BGCs itself (Hengst *et al.*, 2010). The onset of aerial growth can be triggered by variations in multiple conditions such as the availability of nutrients in the environment, changes in metabolism and extra-cellular signals, all of which will also affect secondary metabolism. For example, AdpA is a transcriptional regulator of a large number of sporulation genes that also regulates secondary metabolite BGCs in response to the accumulation of an extracellular hormone-like molecule, γ-butyrolactone (GBL) which is often an indicator of competing microorganisms in the surrounding environment (Chater, 2016).

Although produced during secondary metabolism, the biosynthesis of antimicrobials depends on precursors and cofactors that are derived from the primary metabolism, therefore nutrient availability and secondary metabolism are tightly linked. *Streptomyces* contain a "stringent response" where, during environmental stress, gene transcription is modulated to divert resources away from growth and division and towards amino acid biosynthesis and/or fatty acid biosynthesis to promote survival. This response is regulated by the nucleotide guanosine tetraphosphate (ppGpp), which modulates transcription by binding to RNA polymerase. The concentration of ppGpp increases as nutrients are used up, inducing the transcription of antibiotic biosynthesis genes, therefore increasing the survival advantage of *Streptomyces* against any other microorganisms that might be in the environment and competing for the limited nutrients (Bibb, 2005; Rutledge and Challis, 2015). The availability of carbon, nitrogen and phosphate in the environment have all been shown to affect both morphological differentiation and secondary metabolism in *Streptomyces*. The mechanisms underpinning this remain unclear and are beyond the purpose of this thesis, however, it is important to recognise the importance of global regulatory systems in *Streptomyces* species that can modulate the primary

metabolism and limit the availability of biosynthetic precursors for secondary metabolite biosynthesis (Romero-Rodríguez *et al.*, 2017).

1.4.2.2 Two component systems

Two-component systems (TCS) allow bacteria to change their intracellular processes in response to environmental stimuli by activating or repressing the transcription of specific genes. A classical TCS consists of a sensor histidine kinase, often membrane associated to allow bacteria to respond to extracellular signals, and an associated cytoplasmic response regulator (Jacob-Dubuisson *et al.*, 2018). In these classical TCS, the sensor kinase comprises an extra-cytoplasmic domain that responds to specific environmental stimuli, flanked by two transmembrane helices. On receipt of an activating signal, a conserved histidine residue in the kinase is autophosphorylated using ATP. This phosphate is then transferred to a conserved aspartate in the receiver domain of the corresponding response regulator, which can then alter gene transcription using the effector domain, usually via DNA-binding.

Streptomyces species generally encode a large number of TCSs, even considering their relatively large genome size, presumably to allow them to adapt to their competitive natural environment (Hutchings *et al.*, 2004). Many of the highly conserved TCSs found in *Streptomyces* genomes have been characterised in model organisms such as *S. coelicolor*, *S. lividans* and *S. venezuelae*. During this work, numerous TCSs have been implicated in the control of secondary metabolite production. For example, many secondary metabolites are only produced under phosphate-limited conditions. Phosphate levels are sensed by the membrane-associated PhoR sensor kinase. Under low phosphate conditions, PhoR activates the DNA-binding response regulator PhoP. Mutants lacking *phoP* (or *phoR* and *phoP* in combination) have been shown to overproduce the antibiotics actinorhodin and undecylprodigiosin (Martín, 2004). Similarly, the AfsQ1/Q2 system has been shown to activate the biosynthesis of actinorhodin, calcium-dependent antibiotic, undecylprodigiosin and coelimycin through the binding of the AfsQ1 response regulator to multiple promoter regions within the relative BGCs. AfsQ1 has also been shown to regulate production of the sigma factor SigQ, a negative regulator of

antibiotic production divergently transcribed from *afsQ1/2* (Rodríguez *et al.*, 2013; Chen *et al.*, 2016). Mutation of the CutRS TCS also results in overproduction of actinorhodin, although the mechanism of this is so far unknown (Hutchings *et al.*, 2004; Rodríguez *et al.*, 2013). The signals needed to activate the majority of TCSs also remain unknown; for example, MtrAB is a highly conserved TCS that regulates global expression of genes involved in DNA replication, cell division and antibiotic biosynthesis, but no external stimulus has yet been identified (Som *et al.*, 2017). Usually the genes encoding the sensor kinase and the corresponding response regulator are adjacent in the genome, however this is not always the case, leading to some unpaired sensor kinases and orphan response regulators of unknown function in *Streptomyces* genomes (Hutchings *et al.*, 2004).

1.4.2.3 Cluster situated regulators

In addition to being under the control of global regulators, most secondary metabolite BGCs also contain cluster situated regulators (CSRs) that regulate the expression of genes within the cluster. These CSRs are often under the control of higher regulatory systems like the ones described above (Bibb, 2005). Some clusters, like the actinorhodin cluster, only contain a single regulatory gene, whereas others encode multiple regulators, some of which activate and others that repress production of the metabolite (Aigle and Corre, 2012). Some CSRs can also cross-regulate the expression of genes located in other BGCs elsewhere in the genome (Rodríguez *et al.*, 2013). An example is JadR1 in *S. venezuelae*, the main activator of the jadomycin BGC, which has also been shown to repress chloramphenicol biosynthesis (Chater, 2016). Many CSRs characterised in *Streptomyces* species encode for proteins belonging to the *Streptomyces* antibiotic regulatory protein (SARP) family, all consisting of a winged helix-turn-helix motif responsible for DNA binding, that recognise heptameric repeats within the promoter regions of the genes they regulate (Bibb, 2005; Aigle and Corre, 2012). Other CSRs belong to the LuxR family of proteins and are referred to as large ATP-binding regulators of the LuxR family (LAL) regulators. These regulators contain an N-terminal nucleotide triphosphate (ATP/GTP) binding motif and a LuxR-like helix-turn-helix motif at the C-

terminus for DNA binding (Bibb, 2005; Lu *et al.*, 2017). The FscRI regulator is a LAL regulator situated in the candicidin BGC that controls expression of both the candicidin BGC and the antimycin BGC in *S. albus*. Other common families of regulatory proteins found within *Streptomyces* BGCs include the LysR and TetR families (van der Heul *et al.*, 2018). Some clusters also contain their own cluster-situated TCS that are involved in activating or repressing transcription of genes within that BGC. An examples is the TCS cinKR that controls expression of genes within the cinnamycin BGC in *Streptomyces cinnamoneus* DSM 40646 (O'Rourke, Widdick and Bibb, 2017).

Antibiotics themselves can also act as regulators of secondary metabolism. Many CSRs bind to ligands such as biosynthetic products or intermediates in order to cause changes in the expression of the same or another BGC. For example, JadX, another regulator of the jadomycin BGC can interact with both jadomycins and chloramphenicol and deletion of the *jadX* gene affects production levels of both compounds (Xu *et al.*, 2010; Chater, 2016). MarR regulators, named after the Multiple antibiotic resistance Regulators in *E. coli*, are a family of transcription factors found throughout bacteria; they exist as homodimers where each monomer contributes a winged helix-turn-helix DNA binding motif. MarR proteins often repress transcription of the genes they regulate by binding to palindromic DNA sequences within promoter regions. They are well known to also bind small molecule ligands or phenolic compounds, however, unusually for CSRs, MarR regulators bind DNA and ligands in the same domain. On ligand binding, there is a conformational change in the transcription factor that changes the interaction with DNA and therefore affects target gene expression. Where MarRs are responsible for the regulation of biosynthetic enzymes, the ligand is often the substrate of the enzyme or a closely related compound. MarR regulators are also commonly involved in regulating genes that control the export of antibiotics, such as drug efflux pumps, and therefore the associated ligand is often the molecule required for export (Grove, 2017).

1.5 Tools for genome mining and natural product discovery in *Streptomyces*

Traditional methods of searching for new natural products primarily consists of growing the antibiotic-producing microorganism in pure culture in a laboratory and screening the culture extract for bioactivity. This approach gives low returns and high rates of rediscovery (Rutledge and Challis, 2015). Understanding the biology of antibiotic-producers and the ways in which they regulate the biosynthesis of their secondary metabolites can help to unlock the potential novel biochemistry encoded in cryptic BGCs (Aigle and Corre, 2012). BGCs that are silent under laboratory conditions are more likely to be a source of compounds with novel chemical structures and could therefore be useful in the treatment of drug-resistant infections (Corre and Challis, 2007).

1.5.1 Pleiotropic methods of inducing secondary metabolism

The tight regulatory control of secondary metabolism means that many BGCs are not expressed under laboratory conditions due to missing signals from the surrounding environment. By triggering global changes in gene regulation, the expression of multiple BGCs can be affected in a pleiotropic manner. Although unpredictable, this can allow for the discovery of novel natural products from organisms that are less well-studied and where the regulation of specific BGCs is not well understood. One of the simplest ways to induce global changes in bacterial gene expression is to change the growth conditions; perhaps by using different constituents in the growth media, including or omitting various trace elements, or cultivating the strain with extracts of soil in order to replicate the natural environment (Rutledge and Challis, 2015; Devine, Hutchings and Holmes, 2017). Strains of interest can also be co-cultured with other microorganisms to encourage competition and therefore production of secondary metabolites, for example, cultivation of *S. coelicolor* with the predatory microbe *Myxococcus xanthus* significantly increases the production of actinorhodin (Pérez *et al.*, 2011). Streptomycetes that are co-cultured with mycolic acid-containing bacteria often begin producing secondary metabolites, for example the production of the red antibiotic undecylprodigiosin in *S. lividans* and the discovery of the novel antibiotic alchivemycin A, produced by *Streptomyces endus* in

response to *Tsukamurella pulmonis* (Onaka *et al.*, 2011). Co-culture of actinomycetes with fungal isolates is also particularly successful at inducing the production of novel secondary metabolites. N-acetyl glucosamine is the monomeric form of chitin, a major component of fungal cell walls and insect cuticles which is therefore abundant in the soil. N-acetylglucosamine has been shown to increase the biosynthesis of secondary metabolites via the global regulator DasR (Martín *et al.*, 2010). Adding N-acetylglucosamine to the growth media is an approach that has been used to increase the production of antimicrobials including enhancing actinorhodin production in *S. coelicolor* (Abdelmohsen *et al.*, 2015). Whilst screening a strain under numerous conditions may seem like a daunting task using traditional culturing techniques, modern advances such as the development of microfluidic devices makes antibiotic discovery using these methods more high-throughput than ever before (Baltz, 2018).

Various physical and chemical methods can also be used to induce mutations in producing organisms that can permanently alter the secondary metabolite profile. Chemical mutagenesis using N-methyl-N'-nitro-N-nitrosoguanidine (MNNG), ethyl methanesulfonate (EMS), or nitrous acid (NA) can generate strains with high levels of secondary metabolite production, as well as physical mutagenesis by ultraviolet light or X-rays (Baltz, 2015). The use of histone deacetylase inhibitors activates gene expression in fungi by unwinding the DNA from the histones and increasing transcription. This method has also been used in actinomycetes such as *Pseudonocardia* and *Streptomyces* strains from fungus-farming ants where the addition of sodium butyrate to the growth media induced the expression of antifungal gene clusters, however the activation mechanism in these bacteria is unknown (Moore *et al.*, 2012; Seipke *et al.*, 2012). An alternative approach is to select for mutations in genes encoding RNA polymerase and ribosomal proteins in order to upregulate the transcription and translation of cryptic BGCs. Mutations in the RNA polymerase beta-subunit (*rpoB*) and the ribosomal S12 protein (*rpsL*) can be selected for by culturing strains on either rifampicin or streptomycin, respectively (Ochi and Hosaka, 2013). In a study of over 1000 actinomycetes isolated from soil, 43% of the non-producing *Streptomyces* species gained the ability to produce

antibiotics after a selection on either rifampicin or streptomycin, and by sequentially introducing different mutations in the same strain, up to 180-fold higher levels of secondary metabolite production have been seen in some organisms (Wang, Hosaka and Ochi, 2008; Hosaka *et al.*, 2009). Mutation of global regulators can also be effective at increasing secondary metabolism; deletion of *mtrB* results in increased chloramphenicol production in *S. venezuelae* and deletion of *dasR* in *S. coelicolor* induces a cryptic T1PKS cluster (Aigle and Corre, 2012; Som *et al.*, 2017). However, modifying global regulatory machinery can have detrimental effects on the growth and development of the producing organism and may therefore not be optimal for industrial-scale antibiotic production.

1.5.2 Genome editing of BGCs for discovery of novel natural products

Although effective, pleiotropic methods of inducing secondary metabolism are unpredictable and often result in the production of a complex mixture of metabolites, making isolation and structural identification of novel compounds difficult. Ideally, clusters of interest would be studied in isolation to build an accurate picture of the molecules synthesised by the specific pathway. With the increased availability of genomic data, it is possible to estimate the structure of some secondary metabolites by looking at the modules encoded in the predicted BGCs. However, caution must be taken as even high-quality genomes can contain errors in these highly repetitive regions of sequence. Furthermore, it is extremely difficult to predict any structural features of products from iterative pathways due to the nature of their biosynthesis, so it is usually necessary to express the genes and isolate the metabolite(s) produced (Rutledge and Challis, 2015). Because the genes for biosynthesis, regulation and transport of antibiotics are very often located together in the genome, one experimental approach to determine the product of a novel BGC is to heterologously express the pathway in an organism that has been optimised for the production of secondary metabolites. This can be beneficial if the BGC of interest is in an environmental isolate where not much is known about the optimal culture conditions, or where the producing organism is never isolated, as in the case of metagenomics studies (Feng, Kallifidas and Brady, 2011; Baltz, 2015). Multiple

heterologous hosts have been generated in model *Streptomyces* strains, including *S. coelicolor*, by deleting BGCs for known antibiotics that are produced *in vitro* to remove competing sinks of carbon and nitrogen sources required for antibiotic biosynthesis, and introducing mutations in the ribosomal and RNA polymerase machinery to pleiotropically increase secondary metabolism (Gomez-Escribano and Bibb, 2011). Gene clusters of interest can either be captured using TAR cloning or artificial chromosome libraries for heterologous expression within the optimised host (Nah *et al.*, 2017). Whilst this approach has been effective for the discovery of multiple novel natural products, production levels are often low and many clusters are not expressed at all outside of the native host (Gomez-Escribano and Bibb, 2011; Baltz, 2015). Furthermore, this approach does not allow for methods of cross-cluster regulation to be investigated, as other native BGCs are not present in the heterologous host (Mclean *et al.*, 2019). For these reasons, it is often preferable to genetically modify the regulatory machinery within the producing organism in order to upregulate BGC expression.

Until recently, genome editing in streptomycetes was commonly achieved through homologous recombination, a method that is considered both time- and labour-intensive in comparison with the genetic modification of other microorganisms, especially if an unmarked mutant is required (Tong, Weber and Lee, 2018). Recently however, highly-efficient tools such as CRISPR/Cas9-mediated genome editing have been developed for *Streptomyces* species. The CRISPR/Cas system is a form of acquired immunity in bacterial cells where Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) sequences from viruses that have previously infected cells are used to recognise foreign DNA during subsequent infections so that they can be destroyed by nucleases. In recent years, this system has been developed for generating double stranded breaks in DNA to facilitate targeted genome editing in multiple species (Jiang *et al.*, 2013). The pCRISPomyces-2 system is one example of a CRISPR/Cas9 system that has been developed specifically for genome editing in *Streptomyces* species. CRISPR/Cas9-mediated genome editing requires three components; a Cas9 nuclease enzyme, a short CRISPR RNA (crRNA) to confer target site specificity and a trans-activating CRISPR RNA (tracrRNA) to process crRNA and

recruit the Cas9 enzyme. The pCRISPomyces-2 system uses a codon-optimised Cas9 nuclease from *Streptococcus pyogenes* to generate a double-stranded break at a specified DNA site, known as a protospacer. Any site can be targeted, so long as the required protospacer-adjacent motif (PAM) sequence is present at the 3' end. For the *S. pyogenes* Cas9, this PAM sequence must be NGG, where N is any nucleotide. As *Streptomyces* genomes are GC rich, this motif is usually present within a target DNA sequence, making this system highly suitable for use in these organisms. The benefits of pCRISPomyces-2 are that the crRNA and tracrRNA have been fused into a single synthetic guide RNA (sgRNA) into which the protospacer can be assembled, and the plasmid also contains sites for integration of editing templates for repair by homologous recombination, allowing for highly efficient genome editing in *Streptomyces* species (**Figure 1.7**) (Cobb, Wang and Zhao, 2015; Alberti and Corre, 2019).

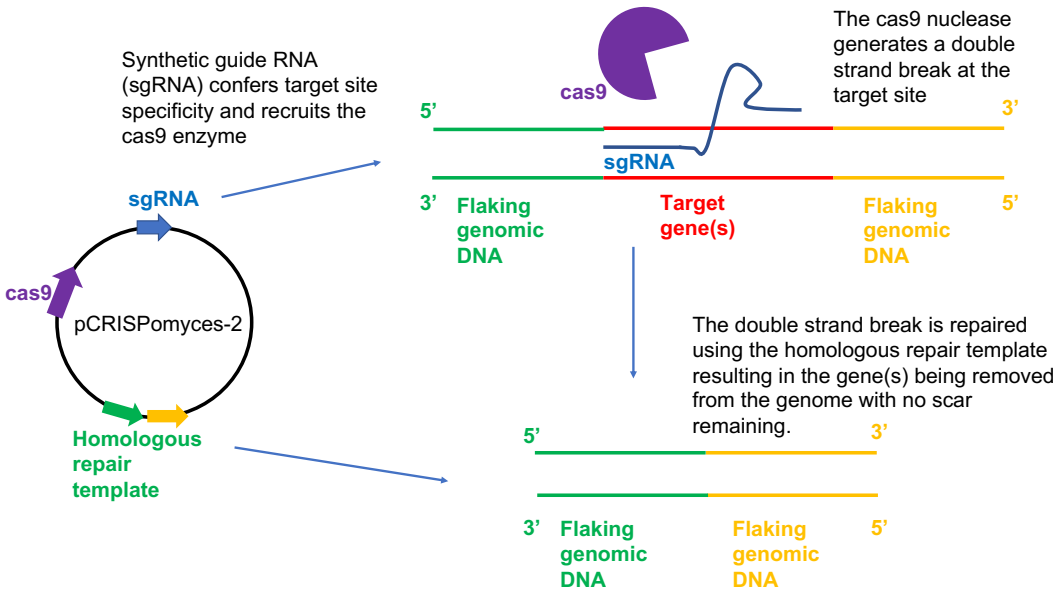


Figure 1.7: The pCRISPomyces-2 system was developed by Cobb *et al*. for specific gene editing in *Streptomyces* species (Cobb, Wang and Zhao, 2015). The synthetic guide RNA (sgRNA) confers target specificity and recruits the Cas9 nuclease which results in a double strand break. One of the main advantages of the pCRISPomyces-2 plasmid is that homology repair arms can also be incorporated into the plasmid so that specific gene edits can be made. When gene deletions are required, no scar or resistance marker is left behind in the genome, meaning that clean mutants can be generated and used for a variety of downstream applications.

Using these new genetic tools, deletion or disruption of other competing pathways in a *Streptomyces* genome can be achieved in order to increase the flux of carbon through the desired BGC and activate the production of a particular natural product (Gomez-Escribano and Bibb, 2014). Additionally, cluster situated regulators can be mutated to switch on expression of the BGC of interest, for example by deleting pathway repressor genes or overexpressing activator genes. A new class of macrolide antibiotics, the stambomycins, were discovered in this way after the overexpression of a LAL regulator from within the previously cryptic pathway in *Streptomyces ambofaciens* (Laureti *et al.*, 2011). The ability to genetically manipulate microbial genomes also allows for the production of novel natural products via metabolic engineering (Corre and Challis, 2009). The modular architecture of many natural product BGCs also means that enzymatic domains can be removed, exchanged or fused to generate novel analogous of existing compounds (Hur, Vickery and Burkart, 2012; Baltz, 2015). This was recently exemplified by the accelerated evolution of the modular PKS BGC that led to the production of novel rapamycin analogues in *Streptomyces rapamycinicus* (Wlodek *et al.*, 2017).

The development of CRISPR for use in *Streptomyces* means that efficient genetic engineering can be achieved in both model organisms and environmental isolates alike, provided a good quality genome sequence is available. Whilst it should be noted that some species remain genetically intractable even with CRISPR, and the deletion of large clusters of $\geq$50 kb and above can still be challenging, these genetic techniques have opened up numerous opportunities for natural product discovery from previously inaccessible sources (Baltz, 2015). For example, the anti-tubercular compound, scleric acid, was recently discovered from the genetically intractable *Streptomyces sclerotialus* by capturing the cluster using TAR cloning, inserting the cloned cluster into a heterologous expression host and de-repressing the pathway by inactivating the repressor *sclM4* using CRISPR (Alberti *et al.*, 2019). These developments mean that we are no longer restricted to mining the cryptic BGCs of model streptomycetes but can also explore the abundant biochemistry available in novel species isolated from the environment.

1.6 Exploring novel environments to find new antibiotic producers

Analysis of biosynthetic genes present within metagenome libraries shows that even ecologically similar environments with similar phylum distributions at the 16S level contain hugely diverse secondary metabolite potential when separated geographically (Reddy *et al.*, 2012). This means that by isolating bacteria from multiple environments across various geographic locations, we can significantly increase the chances of discovering new species with novel secondary metabolite profiles, reducing rates of rediscovery substantially. Unprecedented numbers of antibiotic producing actinomycetes are found not only in soils around the world but in deserts, marine sediments and fresh water ecosystems. Due to the perceived challenges for microbial growth in such harsh environments, exploration of these niches was largely neglected during the Golden Age of antibiotic discovery (Devine, Hutchings and Holmes, 2017; van der Meij *et al.*, 2017). New genera of actinomycetes such as *Salinospora* and *Saccharomonospora* have recently been isolated from marine environments and remain a promising source of novel natural products, such as the cytotoxic salinosporamides and the lipopeptide antibiotic taromycin (Williams *et al.*, 2005; Reynolds *et al.*, 2018). Novel and distinct *Streptomyces* isolates have also been cultured from desert samples such as the chaxalactin- and chaxamycin-producing *Streptomyces leeuwenhoekii*, and *Streptomyces asenjonii* which produces antibiotics with activity against MRSA (Busarakam *et al.*, 2014; Goodfellow *et al.*, 2017).

16S analysis shows that most environmental samples contain thousands of unique bacteria, however most are not culturable using standard microbiological techniques, meaning their biosynthetic potential remains largely unexplored. The recently developed iChip allows growth components and signalling molecules from the natural environment into small growth chambers via diffusion for improved culture and isolation of bacteria from soil and aquatic sediments (Nichols *et al.*, 2010). Culture-independent methods are also becoming increasingly successful, particularly the use of metagenomics, which allows genetical material to be cloned directly from all varieties of environmental samples, and BGCs of interest can be

heterologously expressed in laboratory-grown organisms (Milshteyn, Schneider and Brady, 2014). The malacidins are antibiotics that are broadly active against Gram-positive bacteria including MRSA. The producing organism of the malacidins has so far not been identified; instead, metagenome libraries from diverse soil samples were screened for NRPS clusters containing conserved Asp-X-Asp-Gly motifs that are characteristic of calcium-binding antibiotics like daptomycin (Hover *et al.*, 2018). Whilst this approach will allow for the discovery of many more novel natural products, it is predicted that most unculturable bacteria have small genomes of less than 3.5 Mb, whereas the most prolific antibiotic producers are known to have large genomes. Focussing on natural product discovery from novel but culturable bacteria with large genomes is predicted to identify greater numbers of novel natural products in the near future (Baltz, 2018).

Increasingly, antibiotic-producing actinomycetes are being recognised for their important interactions with higher eukaryotes and these previously neglected environments represent a particularly promising source of novel natural products. Although some *Streptomyces* strains are plant pathogens, many can act as beneficial endosymbionts. The secondary metabolites produced by *Streptomyces* in plant roots have been shown to have plant growth promotion properties, as well as antimicrobial activity against prominent plant diseases and may therefore serve as promising biocontrol agents in the future as well as a source of novel natural products (Vurukonda, Giovanardi and Stefani, 2018). *Streptomyces* species also associate with marine sponges and multiple novel strains with significant biosynthetic potential have been isolated from them (Khan *et al.*, 2011; Li *et al.*, 2011). Some actinomycetes, including *Streptomyces* species, are also known to form mutualistic symbioses with insects. The antennal glands of female digger wasps are populated with antibiotic-producing *Streptomyces* that are applied to the surface of larvae to protect them from infection and chemical analysis of the secondary metabolites produced by the symbiotic *Streptomyces* species reveals a broad range of bioactivity (Poulsen *et al.*, 2011; Nechitaylo, Westermann and Kaltenpoth, 2014).

Some insects cultivate a fungal food source by foraging for plant material and many who display this behaviour are known to make use of the natural products produced

by streptomycetes and other actinobacteria to help them in this mutualism. The southern pine beetle, *Dendroctonus frontalis* grows its symbiotic fungus in the hollowed out bark of pine trees and recruits antibiotic producing actinomycetes to help out-compete antagonistic fungi which threaten the survival of its fungiculture (Scott *et al.*, 2008). Termites also cultivate a fungus, known as *Termitomyces*, and termite nests are also associated with antibiotic producing *Streptomyces*, however the nature of this symbiosis is less well characterised (van der Meij *et al.*, 2017). One of the most well studied insect-fungiculture systems is that of the attine leafcutter ants. Found in central and southern America, these ants cultivate the *Leucoagaricus gongylophorus* fungus as their primary food source by scavenging for and breaking down plant material from their surroundings. The *Leucoagaricus* produces nutrient-rich fruiting bodies, or gongylidia, that the ants can digest. In return, the ants tend to the *Leucoagaricus* fungus and protect it from pathogens. *Escovopsis* species are specialized pathogens of the attine ant fungus garden and infection can cause colony collapse within just two days (Currie *et al.*, 1999). To combat this, the ants have evolved several grooming techniques to weed out infected parts of their fungal cultivar, but more importantly, they have also co-evolved a secondary symbiosis with antibiotic producing actinomycetes that cover their cuticles. A single strain of *Pseudonocardia* is vertically transmitted by the queen to all new worker ants and *Streptomyces* species are acquired from the ants' environment. These bacteria cluster around crypts in the cuticle that have exocrine glands below, suggesting that the ants can provide some sort of nutritional benefit to the actinomycetes that reside there (**Figure 1.8**) (Currie *et al.*, 2006).

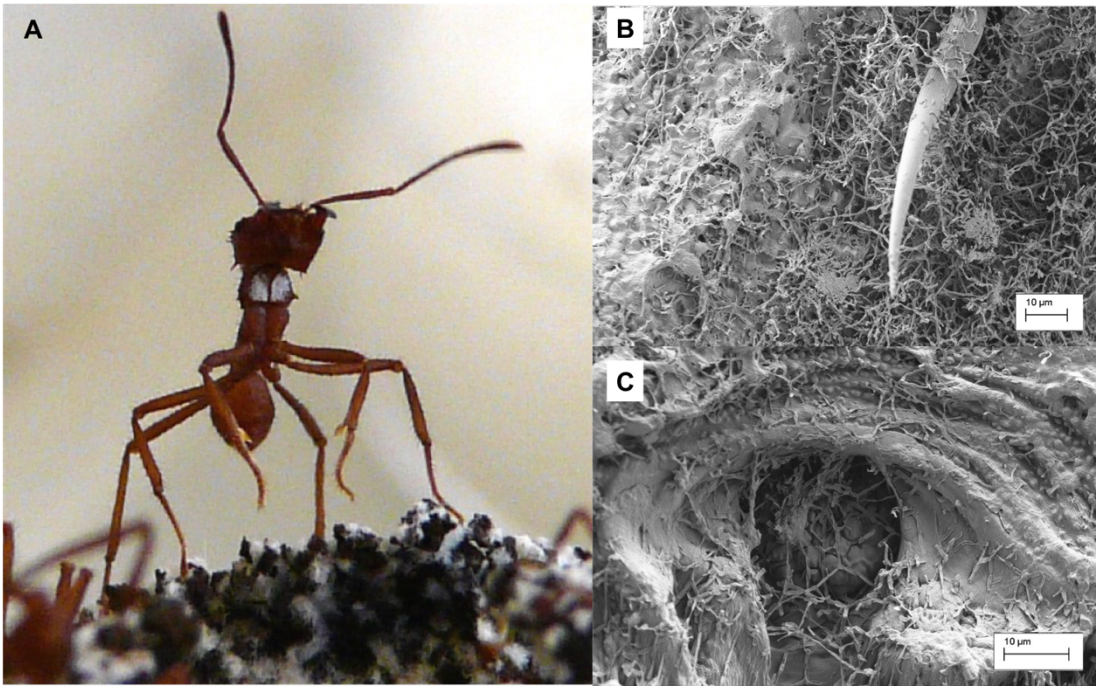


Figure 1.8: A) Attine leafcutter ants have antibiotic-producing actinobacteria on their cuticle that help protect the symbiotic food fungus, *Leucoagaricus,* from infection. B) SEM imaging shows the presence of filamentous actinomycetes such as *Pseudonocardia* and *Streptomyces* on the ant cuticle. C) The actinobacteria in the ant microbiome cluster around crypts on the ants' cuticle; it is presumed that glands below these crypts provide nutritional benefit to the actinomycetes that might also encourage the production of secondary metabolites. Images courtesy of Dr Neil Holmes (formerly Hutchings Laboratory) and Sarah Worsley (Hutchings Laboratory). Scanning electron microscopy conducted by Kim Findley (JIC Imaging Facility).

The strains associated with the attine leaf-cutter ants are capable of producing multiple antimicrobials, many of which have potent activity against the *Escovopsis weberi* pathogen *in vitro* (Currie, Bot and Boomsma, 2003; Barke *et al.*, 2010). *Streptomyces albus S4,* for example, produces the antifungals candicidin and antimycin in addition to encoding for several other NRPS and PKS compounds within its genome (Seipke *et al.*, 2011). This multi-drug approach could explain why the ants have never faced an antimicrobial resistance problem like the one humans are currently facing, even though they are predicted to have evolved this symbiotic relationship over 50 million years ago. These insect microbiomes are evolutionarily primed for the production of antimicrobials and recent data suggest that strains of *Streptomyces* from these systems are often more promising as a source of novel natural products compared to their soil-dwelling relatives (Chevrette *et al.*, 2019). Furthermore, recent work has also shown that it is not just the bacteria in this system that produce specialised metabolites; the *Escovopsis* pathogen also produces

compounds that inhibit the antibiotic producing bacteria and worker ants, implying that this complex system is the subject of a competitive evolutionary arms race which likely increases the rate of evolution of novel natural product BGCs in microorganisms present (Heine *et al.*, 2018).

1.7 *Streptomyces formicae*

Some species of plant-ants have also been shown to use fungiculture, although the system is less evolutionarily developed than the attines. *Tetraponera* ants colonise the African *Acacia drepanolobium* plant (known as swollen thorn acacia) which are common in sub-Saharan Africa. These ants cultivate a fungus of the *Chaetomium* species inside the domatia of the acacia thorns and in return, provide the host plant with some protection against herbivores. Until recently, this system had not been investigated as a source of novel antimicrobials, as the focus had mostly been on the attine ants. Sequencing of DNA from the bacterial community on the surface of *Tetraponera penzigi* ants from Kenya, as well as the bacteria in the domatia of the host plant, shows members of various phyla are present including Proteobacteria, Firmicutes and Bacteroidetes, but Actinobacteria with promising antimicrobial activities could also be easily isolated (Seipke *et al.*, 2013). One such isolate, named *Streptomyces formicae*, displayed a unique range of antimicrobial activity against drug-resistant bacteria and fungi, including MRSA and *Lomentospora prolificans* (**Figure 1.9**). Based on sequence analysis of six phylogenetic markers (16S RNA, *atpD, rpoB, gryA, recA,* and *trpB*), *S. formicae* is thought to be of unique lineage, with its closest relative being *Streptomyces* sp. NRRL S-920, a soil isolate of unknown origin (Qin *et al.*, 2017).

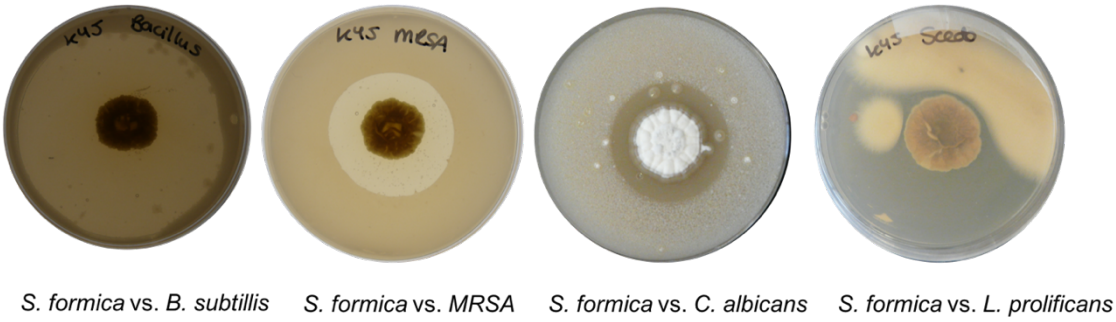


Figure 1.9: The new species, *Streptomyces formicae*, isolated from the fungus-farming plant ant, *Tetraponera penzigi*, has potent antibacterial and antifungal activity, including against drug resistant strains. The bioassays imaged above were carried out by Rebecca Devine during this thesis project.

The genome of *S. formicae* was sequenced using the PacBio and 454 platforms, with Illumina MiSeq to correct for base changes, revealing a 9.6 Mbp linear chromosome with 71.38% GC and 8162 protein coding sequences. Analysis by AntiSMASH version 4.0 predicts the presence of 34 secondary metabolites within the *S. formicae* genome, however, manual inspection reveals that many of these predicted clusters actually contain the biosynthetic genes for multiple distinct compounds. We therefore estimate that this talented strain in fact contains at least 45 secondary metabolite BGCs, almost double the average for a strain of *Streptomyces*, and many look as though they may produce novel compounds (**Table 1.1**) (Holmes *et al.*, 2018). There are multiple PKS, NRPS and RiPP BGCs predicted within the genome sequence, including a bacteriocin, a glycopeptide-like antibiotic and some terpenes which all may be responsible for the production of antimicrobial secondary metabolites. Siderophores and osmolyte BGCs are also present for example those encoding for desferrioxamine and ectoine, which are likely to produce compounds that help survival during environmental stress. The *S. formicae* genome also contains a geosmin synthase, a metabolite produced by all known *Streptomyces* species. As with other strains of *Streptomyces*, these 'essential' and well-conserved BGCs are located towards the centre of the genome, whereas BGCs for antimicrobials sit on the chromosome arms. From this analysis, it was not possible to identify which BGCs are responsible for the antibacterial and antifungal activities seen in the bioassays above.

Table 1.1: AntiSMASH (version 4.0) analysis of the *S. formicae* genome predicted the presence of 34 BGCs, however, manual inspection revealed that some of these predicted gene clusters are likely to contains two or more BGCs that make distinct products. We therefore predict that *S. formicae* contains at least 45 BGCs. Where AntiSMASH predicted clusters are likely to encode more than one product, they are annotated a, b, c and d (Holmes *et al.*, 2018).

BGC	BGC type	Notes
1	Other	-
2	RiPP (lantipeptide)	Single A-gene
3	Mixed NRPS-Type 1 PKS	transAT type PKS module
4a	RiPP (lantipeptide)	Two A-genes
4b	NRPS	Two siderophore-like ORFs
4c	Type 3 PKS	Phloroglucinol synthase-like
4d	Possible RiPP	Two radical SAM genes
5	NRPS	Probable pentapeptide
6a	NRPS	Telomycin-like BGC
6b	Terpene	2-methyl isoborneol BGC
6c	Aminoglycoside	-
6d	NRPS	Probable decapeptide
7	NRPS	Probable pentapeptide
8	Mixed Type 1 PKS-NRPS	-
9	Mixed RiPP (lantipeptide)-Type 1 PKS	Abyssomicin-like BGC
10	Terpene	-
11	Ectoine	Ectoine BGC
12	Mixed Type 1 PKS-NRPS-RiPP	-
13	RiPP	SapB BGC
14	Melanin	-
15	Siderophore	Desferrioxamine B BGC
16	RiPP	-
17a	NRPS	Griseobactin-like BGC
17b	Laspartomycin	Probably tridecapeptide

18	Mixed NRPS-FAS/PKS	-
19	Terpene	Pentalene synthase-like
20	Terpene	Albaflavenone BGC
21	Siderophore	Aerobactin like
22	Butyrolactone	-
23	Bacteriocin	-
24	NRPS	Calcium dependent antibiotic-like lipo-hexapeptide
25	Terpene	Geosmin BGC
26	Type 2 PKS	-
27	NRPS - Type 1 PKS	-
28	Terpene	Hopene BGC
29a	RiPP (lantipeptide)	Single A-gene
29b	Mixed Type 3 PKS-FAS	-
29c	RiPP (lassopeptide)	Single precursor gene
30	Type 1 PKS	Polyene-like, probable octadecaketide
31	NRPS	Glycopeptide antibiotic like BGC
32	NRPS	-
33	Type 1 PKS	Lasaloic acid, probable undecaketide
34a	NRPS	Single module, terminal thioester reductase domain
34b	Clavam (beta-lactam)	Clavam-like
34c	Type 1 PKS	-

1.8 Novel antibiotics from *Streptomyces formicae*

Bioassay guided fractionation of culture extracts of *S. formicae* led to the isolation of both antibacterial and antifungal fractions (**Figure 1.10**). Although purification of the antifungal compounds was possible, structural elucidation has proved challenging. In contrast, it was possible to isolate and purify the compounds from the antibacterial fraction, and the structures of these were solved using nuclear magnetic resonance

(NMR) by Dr Zhiwei Qin (post-doctoral researcher, Wilkinson Laboratory). This revealed thirteen new antibacterial type 2 polyketide natural products from *S. formicae* (Qin *et al.*, 2017). These new compounds were classified into two groups; the first group of compounds (**1-3**) were named fasamycins C-E, due to their structural similarity to fasamycins A and B described previously (Feng, Kallifidas and Brady, 2011). The remaining compounds (**4-13**) are significantly modified compared to the fasamycins and were named the formicamycins. The formicamycins are pentacyclic compounds that can be halogenated at up to four positions on the carbon backbone. In addition, supplementation of the *S. formicae* growth media with sodium bromide resulted in the incorporation of bromine to yield three additional formicamycin congeners (**14-16**), further increasing the novel biochemistry already identified from this strain (**Figure 1.11**).

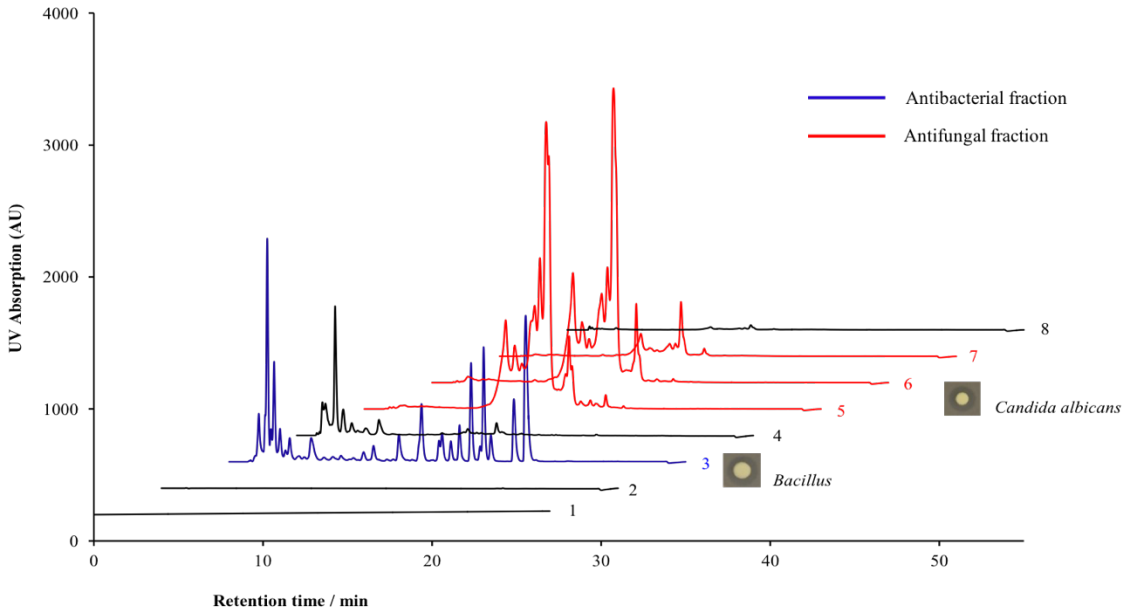
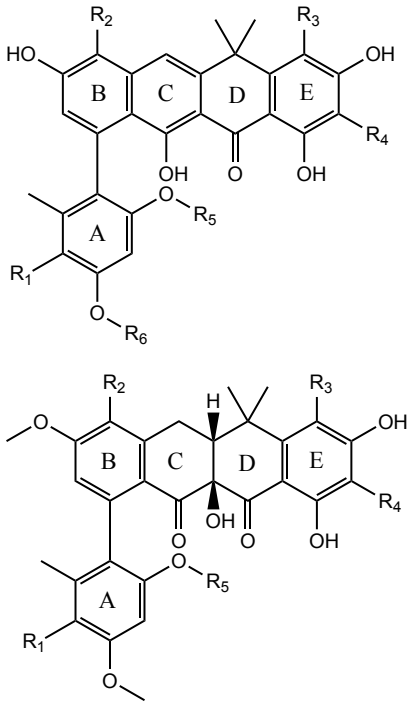


Figure 1.10: Bioassay guided fractionation using silica gel chromatography of culture extracts of *S. formicae* for the separation of the antifungal and antibacterial fractions. UV absorbance of fractions measured at 280 nm. Experiment performed by Dr Zhiwei Qin at the John Innes Centre. Figure reproduced with permission.



Fasamycin A R_1=H, R_2=H, R_3=Cl, R_4=H, R_5=H, R_6=H
Fasamycin B R_1=H, R_2=H, R_3=Cl, R_4=Cl, R_5=H, R_6=H
1.Fasamycin C R_1=H, R_2=H, R_3=H, R_4=H, R_5=H, R_6=CH_3
2.Fasamycin D R_1=H, R_2=Cl, R_3=H, R_4=H, R_5=H, R_6=CH_3
3.Fasamycin E R_1=Cl, R_2=Cl, R_3=H, R_4=H, R_5=H, R_6=CH_3

4.Formicamycin A R_1=H, R_2=Cl, R_3=H, R_4=H, R_5=H,
5.Formicamycin B R_1=Cl, R_2=Cl, R_3=H, R_4=H, R_5=H,
6.Formicamycin C R_1=H, R_2=Cl, R_3=Cl, R_4=H, R_5=CH_3,
7.Formicamycin D R_1=Cl, R_2=Cl, R_3=Cl, R_4=H, R_5=H
8.Formicamycin E R_1=Cl, R_2=Cl, R_3=Cl, R_4=H, R_5=CH_3
9.Formicamycin F R_1=Cl, R_2=Cl, R_3=H, R_4=Cl, R_5=CH_3
10.Formicamycin G R_1=H, R_2=Cl, R_3=Cl, R_4=Cl, R_5=CH_3
11.Formicamycin H R_1=Cl, R_2=H, R_3=Cl, R_4=Cl, R_5=CH_3
12.Formicamycin I R_1=Cl, R_2=Cl, R_3=Cl, R_4=Cl, R_5=H
13.Formicamycin J R_1=Cl, R_2=Cl, R_3=Cl, R_4=Cl, R_5=CH_3
14.Formicamycin K R_1=H, R_2=Cl, R_3=Br, R_4=Cl, R_5=CH_3
15.Formicamycin L R_1=Cl, R_2=Cl, R_3=Br, R_4=Cl, R_5=CH_3
16.Formicamycin M R_1=H, R_2=Br, R_3=H, R_4=H, R_5=CH_3

Figure 1.11: Structures of previously reported fasamycins A and B (Feng, Kallifidas and Brady, 2011) along with the new fasamycin congeners C-E and the new formicamycins A-M from *S. formicae* (Qin *et al.*, 2017). For the purpose of labelling absorbance data in this thesis, the compounds isolated from *S. formicae* are numbered **1-16** as shown. Compounds **1-16** were isolated at the John Innes Centre and their structures solved by NMR conducted by Dr Zhiwei Qin.

With the purified compounds in hand, Dr John T Munnoch (former PhD student in the Hutchings Laboratory) investigated the structure activity relationship (SAR) of the formicamycins by examining the growth of *Bacillus subtilis* in liquid medium supplemented with micromolar concentrations of the formicamycins. An example growth curve for formicamycin I (**12**) is shown (**Figure 1.12**). These experiments revealed that the formicamycins were at least as potent as other known antibiotics like apramycin and ampicillin.

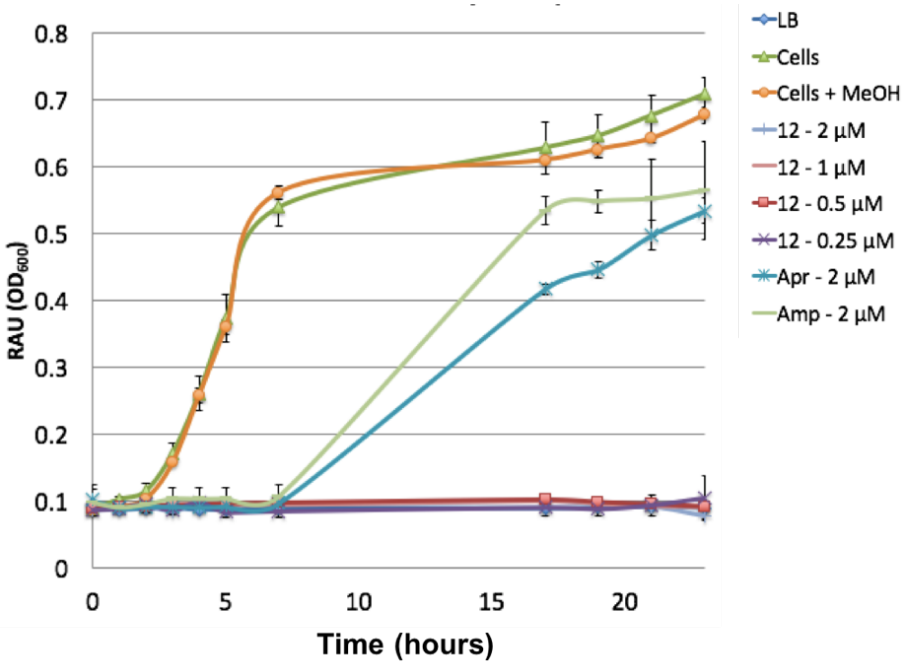


Figure 1.12: Growth curve of *Bacillus subtilis* in the presence of various concentrations of formicamycin I (**12**) as well as apramycin (Apr), ampicillin (Amp) or methanol (MeOH) to control for the solvent used. Assays conducted by Dr John T Munnoch and figure reproduced with permission (Qin *et al.*, 2017).

The minimum inhibitory concentrations (MICs) for each compound against *B. subtilis,* MRSA and VRE were identified (**Table 1.2**).

Table 1.2: MIC data for the fasamycins and formicamycins from *S. formicae* against *B. subtilis*, and MRSA and VRE clinical isolates. Values indicating "Not tested" or with "<" or ">" indicate issues with compound availability and/or a decision not to test further concentrations, i.e. they represent the lowest/highest concentrations tested. Assays conducted by Dr John T Munnoch and reproduced with permission (Qin *et al.*, 2017).

Compound	Minimum Inhibitory Concentration (μM)		
	B. subtilis	MRSA	VRE
Fasamycin C	<20	40	40
Fasamycin D	10	10	10
Fasamycin E	5	80	80
Formicamycin A	5	>80	>80
Formicamycin B	10	10	10
Formicamycin C	5	1.25	80
Formicamycin D	10	20	10
Formicamycin E	10	20	10
Formicamycin F	5	20	2.5
Formicamycin G	5	Not tested	Not tested
Formicamycin H	10	Not tested	Not tested
Formicamycin I	<2.5	<2.5	1.25
Formicamycin J	<20	0.625	1.25
Formicamycin K	<2.5	2.5	5
Formicamycin L	<2.5	1.25	2.5

The formicamycins are all potent inhibitors of Gram-positive bacteria, including drug-resistant clinical isolates of MRSA and VRE. The potency of the compounds appears to increase as the number of chlorine atoms on the compound increases. In addition, formicamycins appear to be generally more bioactive than fasamycins. This could be due to the saturation of ring C and/or the ability to add more halogen atoms to the scaffold, as fasamycins are only halogenated at two positions on the backbone whereas formicamycins can be halogenated at up to four different positions. Addition of bromine to the scaffold appears to make the compounds marginally more potent than the equivalent chlorinated formicamycin.

To investigate whether bacteria can acquire spontaneous resistance to formicamycins *in vitro*, Dr John T Munnoch grew MRSA for 20 generations in the presence of half-MIC concentrations of formicamycins A, J and L along with a no compound control. No spontaneous resistant mutants were isolated on agar and there appeared to be no change in MIC for these compounds, suggesting they exhibit a high natural barrier for the development of resistance, at least under the conditions tested here (Qin *et al.*, 2017). Whilst this is a beneficial property for a compound designed to be used in a clinical setting, it does raise challenges in determining the mechanism of action of the formicamycins. Previous studies have suggested that fasamycins A and B act by inhibiting type 2 fatty acid biosynthesis and it is possible that the formicamycins act in the same way. Overall, these preliminary results show that the formicamycins are attractive compounds for further investigation.

1.9 Aims and objectives of this thesis

The aim of this thesis was to investigate antibiotic production in *Streptomyces formicae* by applying new genomic tools to a novel environmental isolate. During the work presented, the biosynthesis of the formicamycins was fully characterised by identifying the genes responsible for production within the host and assigning functions to the enzymes present in the pathway. Regulation of formicamycin production and mechanisms of host resistance were also investigated. Using this knowledge, we have engineered strains of *S. formicae* that overproduce the formicamycins as well as their biosynthetic precursors, intermediates and side-products. This has led to the discovery of several novel chemical structures with a range of antibiotic activities. The mechanism of bioactivity of the formicamycins has also been investigated. Furthermore, experiments have been conducted to identify the BGC responsible for the production of the antifungal compounds from *S. formicae*. This work has revealed a prolific potential for the discovery of novel antimicrobials from this talented environmental isolate. The development of protocols for the genetic manipulation of *S. formicae* during this project will allow further work to be conducted on this organism in the future to identify further novel natural products.

2 Materials and Methods

2.1 Chemicals and Reagents

Chemicals and reagents used are laboratory standard grade or above, purchased from Sigma Aldrich (UK) or Thermo Fisher Scientific (UK) unless otherwise stated. All media and solutions were made using deionised water (dH_2O) except where stated otherwise.

2.2 Bacterial strains

The bacterial strains used or generated in this study are listed in the appendix of this thesis (**Chapter 8.1**). Growth media are listed in **Table 2.1** with selection concentrations of antibiotics listed in **Table 2.2**. *E. coli* strains were routinely grown shaking at 220 rpm, in LB broth or on LB agar at 37°C unless stated otherwise. Agar plates of *Streptomyces formicae* were grown at 30°C.

Table 2.1: Growth media used in this thesis and their constituents

Media	Recipe (per litre)	Water	pH
SFM	20 g soy flour 20 g mannitol 20 g agar	Tap	
MYM	4 g maltose 4 g yeast extract 10 g malt extract +/- 18 g agar	50:50 Tap:deionised	7.3
ISP2/YEME	4 g yeast extract 10 g malt extract 4 g glucose 20 g agar	Tap	

LB	10 g tryptone 5 g yeast extract 10 g NaCl (omitted when selecting with Hygromycin) +/- 20 g agar	Deionised	7.5
2YT	16 g tryptone 10 g yeast extract 5 g NaCl	Deionised	7.0
Oat	25 g oats (crushed in a pestle and mortar) 20 g agar	Tap	
TSB	17 g tryptone 3 g soya peptone 5 g NaCl 2.5 g dipotassium phosphate 2.5 g glucose	Deionised	7.3
PGA	4 g potato extract 20 g glucose 17 g agar	Deionised	7.5

Table 2.2: Antibiotics used in this thesis and their selection concentrations

Antibiotic	Selection concentration (μg/ml)
Ampicillin/Carbenicillin	100
Apramycin	50
Chloramphenicol	30
Erythromycin	10
Hygromycin	50
Kanamycin	50
Nalidixic Acid	25
Thiostrepton	30

2.3 Preparation of *Streptomyces* spores

A single colony of *Streptomyces* was picked and plated to give a confluent lawn on SFM or MYM agar and grown at 30°C for 7-10 days until spores were visible. A sterile cotton bud was then used to scrub the spores from the surface of the plate with 1.5 ml 20% glycerol. Harvested glycerol and spore solutions were preserved at -80°C.

2.4 Glycerol Stocks

Glycerol stocks of *E. coli*, *B. subtilis, S. aureus, Enterococcus* and *C. albicans* were produced by resuspending 1-3 ml of overnight culture in LB and glycerol (final concentration 20%). Stocks were stored at -80°C.

2.5 Microscopy

A standard light microscope was used to check for sporulation in *Streptomyces* cultures. Scanning electron microscopy was conducted by Elaine Barclay the bioimaging facility at the John Innes Centre. Briefly, the sample was cryo-fixed by plunging into sub-cooled nitrogen and transferred into the cryo-preparation chamber under vacuum. Charge was reduced by coating the samples in a fine layer

of platinum and the sample was transferred onto the cold-stage within the SEM chamber for imaging.

2.6 DNA Extraction

Phage-derived artificial chromosome (PAC) DNA (**Chapter 8.2**) and genomic DNA was extracted from *E. coli* and *Streptomyces* species respectively by pelleting 1 ml of overnight culture in a benchtop microcentrifuge at 13000 rpm for 5 minutes and resuspending in 100 μl solution I (50 mM Tris/HCl, pH 8; 10 mM EDTA). Alkaline lysis was performed by adding 200 μl solution II (200 mM NaOH; 1% SDS) and mixing by inverting ten times. 150 μl solution III (3M potassium acetate, pH 5.5) was added and samples mixed by inverting five times before being centrifuged at 13 000 rpm for 5 minutes. The supernatant was extracted in 400 μl phenol:chloroform:isoamyl alcohol by vortexing for 2 minutes and centrifuging at 13 000 rpm for 5 minutes. The upper phase was transferred to a fresh microcentrifuge and 600 μl 2-propanol added. The samples were then incubated on ice for 10 minutes to aid precipitation of the DNA. Samples were centrifuged at 13 000 rpm for 5 minutes and the DNA pellet washed with 200 μl 70% ethanol. After centrifuging again at 13 000 rpm for 5 minutes, the DNA pellet was air-dried for 5 minutes before being resuspended in autoclaved dH_2O.

2.7 DNA Quantification

Isolated DNA was analysed using the Nanodrop 2000 UV-Vis Spectrophotometer and the Qubit assay using the Qubit® fluorimeter 2.0. Both the high-sensitivity and the broad-range kits were used depending on the sample concentration estimated by Nanodrop.

2.8 Primers

All primers were designed manually in ApE (A Plasmid Editor) and ordered from Integrated DNA Technologies (IDT). A table of all primers used in this work is available in the appendix of this thesis (**Chapter 8.3**).

2.9 Polymerase Chain Reaction (PCR)

Two different DNA polymerases were used depending on the application. PCRBIO® Taq DNA Polymerase (from PCR Biosystems) was used for diagnostic PCR reactions, whereas Q5® High-Fidelity DNA polymerase was used to amplify DNA fragments for subsequent cloning. PCRs were generally conducted as below using either a DNA engine PTC 300 (BIORAD®) or Prime (Techne) PCR machine (**Table 2.3**).

Table 2.3: Conditions used for PCRs conducted during this thesis

Cycles	Temperature	Time	Notes
1	95°C	2 min	Initial denaturation
30	95°C	30 sec	Denaturation
	55°C to 72°C	30 sec	Anneal (T_m was calculated using NEB calculator at http://tmcalculator.neb.com)
	72°C	30 sec per kb	Extension
1	72°C	10 min	Final extension
1	4°C	hold	Final hold

2.9.1 PCRBIO® Taq DNA Polymerase

For diagnostic PCR, the following reaction mix was generated (**Table 2.4**).

Table 2.4: Reaction constituents for PCRBIO Taq reactions conducted during this thesis

Reagent	20 µl reaction	Final concentration	Notes
2x PCRBIO Taq Mix	10 µl	1x	
DMSO	1 µl	5%	To help primer annealing
Forward primer (5µM)	0.5 µl	125nM	
Reverse primer (5µM)	0.5 µl	125nM	
Template DNA	0.5 µl	variable	<10 ng plasmid or gDNA
dH_2O	7.5 µl		

2.9.2 Q5® High-Fidelity DNA polymerase

To amplify DNA fragments for cloning, the following reaction mix was generated (**Table 2.5**).

Table 2.5: Reaction constituents for Q5 reactions conducted during this thesis

Reagent	50 µl reaction	Final concentration	Notes
5X Q5 Reaction Buffer	10 µl	1x	
5X Q5 High GC Enhancer	10 µl	0.5%	Reduces T_m
10 mM dNTPs	1 µl	200 µM	
Forward primer (10 µM)	0.5 µl	100 nM	
Reverse primer (10 µM)	0.5 µl	100 nM	
Template DNA	0.5 µl	variable	<10ng plasmid or gDNA
Q5 High-Fidelity Polymerase	0.5 µl		
dH_2O	27 µl		

2.10 Agarose gel electrophoresis

Gels were made with 1% agarose in TBE buffer (90 mM Tris HCl, 90 mM Boric Acid, 2 mM EDTA) with 2 µg/ml ethidium bromide. DNA samples and loading buffer (5x) (0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene-cyanol blue, 40% (w/v) sucrose in water) were run alongside a 1 kb plus DNA ladder plus loading dye for size determination. Electrophoresis occurred at 120 V (Sub-Cell GT electrophoresis system, BIOLINE) for 30-60 minutes depending on size (larger fragments were run longer for clearer separation). DNA was visualised by UV-light using a Molecular Imager Gel Doc System (BIO-RAD).

2.11 Gene Synthesis

Gene synthesis was conducted by GenScript. When appropriate, *Streptomyces* genes were codon-optimised for expression in *E. coli*.

2.12 Plasmid preparation

The plasmids used or generated in this study are listed in the appendix of this thesis (**Chapter 8.2**). Plasmid DNA was prepared using QIAprep Spin Miniprep kits (QIAGEN) from 3-5 ml overnight cultures as per manufacturer's instructions. Plasmids where eluted from the column using 30 µl autoclaved dH_2O.

2.13 Restriction digest

Both Roche and NEB restriction enzymes were used to digest plasmid DNA and PCR fragments in 50-100 µl total volumes in accordance with manufacturer's guidelines. Digests were carried out with optimal buffer, which was outlined by Roche or NEB. Digestion of 1 µg of DNA was typically performed at 37°C for 1 hour by adding 1 unit of the appropriate restriction enzyme. Then the restriction enzymes were heat inactivated at 65°C for 10 minutes and 2 µl shrimp alkaline phosphatase was added to dephosphorylate the digested plasmid DNA to prevent re-ligation. Digests were then analysed by gel electrophoresis; the desired bands were excised and gel extracted for downstream applications.

2.14 Gel extraction

Gel fragments containing DNA bands of interest were excised using a scalpel and extracted using a QiaQuick Gel Extraction Kit (QIAGEN), according to the manufacturer's instructions. DNA was eluted in 30 µl autoclaved dH_2O.

2.15 Ligation

Ligation reactions were carried out using T4 DNA Ligase according to the manufacturers' instructions. A standard ratio for reactions was 1:3 of plasmid to

insert. Volumes were calculated using the (NEB) NeBio Ligation Calculator (at https://nebiocalculator.neb.com) to account for differences in DNA size and concentration.

2.16 Golden Gate

Constructs were assembled in 20 µl reactions with 100 ng purified backbone and 0.3 µl insert in the presence of 2 µl T4 ligase buffer (NEB) and 1 µl T4 ligase (NEB) with 1 µl BbsI (NEB) and dH_2O. This was run in a thermocycler under the following conditions:

- 10 cycles of the following:
    - 10 minutes at 37°C
    - 10 minutes at 16°C
- 5 minutes at 50°C
- 20 minutes at 65°C
- 4°C hold

Assembly of small inserts such as 20 nucleotide synthetic protospacers were confirmed by blue/white screening using X-Gal and sequencing by Eurofins. Larger inserts were confirmed using colony PCR (**Chapter 2.21**) or restriction digestion (**Chapter 2.13**) followed by gel electrophoresis (**Chapter 2.10**).

2.17 Gibson Assembly

Multiple DNA fragments were assembled into digested vector backbones using designed overlaps of between 18 and 24 nucleotides. Gel extracted DNA fragments were incubated in a ratio of 1:3 of plasmid to insert (1:5 for inserts smaller than 300 nucleotides) in the presence of Gibson Assembly master mix (NEB) at 50°C for 1 hour.

2.18 Preparation and transformation of electrocompetent *E. coli*

Single colonies of *E. coli* ET12567/pUZ8002 and Top10 were grown as 10 ml overnight cultures in LB containing appropriate antibiotics (**Table 2.2**) and sub-cultured to an

OD_{600} of approximately 0.4. Cultures were pelleted by centrifugation at 4000 rpm for 5 minutes at 4°C and resuspended in 10 ml ice-cold 10% v/v glycerol. Pellets were washed in ice-cold 10% v/v glycerol twice before being either flash frozen in liquid nitrogen and stored at -80°C or being used for transformation immediately. For transformation, approximately 2 μg of DNA was added to 50 μl of cells and the mixture transferred to an ice-cold electroporation cuvette. Electroporation was carried out using the BioRad Electroporator set to: 200 Ω, 25 μF and 2.5 kV. The electroporated cells were diluted in LB and transferred to a micro-centrifuge tube for recovery at 37°C for 1 hour with 220 rpm shaking before plating on selective media for overnight incubation and colony selection.

2.19 Preparation and transformation of chemically competent *E. coli*

Single colonies of Top10 *E. coli* were grown as 10 ml overnight cultures in LB containing appropriate antibiotics, and sub-cultured to an OD_{600} of approximately 0.4. Cultures were pelleted by centrifugation at 4000 rpm for 5 minutes at 4°C and resuspended in 10 ml of ice-cold 100 mM $CaCl_2$. Pellets were washed in ice-cold 100 mM $CaCl_2$ twice before being either flash frozen in liquid nitrogen and stored at -80°C or being used for transformation immediately. For transformation, approximately 2 μg of DNA was added to 50 μl of cells, mixed gently and incubated on ice for around 30 minutes. This mixture was then heat-shocked for 30 seconds at 42°C then immediately cooled on ice for 2 minutes. The transformed cells were diluted in LB and recovered at 37°C for 1 hour with 220 rpm shaking before plating on selective media for overnight incubation and colony selection.

2.20 Tri-parental mating

The three strains required for tri-parental mating (usually *E. coli* DH10β containing the PAC clone, Top10 with pR9604 and ET12567/pUZ8002) were inoculated into liquid LB with appropriate antibiotic selection and grown overnight at 37°C, 250 rpm. The following day, these were sub-cultured and grown to exponential phase (measured by OD_{600} 0.4-0.6) and washed by centrifugation to remove antibiotics. 20 μl of each strain was spotted onto the centre of a LB agar plate with no antibiotic and

incubated overnight at 37°C. The spot was then re-streaked onto LB + antibiotics to select for the intended strain containing the desired combination of plasmids. The presence of the required plasmids was tested and confirmed by colony PCR.

2.21 Colony PCR in *E. coli*

After overnight incubation, single colonies were picked with a sterile pipette tip and transferred to the base of a 0.2 ml PCR Tube. The tip was then discarded into an Eppendorf containing 500 µl sterile LB media. Biotaq PCR was conducted as described above (**Chapter 2.9.1**) and reactions analysed by gel electrophoresis. Positive colonies were then transferred to 10 ml LB supplemented with antibiotics for overnight incubation and plasmid amplification.

2.22 Constructing gene deletions using Lambda λ RED methodology (ReDirect)

PAC clones were PCR targeted by ReDirect by replacing genes of interest with an apramycin resistance gene as described previously (Gust *et al.*, 2003). The published methodology was edited to remove the *oriT* from the apramycin cassette, as this sequence is also present in the PAC and therefore encouraged undesirable recombination events. A linear PCR product (FRT – flanked resistance gene construct) was purified and electroporated into *E. coli* BW25113 pIJ1790 containing the desired cosmid for PCR targeting. The expression of Lambda *red* genes was induced by addition of L-arabinose (final concentration 10 mM) to the growth medium to allow the introduction of linear DNA. Edited cosmids were PCR confirmed before conjugation into *Streptomyces* via ET12567/pUZ8002.

2.23 Constructing gene knockouts using CRISPR/Cas9 genome editing

Clean deletions were made using the pCRISPomyces-2 system as described previously (Cobb *et al.*, 2015) (**Figure 2.1**). Approximately 20 base pair protospacers for use in the synthetic guide RNA (sgRNA) were designed so that the last 15 nucleotides, including the NGG sequence, were unique in the genome to minimise off target effects. The forward and reverse sequences were ordered as oligos from

IDT and annealed by heating to 95°C for 5 minutes followed by ramping to 4°C at 0.1°C/second. Annealed protospacers were then assembled into the BbsI site of the pCRISPomyces-2 vector using golden gate assembly (**Chapter 2.16**). This vector was then digested with XbaI (**Chapter 2.13**) and a 2kb PCR-amplified homology repair template (1 kb from either side of the target region) was assembled into the vector using Gibson assembly (**Chapter 2.17**). The final vector was cloned in *E. coli*, isolated, confirmed by PCR and sequencing, and transformed into the desired *Streptomyces* strain by conjugation via the non-methylating *E. coli* strain ET12567/pUZ8002 (**Chapter 2.25**). Once the required deletion event had taken place, loss of the temperature-sensitive pCRISPomyces-2 plasmid was encouraged by plating mutants on media lacking antibiotic selection at 37°C for multiple generations.

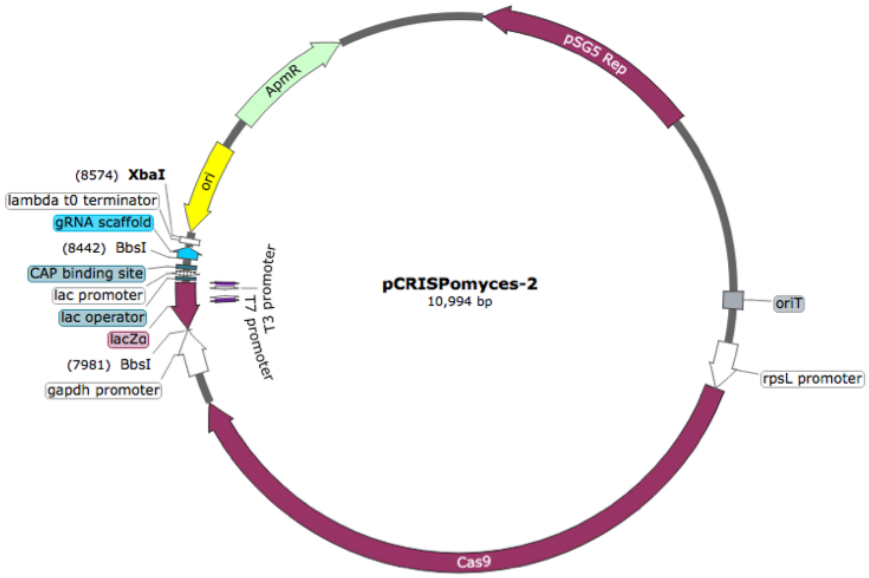


Figure 2.1: The pCRISPomyces-2 vector (Cobb, Wang and Zhao, 2015) visualised and annotated in SnapGene. The synthetic gRNA is seen, along with the BbsI site for integration of the desired protospacer. The Cas9 enzyme causes the double stranded break in DNA at the target site specified by the protospacer. The chromosome is then repaired according to the homologous repair template assembled into the XbaI site of the plasmid.

To confirm the desired deletion event had occurred within the genomic DNA, two PCR products of just over 1kb each were generated using primer pairs 1F/1R and 2F/2R from either genomic DNA prepped by the phenol:chloroform method (**Chapter 2.6**) or lysed mutant colonies (**Chapter 2.26**) (**Figure 2.2**). This approach allows PCR confirmation to be conducted on mutants that still contain the pCRISPomyes-2

plasmid (i.e. before multiple generations to cure the plasmid have been undertaken) as the external primers bind within the genomic DNA and not within the 2 x 1 Kb homologous repair template in the deletion vector. PCR products in the wild-type (WT) strain will be longer than the mutants as the gene(s) of interest are still present. For large deletions such as whole clusters, the WT strain does not generate a band at all, because the region of DNA would be too long for amplification by standard PCR. Using this experimental design, primers 2F and 1R can also be used together to generate a smaller product, however, this is only suitable on apramycin-sensitive strains that no longer contain the pCRISPomyces-2 plasmid, as this region would also be present here.

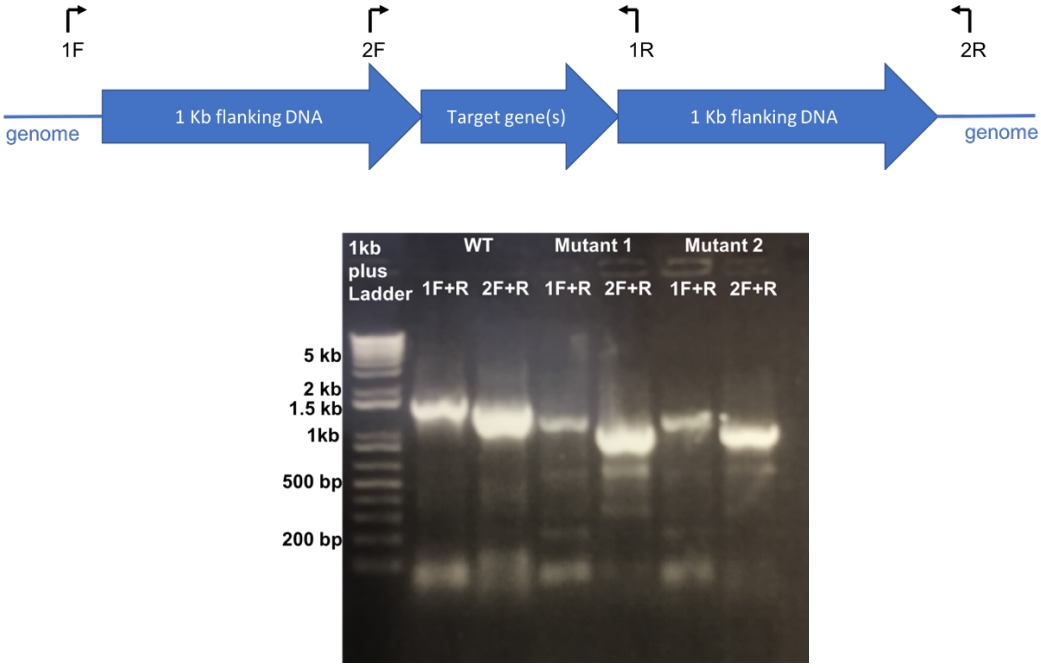


Figure 2.2: For all CRISPR/Cas9 deletions generated in this thesis, two products of just over 1kb each were generated in order to confirm the desired deletion event had occurred within the genomic DNA. The example shown here illustrates two mutants that have had an approximately 300 bp region of the genome deleted. The bands from the WT template are approximately 300 bp longer than the bands from the mutant strains. Some non-specific binding is present as the 'template' here is lysed *Streptomyces* colonies. The bands were excised and confirmed by sequencing.

2.24 Sequencing of small DNA fragments and plasmids

Cloned DNA constructs were confirmed by Sanger sequencing using the Mix2Seq service from Eurofins Genomics. Plasmid DNA was diluted according to the manufactures instruction and test primers were added directly with sterile dH_2O and a final concentration of 5% DMSO.

2.25 Conjugation

Single colonies of non-methylating *E. coli* ET12567/pUZ8002 containing the required plasmid were selected from plates and grown in 10 ml LB broth plus antibiotics at 37°C overnight at 250 rpm. Subcultures of OD_{600} between 0.4 and 0.6 were washed twice in LB to remove antibiotics. Between 10 and 200 μl of *Streptomyces* spores were heat shocked at 50°C for 10 minutes in 500 μl 2YT to encourage germination and added to the washed *E. coli* cells. The cell mixture was pelleted by centrifugation at 13 000 rpm for 1 minute, the supernatant was removed and cells were resuspended in the residual liquid. This was plated on SFM + 10 mM $MgCl_2$ at various dilutions and incubated at 30°C for 16-20 hours. For selection of desired ex-conjugants, 0.5 mg Nalidixic acid and an appropriate concentration of the selection antibiotic (to give final concentrations as described in **Table 2.2** was added in 1 ml dH_2O to each plate and cultures returned to the 30°C incubator for 5 days or until colonies appeared.

2.26 Colony PCR in *Streptomyces* species

Following purification by re-streaking on antibiotic selective media, *Streptomyces* ex-conjugants were confirmed by colony PCR. Single colonies were picked and soaked in 100μl 50% DMSO at 50°C for 1 hour. This was then used as template for BioTaq PCR (**Chapter 2.1**) at 10% of the final volume of the reaction (usually 2.5 μl in 25 μl).

2.27 Genetic complementation

Complementation of individual gene deletions were achieved by fusing the gene to either a native *S. formicae* promoter in pMS82 by Gibson assembly or by ligating the digested gene product downstream of the *ermE** promoter in pIJ10257. The complementation construct was then conjugated into the relevant mutant as described (**Chapter 2.25**).

2.28 Chemical Extraction of Secondary Metabolites from agar plates

Agar plates were inoculated with *S. formicae* and incubated at 30°C for 9 days. For small scale extractions, three circular regions of the agar were sampled using the lid of a falcon tube and a razor to excise the agar from the plate. The weight of the samples was measured to three decimal places. 1mL of ethyl acetate was added to the samples and shaken for 1 hour. For larger scale extractions for compound isolation, at least 40 plates of each strain were inoculated and grown as above. The agar was then crushed and extracted in a minimal volume of ethyl acetate. The ethyl acetate was evaporated in a Rotovap or Genevac depending on the sample volume and the sample resuspended in a minimal volume of HPLC grade methanol. For both extraction methods, the samples were then centrifuged to remove insoluble material, dried using the Genevac and resuspended in methanol for analysis.

2.29 High-performance Liquid Chromatography (HPLC) and Mass Spectrometry (LCMS)

For HPLC analysis, extracts in 100% methanol were run on the Agilent 1100 HPLC system using a Gemini-NX C18 00F0-4453-EO column (150 X 4.6 mm, 3 μm; Phenomenex). Solvent A was water + 0.1% formic acid and solvent B was 100% methanol. Compounds were separated through the column as follow: initially, 50% B (methanol) was run through the column to equilibrate, then, an elution of 50-100% B (methanol) was run through the column over 14 minutes. The concentration of B (methanol) was then reduced to 50% to wash the column.

For HPLC LCMS analysis, samples were run on a Shimadzu Nexera XR ultra high-performance liquid chromatographer (UHPLC) attached to a Shimadzu ion-trap time-of-flight (IT-ToF) mass spectrometer. For UHPLC, the Phenomenex Kinetex C_{18} column was used (100 x 2.1 mm, 100 Å) with solvent A as water + 0.1% formic acid and solvent B as 100% methanol. Compounds were separated through the column as follows; initially, 20% B (methanol) was run through the column to equilibrate, then, an elution of 20%-100% B (methanol) was run through the column over 12 minutes, followed by 2 minutes of 100% B to fully elute all the compounds off the column. The flow-through was then reduced back down to 20% B, which was run through the column for a total of 3 further minutes to fully clean the column. The flow rate during these experiments was set to 0.6 mL/minute, the injection volume was 10 μL and positive-negative mode switching was used. The spray chamber was calibrated using sodium trifluoroacetate cluster ions according to the manufacturer's instructions. The heat block was set to 300 °C, with the nebulizer gas rate at 1.5 L/min and the drying gas on.

All HPLC and LCMS experiments described in this thesis were run by Dr Zhiwei Qin and/or Hannah McDonald from Professor Barrie Wilkinson's group at the John Innes Centre. All strains were generated by Rebecca Devine and strain fermentation and extractions were conducted by both groups.

2.30 Protein overexpression and purification

The gene of interest was codon optimised and synthesised by GenScript and sub-cloned into pGS-21a for expression as a His-GST fusion protein. Protein was over-produced in 2L *E. coli* BL21 grown at 37°C for 3 hours and induced overnight with 0.1 mM IPTG at 20 °C, 200 rpm shaking. The following day, cells were harvested by centrifugation at 6000 rpm for 30 minutes. The pellets were frozen at -80°C for purification at a later date or processed immediately.

Pellets were thawed on ice and resuspended in PBS + lysozyme (100 μg/ml) + EDTA-free protease inhibitor. Cells were passed through a French press operated at x16000 lb/in^2 twice to lyse and separated into soluble and insoluble fractions by

ultracentrifugation at 42 000 rpm for 1 hour 30 minutes. A His-trap (1ml) column was loaded onto an AKTA FPLC (UPC-900, Frac-950, INV-907, M-925) and washed in 20% ethanol before the soluble fraction was loaded. The column was then washed with 40 mls of PBS. After that, a 40 mM solution of imidazole was passed through the column for 15 minutes to reduce non-specific binding before a fast gradient up to 500 mM imidazole was used to elute the bound protein. Fractions were collected every minute and analyzed by SDS page. Fractions that contained relatively pure protein were concentrated in Amicon Ultra 50 KDa centrifugal filters (Millipore) and exchanged into PBS + 10% (v/v) glycerol for storage. Protein was flash frozen in liquid nitrogen in aliquots for further use.

2.31 Protein analysis by SDS-PAGE

A standard resolving gel of 16.6 % (w/v) Acrylamide:Bis-Acrylamide 37.5:1 (Fisher BioReagents) was made as per the table below and cast using Mini-PROTEAN® Tetra hand cast systems (BIO-RAD) using 0.75mM glass plates. The resolving gel was left to polymerise at room temperature for a minimum of 30 minutes. Subsequently, a short stacking gel of 8% (w/v) Acrylamide was cast on top of the resolving gel and a 0.75mm comb used to set the lanes (**Tables 2.6** and **2.7**).

Samples to be analysed were defrosted on ice or analysed immediately post collection and resuspended in 50-200 µl SDS loading buffer (950 µl Bio-Rad® laemmli buffer, 50 µl β-mercaptoethanol) before boiling (100°C, 10 minutes). Cell debris was pelleted by centrifugation (5 minutes, 13000 rpm) and 20-50 µl of sample loaded onto SDS-PAGE gels with 3 µl PageRule Prestained Protein Ladder® (Thermo Scientific) as a marker.

Gels were electrophoresed (150V, 60 minutes) in 1 x TGS running buffer (0.025 M Tris-HCl, 0.192 M glycine, 1% SDS v/v) and stained with InstantBlue Protein Stain (Expedeon) with agitation at room temperature for 1 hour or overnight. Gels were de-stained in dH_2O for at least 1 hour and imaged using white light on a Molecular Imager Gel Doc System (BIORAD).

Table 2.6: Constituents of buffers used for making SDS-PAGE gels in this thesis

4x Resolving Buffer	4x Stacking Buffer
1.5 M Tris-HCl, pH 8.8	0.5 M Tris-HCl, pH 6.8
0.4% SDS	0.4% SDS

Table 2.7: Constituents of SDS-PAGE gels used in this thesis

Constituent	Stock solution	Volume	Final concentration
Resolving gel			
Resolving buffer	4x	2.5 ml	1x
Acrylamide	40% (w/v)	4.15 ml	16.6%
dH_2O		3.35 ml	
Tetramethylethylenediamine (TEMED)	>99%	10 µl	0.1%
Ammonium persulfate (APS)	10% (w/v)	100 µl	1%
Stacking gel			
Stacking buffer	4x	2.5 ml	1x
Acrylamide	40% (w/v)	2 ml	8%
dH_2O		5.5 ml	
TEMED	>99%	10 µl	0.1%
APS	10% (w/v)	100 µl	1%

2.32 Quantification of protein from whole cell lysates

Protein samples of unknown concentration were compared to samples of Bovine serum albumin (BSA) at a known concentration using the Bradford assay (Bradford, 1976). All samples were diluted in Bradford Dye Reagent solution (BIO-RAD) and dH_2O as below in 1.6 mL cuvettes (semi-micro disposable polystyrene 10 mm path length; Fisherbrand) and mixed by inverting (**Tables 2.8** and **2.9**). Absorbance was measured at A_{595} and a standard curve produced by plotting the absorbances of BSA

standards against their known protein concentration (mg/ml). Using this, the protein concentrations of unknown samples was calculated based on their absorbances.

Table 2.8: Constituents of standards used for protein quantification in this thesis

BSA (µl)	Bradford Reagent (µL)	dH_2O
0	200	800
1	200	799
3	200	797
5	200	795
8	200	792
10	200	790
15	200	785

Table 2.9: Constituents of unknown samples generated for protein quantification in this thesis

Sample (µl)	Bradford Reagent (µL)	dH_2O
1	200	800
3	200	797
5	200	795

2.33 Western blot

Protein samples were subjected to SDS-PAGE as above (**Chapter 2.31**) before being transferred to a nitrocellulose Biodyne A membrane (Pall Corporation) in a Trans-Blot SD Semi-Dry Transfer Cell (BIORAD). Three layers of blotting paper, equal size to the gel, were soaked in 1 x transfer buffer (25 mM Tris, 192 mM Glycine, 0.1 % SDS + 20% methanol) and placed on the transfer cell anode plate. Nitrocellulose membrane was soaked in 100% methanol (1 minute), followed by washing in transfer buffer (5 minutes) and placed on top of the blotting paper. The SDS polyacrylamide gel of proteins to be transferred was placed on top of the membrane, followed by three more layers of soaked blotting paper before transfer took place (10 V, 1 hour). A

blocking solution of 5% (w/v) fat-free skimmed milk powder in 1 x TBST (50 mM Tris Cl pH 7.5, 150 mM NaCl, 1% Tween) was poured over the membrane and incubated at room temperature overnight with gentle agitation. Anti-Flag antibody conjugated to horseradish peroxidase (HRP) was diluted 1: 20,000 in 1 x TBST and used to check strains containing 3xFlag tagged proteins. The membrane was incubated in 20ml of the respective antibody suspension at room temperature for 1 hour and washed 3 times for 10 minutes in TBST. Membranes were developed for 1 minute in a 50:50 mix of solutions A and B (**Table 2.10**) and fluorescence was detected using the ECL setting for imaging using a SYNGENE G:Box.

Table 2.10: Constituents of developing solutions used in this thesis

Developing Solution A	Developing Solution B
10 ml 100 mM Tris pH 8.5	10 ml 100 mM Tris pH 8.5
100 μl luminal	6 μl 30% hydrogen peroxide
45 μl coumaric acid	

2.34 Analysis of compounds by UV-Vis

Samples were analysed by UV-Vis using the Nanodrop 2000 UV-Vis spectrophotometer to generate wavelength scans between 200 and 800 nm.

2.35 RNA Extraction

For all RNA work, RNase-free water was prepared by DEPC treating (0.1% v/v) at 37°C O/N and autoclaving twice before use. All other equipment required for the processing of RNA was double autoclaved before use.

Samples were harvested using a sterile spatula from plates with cellophanes and flash frozen in liquid nitrogen for storage at -80°C before extraction. Pellets were thawed and crushed in liquid nitrogen using a sterile pestle and mortar on dry ice. Samples were resuspended in 1 mL RLT Buffer (Qiagen) supplemented with β-mercaptoethanol (10 μl in every 1 ml buffer) and vortexed for 1 minute. Samples were then applied to a QIA-shredder column (Qiagen) and centrifuged for 2 minutes.

Flow through was collected (leaving the pellet behind) and mixed with 700 μl acidic phenol-chloroform for 1 minute. Samples were then incubated at room temperature for 3 minutes before centrifuging at 13000 rpm for 20 minutes. The upper phase was collected and mixed with 0.5 volumes of 96% ethanol. This was then applied to a RNeasy Mini spin column (Qiagen) and purified following the manufacturers protocol including on column DNase treatment. Following elution, the Turbo-DNase kit was then used according to the manufacturers protocol and a further Qiagen RNeasey mini clean-up was conducted. Samples were then aliquoted for quantification or other downstream processes and flash frozen in liquid nitrogen for storage at -80°C.

2.36 RNA Quantification

For quantification, RNA was analysed by Nanodrop and formaldehyde gel electrophoresis. To prepare the formaldehyde gel, 1.2 g agarose was mixed with 10 ml 10x FA buffer (200mM 3-[N-morpholino]propanesulfonic acid (MOPS), 50 mM sodium acetate, 10 mM EDTA, pH to 7.0 with NaOH) and RNase free water to 100 ml, heated and cooled to 65°C in a water bath. 1.8 ml 37% formaldehyde and 1ul 10 mg/ml ethidium bromide was added and the gel poured into a pre-treated, RNase-free gel support. 1 volume of RNA loading buffer (16 μl bromophenol blue, 80 μl 500 mM EDTA pH 8.0, 720 μl 375 formaldehyde, 2 ml 100% glycerol, 3.084 ml formamide, 4 ml 10 x FA gel buffer, RNase free water to 10 ml) was added to 4 volumes of RNA. Samples were incubated at 65°C for 3-5 minutes, chilled on ice and loaded onto the gel. The gel was run at 70V in 1 x FA gel running buffer (100 ml 10X FA gel buffer, 20 ml 37% formaldehyde, 880 ml RNase-free water) for approximately 1 hour.

2.37 RT-PCR

RT-PCR reactions were conducted using the Qiagen OneStep RT-PCR Kit. Reactions were set up in 50 µl volumes (**Table 2.11**). To each reaction, 1 µl RNA was added.

Table 2.11: Constituents of RT-PCR reactions used in this thesis

Reagent	50 µl reaction	Final concentration
RNase-free water	variable	
5x Qiagen OneStep RT-PCR Buffer	10.0 µl	1x
dNTP mix (containing 10 mM of each dNTP)	2.0 µl	400 µM of each dNTP
Forward primer	variable	0.6 µM
Reverse primer	variable	0.6 µM
Qiagen OneStep RT-PCR Enzyme Mix	2.0 µl	

Thermocyclers were programmed as follows including an initial heating step at 95°C for 15 minutes to activate HotStarTaq DNA Polymerase (**Table 2.12**).

Table 2.12: Reaction conditions for RT-PCR reactions conducted in this thesis

Cycles	Temperature	Time	Notes
1	50°C	30 min	Reverse transcription
2	95°C	15 min	Initial PCR activation step
30	95°C	30 sec	Denaturation
	55°C to 72°C	30 sec	Anneal (Tm was calculated using NEB calculator (http://tmcalculator.neb.com/#!/))
	72°C	30 sec per kb	Extension
1	72°C	10 min	Final extension
1	4°C	hold	Final hold

PCR products were analysed by gel electrophoresis with no-RT controls to confirm no contamination of gDNA (**Chapter 2.10**).

2.38 Cappable RNA Sequencing

RNA was sent to Vertis Biotechnologie AG for sequencing and analysed by capillary electrophoresis. The RNA was then enriched by capping the 5' triphosphorylated RNA with 3'-desthiobiotin-TEG-guanosin 5' triphosphate (DTBGTP) (NEB) using the vaccinia capping enzyme (VCE) (NEB) for reversible binding of biotinylated RNA species to streptavidin. Enriched RNAs were poly(A)-tailed using poly(A) polymerase. 5'Illumina TruSeq sequencing adapters (carry 50% of each tag; ATTACTCG and TCCGGAGA) were then ligated to the 5' mono-phosphate groups of processed transcripts. The samples were then treated with CapClip Acid Pyrophosphatase (Cellscript) to convert 5' triphosphate (5'PPP) structures into 5' monophosphate ends. To the newly formed 5'P groups, the 5' Illumina TruSeq sequencing adapters (carry 50% of each tag; CGCTCATT and GAGATTCC) were ligated. First-strand cDNA synthesis was performed using an oligo(dT)- adapter primer and the M-MLV reverse transcriptase. The resulting cDNAs were PCR-amplified to about 10-20 ng/μl using a high-fidelity DNA polymerase.

For Illumina sequencing, 100-300 bp long 5' fragments were isolated from the full-length cDNAs. The cDNA was fragmented and the 5'-cDNA fragments were bound to streptavidin magnetic beads. The bound cDNAs were blunted and the 3' Illumina sequencing adapter ligated to the 3' ends before PCR amplification. The cDNA libraries were pooled in approximately equimolar amounts and confirmed to be around 200-500 bp. The cDNA pool was sequenced on an Illumina NextSeq 500 system using a 75 bp read length.

FASTQ sequencing files were sorted, quality checked, trimmed to remove rRNA reads, mapped and annotated by Vertis Biotechnologie AG.

2.39 Chromatin Immuno-precipitation (ChIP) Sequencing

For ChIP-seq experiments, spores were inoculated onto cellophane disks on SFM plates and grown for 2-4 days at 30°C. The discs were removed and the mycelium submerged in 10 ml of a 1% (v/v) formaldehyde solution for 20 minutes at room

temperature to cross-link proteins to DNA. The discs were then incubated in 10 ml of 0.5 M glycine for 5 minutes before the mycelium was harvested, washed twice with 25 ml ice-cold PBS (pH 7.4) and flash frozen at -80°C until needed. At the same time, a small sample was taken for western blot analysis to confirm expression of the tagged protein of interest.

For all downstream steps, EDTA-free protease inhibitor tablets were added to working solutions (2 tablets per 10 ml buffer). To lyse the cells, pellets were resuspended in 2 ml lysis buffer (10 mM Tris-HCl pH 8.0, 50 mM NaCl, 10 mg/ml lysozyme) and incubated at 37°C for 30 minutes. To fragment the DNA, 1 ml IP buffer (100 mM Tris-HCl pH 8.0, 500 mM NaCl, 1% v/v Triton-X) was added and samples mixed by pipetting up and down. Samples were sonicated 20 times at 50 Hz, for 10 seconds per cycle, and incubated on ice for >2 minutes between pulses.

25 µl of crude lysate was mixed with 75 µl TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA) and extracted with 200 µl phenol:chloroform. Of this extract, 25 µl was harvested and 2µl RNaseA (1mg/ml) was added. This was incubated at 37°C for 30 minutes and run on a 1% agarose gel to confirm the DNA fragments were within the desired size range for sequencing.

Once fragmentation was confirmed, the remaining crude lysate was cleared by centrifugation at 4000 rpm for 15 minutes at 4°C. For binding, 500µl Anti-FLAG M2 magnetic beads were prepared by washing in 2.5 ml 0.5 IP buffer. 40µl of the prepared beads were incubated with the cleared lysate for each sample on a vertical rotor overnight at 4°C. The beads were washed with 0.5 IP x 4 for 10 minutes each at 4°C.

To elute the bound DNA, 100 µl elution buffer (50mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS) was added and samples incubated at 65°C overnight. Another 50 µl elution buffer was added and the samples incubated for a further 5 minutes at 65°C. The 150 µl total eluate was purified by adding 2 µl proteinase K and incubating at 55°C for 1.5 hours. DNA was extracted in 150 µl phenol-chloroform and purified on a QIAquick column (Qiagen). DNA was eluted in 50 µl EB buffer (10 mM Tris-HCl pH

8.5); 45 µl was frozen for sequencing and 5 µl retained for quantification by nanodrop and Qubit (**Chapter 2.7**). The DNA was then sent to Genewiz® for sequencing using the Illumina HiSeq platform.

2.40 Analysis of ChIP-Sequencing data

Raw sequencing data were received as FASTQ files from Genewiz® and processed by Dr Govind Chandra at the JIC. Raw reads were aligned to the reference genome and extracted coordinates were listed in .bed files. These .bed files were visualised in Integrated Genome Browser (IGB) to show peaks of enrichment present across the whole genome. A cut off of around 2000 reads was enough to remove background noise and visualise only the significant peaks that indicate binding sites for each protein.

2.41 Solid culture colony bioassay

To assess the antibiotic production by a strain in solid culture, an agar plate (usually MYM, SFM for growth of *C. albicans*, see **table 2.1**) was inoculated in the centre with 2 µl *Streptomyces* spores and incubated at 30°C for 5-10 days. To overlay the bioassays with *E. coli*, *MRSA*, *Bacillus subtilis* or *Candida albicans*, the indicator strains were grown overnight in 10 ml LB at 30°C, and sub-cultured for 3-4 hours the following day to ensure the cells were in exponential growth phase. The subculture was diluted 1 in 10 into soft LB (0.5% agar) at 50°C which was used to overlay the plate. The hydrophobic nature of the *Streptomyces* colony means the overlay will not go over the top of the colony, but rest around the outer edge.

For *Lomentospora prolificans*, the fungus was re-streaked from a PGA stock plate onto the plate with the growing *Streptomyces* colony. *E. coli*, *MRSA* and *B. subtilis* assays were incubated at 30°C overnight before being visually examined. *C. albicans* plates were incubated at room temperature for 2-3 days until visible growth/inhibition was evident. Bioassays inoculated with *L. prolificans* were incubated at 30°C for 1 week or until sufficient growth/inhibition was evident.

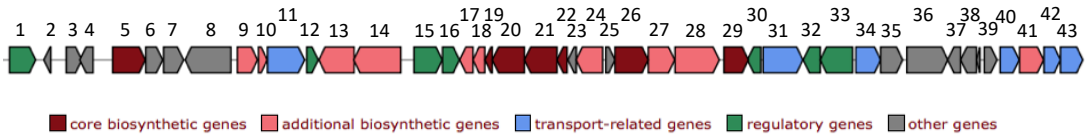
3 Identifying the formicamycin biosynthetic gene cluster and the mode of action of the formicamycins

Understanding the biosynthesis of a natural product depends on the characterisation of the biosynthetic gene cluster (BGC). The aim of this work is to identify the genes responsible for the biosynthesis of the formicamycins in *Streptomyces formicae*. Finding the genes responsible for the assembly of a particular microbial metabolite is extremely important; in order to obtain high yields of a compound of interest it is often necessary to generate mutations in the BGC in the native producer or even heterologously express the genes in a microorganism that is more genetically tractable (Ward and Allenby, 2018). Furthermore, increasing our understanding of secondary metabolite production in nature can inform future drug discovery and synthesis efforts by medicinal chemists. Identifying a BGC in the native host usually requires that the organism is genetically tractable. When working with environmental isolates of *Streptomyces*, this can be especially challenging as the biomass available for laboratory investigation is often limited due to difficulties in culturing the organism and many of the traditional genetic tools are not directly applicable. The availability of a high-quality genome sequence for *S. formicae* meant that bioinformatic analysis could be applied in combination with a newly developed CRISPR/Cas9 tool for genetic manipulation of this environmental isolate in order to identify the genes responsible for the biosynthesis of the formicamycins.

When working with novel natural products, it is also important to identify their mechanism of action, as this can provide an insight into the potential toxicity of the agent and therefore its suitability for development into a clinically useful drug. As described in **Chapter 1.8**, the formicamycins display clear antimicrobial activity against Gram-positive pathogens, however, no mechanism of action has previously been defined, so this is also investigated during this chapter.

3.1 Identifying the formicamycin biosynthetic gene cluster

Using the previously described AntiSMASH analysis of the *S. formicae* genome, the predicted BGCs were examined to identify potential genes that may be involved in the biosynthesis of the formicamycins. The halogenation of the compounds enabled any clusters without halogenase enzymes to be eliminated. This left three BCGs (8, 26 and 27), only one of which was predicted to be a T2PKS gene cluster (BGC 26) (**Table 1.1**). Based on the aromatic structure of the compounds, it was predicted that this cluster must be responsible for the production of the formicamycins. Putative gene products were assigned using the National Centre for Biotechnology Information (NCBI) Basic local alignment search tool (Blast) (Altschup *et al.*, 1990) (**Figure 3.1**). The predicted T2PKS cluster contains genes encoding for the ACC carboxylase enzymes necessary for making the extender units for polyketide biosynthesis, as well as the essential acyl-carrier protein and ketosynthase enzymes for building the polyketide backbone. In addition, there are 5 putative cyclase enzymes encoded in the BGC, which may correspond with the pentacyclic structure of the formicamycins, as well as several methyltransferases, including two O-methyltransferases, which could be responsible for the O-methylation at R_5 and R_6 (**Figure 1.5**). There is also a putative halogenase enzyme encoded within this cluster. The formicamycins can be chlorinated at up to four positions on the carbon chain, which would require at least one halogenase enzyme. Finally, there are also several regulatory genes and transporters that are common within BGCs encoding antimicrobial secondary metabolites.



#	Gene	AA	Annotation	#	Gene	AA	Annotation
1	*orf4*	306	NAD-dependent epimerase/dehydratase	23	*forS*	106	Monooxygenase/cyclase
2	*orf3*	336	MarR family transcriptional regulator	24	*forT*	342	O-Methyltransferase
3	*orf2*	199	Hypothetical protein	25	*forU*	119	Monooxygenase/cyclase
4	*orf1*	170	Transposase	26	*forV*	430	Halogenase
5	*forQ*	422	Decarboxylase	27	*forW*	341	O-Methyltransferase
6	*forP*	217	β-lactamase (metallohydrolase)	28	*forX*	571	Monooxygenase
7	*forO*	259	Exodeoxyribonuclease III	29	*forY*	315	Oxidoreductase
8	*forN*	590	Acylhydrolase	30	*forZ*	172	MarR family transcriptional regulator
9	*forM*	261	Methyltransferase	31	*forAA*	513	Multidrug resistance protein
10	*forL*	113	PKS cyclase	32	*forBB*	220	LuxR family response regulator
11	*forK*	478	Na+/H+ exchanger	33	*forCC*	417	Sensor histidine kinase
12	*forJ*	149	MarR family transcriptional regulator	34	*orf6*	321	ABC transporter
13	*forI*	455	ACC biotin carboxylase	35	*orf7*	284	ABC transporter permease
14	*forH*	607	ACC carboxyl transferase	36	*orf8*	529	Glutamate synthase
15	*forG*	363	Sensor histidine kinase	37	*orf9*	166	Hypothetical protein
16	*forF*	219	LuxR family response regulator	38	*orf10*	203	Hypothetical protein
17	*forE*	171	ACC biotin carboxy carrier protein	39	*orf11*	164	Hypothetical protein
18	*forD*	153	PKS cyclase/dehydratase	40	*orf12*	247	Glutamate ABC transporter
19	*forC*	96	PKS ACP	41	*orf13*	310	ABC transporter substrate binding protein
20	*forB*	415	KS_{β}	42	*orf14*	214	ABC transporter permease
21	*forA*	420	KS_{α}	43	*orf15*	289	ABC transporter permease
22	*forR*	131	Cupin (cyclase/monooxygenase)				

Figure 3.1: Predicted T2PKS BGC by AntiSMASH 4.0 and annotation of putative gene products assigned using NCBI Blast. AA = number of amino acids; ACC = acetyl-CoA carboxylase; PKS = polyketide synthase (Qin *et al.*, 2017).

Heterologous expression of the predicted formicamycin BGC in multiple S*treptomyces* superhosts including *S. coelicolor* M1152, M1146, M1154 and *S. venezuelae* M1714, M1702 and M1711 did not result in the production of any compounds (data not shown). Therefore, to confirm this predicted T2PKS gene cluster is responsible for formicamycin biosynthesis, experiments were designed to delete the entire predicted BGC in the producing organism. As with many environmental isolates, little was known about the ideal growth conditions for *S. formicae* and no attempts had previously been made to genetically manipulate the strain. A cosmid library was not available, so traditional methods of gene editing were not applicable, therefore CRISPR/Cas9 was chosen as the most appropriate method to optimise for use in *S. formicae*.

Generating the required plasmid for deletion of the predicted formicamycin BGC was relatively simple, as good quality genomic DNA could be isolated from *S. formicae* grown in liquid medium and used for cloning into the pCRISPomyces-2 backbone. However, in order to move the finished plasmid into *S. formicae*, it was necessary to obtain dense spore stocks for use in conjugation reactions with *E. coli*, therefore media screens were carried out to identify the optimal growth conditions for sporulation of *S. formicae* under laboratory conditions. Eight media recipes were initially tested and the resulting biomass was examined under a light microscope. Methods of stressing the bacteria into sporulation by nutrient starvation were unsuccessful as no growth was evident on water agar after 1 month's incubation at 30°C and there was limited growth on minimal media under the same conditions. Growth on SFM, SMASH agar and ISP4 was good after 10 days, however, limited signs of sporulation were visible under the light microscope (data not shown) (**Table 2.1**). Good growth and sporulation were evident on MYM, ISP2/YEME and oat agar after 10 days at 30°C, so these samples were further analysed using scanning electron microscopy (SEM) conducted by Elaine Barclay at the John Innes Centre to determine which medium would be optimal for generating spore stocks of *S. formicae* (**Figures 3.2-3.4**).

While all three samples did contain sporulating *S. formicae*, samples grown on MYM showed the most even and healthy spores. The spore chains cultured on ISP2/YEME

have a thin and ropey appearance. There is also a variety of different spore sizes present, which may be indicative of uneven chromosomal segregation. Similarly, the samples grown on oat agar appear thin and under-developed. Although growth seems to be denser on oat agar than on ISP2/YEME, some of the hyphae have still not differentiated into spores at day 10. Furthermore, there is evidence of uneven septum distribution when sporulation does occur. In contrast, the MYM samples show dense growth with smoother chains of spores that are all even in size, suggesting regular chromosomal segregation.

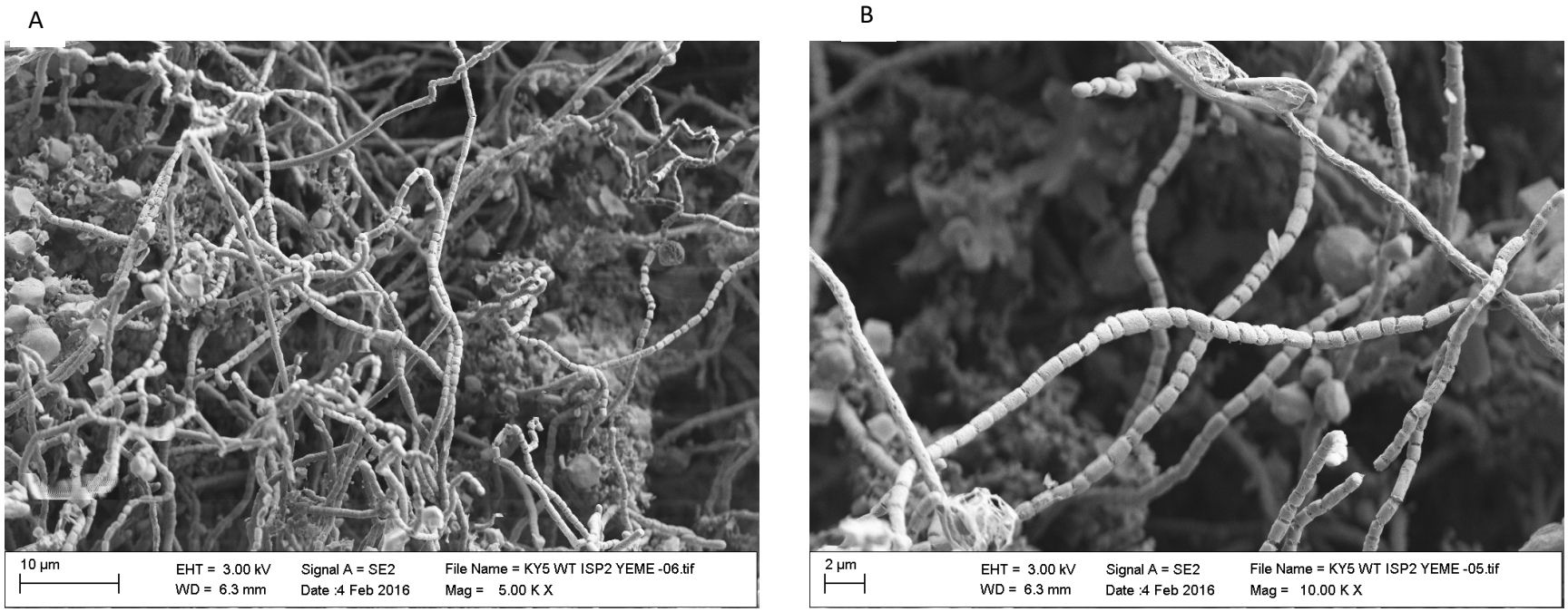
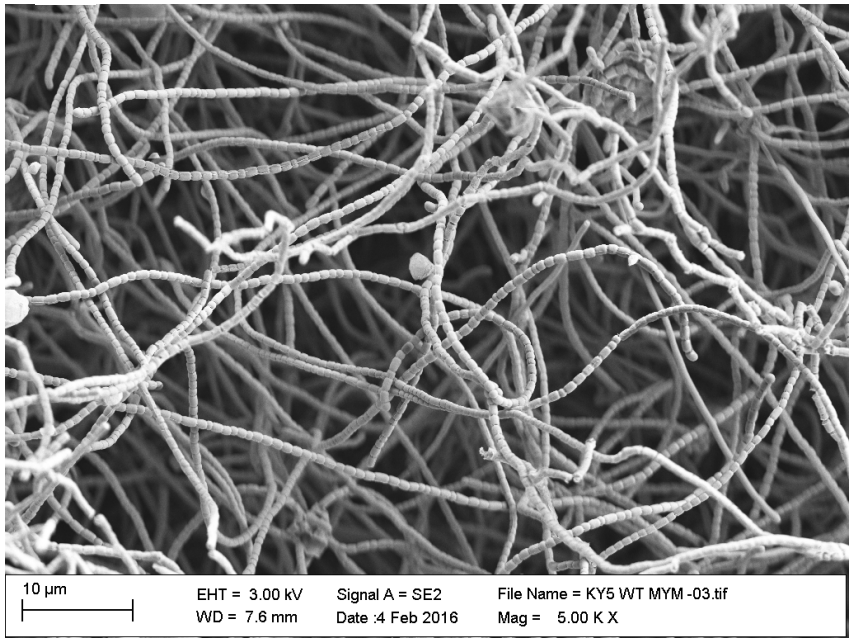


Figure 3.2: *Streptomyces formica* was grown on ISP2 YEME agar for 10 days at 30°C, flash frozen in liquid nitrogen and visualised by cryo-SEM at the John Innes Centre by Elaine Barclay. Samples were visually inspected for signs of sporulation at a variety of magnifications. Samples are pictured here at A: 5000x magnification and B: 10 000x magnification. Sporulation is evident although septum distribution appears uneven in places.

A



B

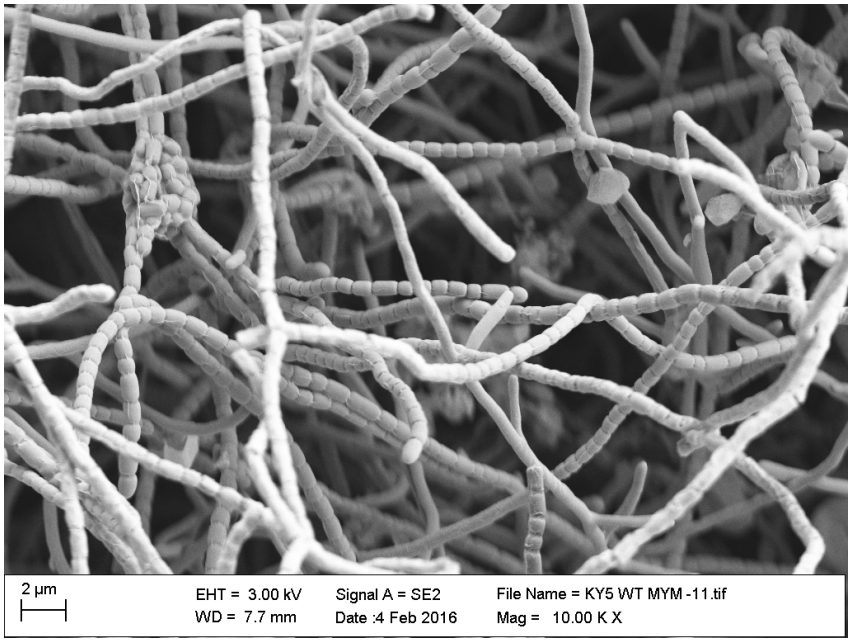
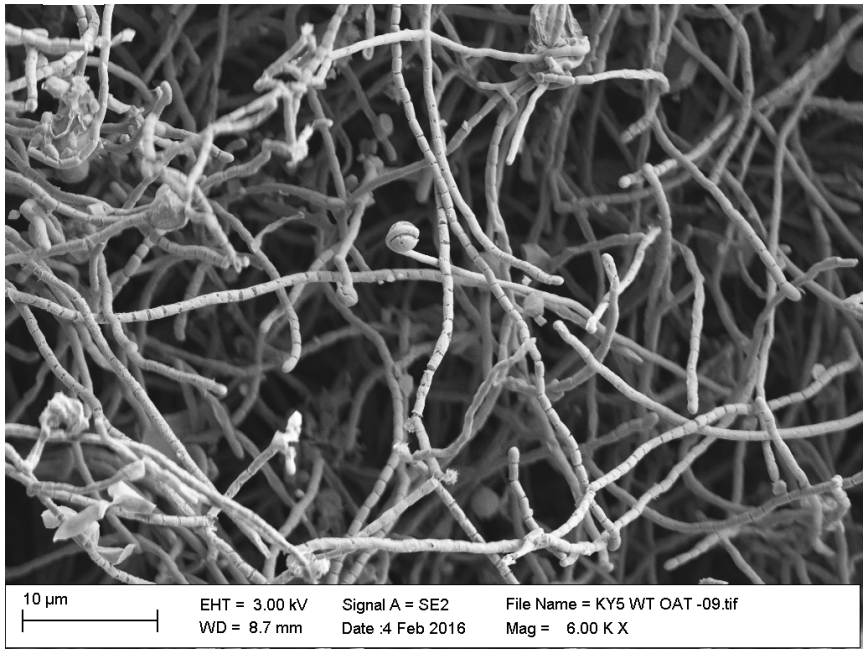


Figure 3.3: *Streptomyces formica* was grown on MYM agar for 10 days at 30°C, flash frozen in liquid nitrogen and visualised by cryo-SEM at the John Innes Centre by Elaine Barclay. Samples were visually inspected for signs of sporulation at a variety of magnifications. Samples are pictured here at A: 5000x magnification and B: 10 000x magnification. Healthy sporulation is evident and growth is dense.

A



B

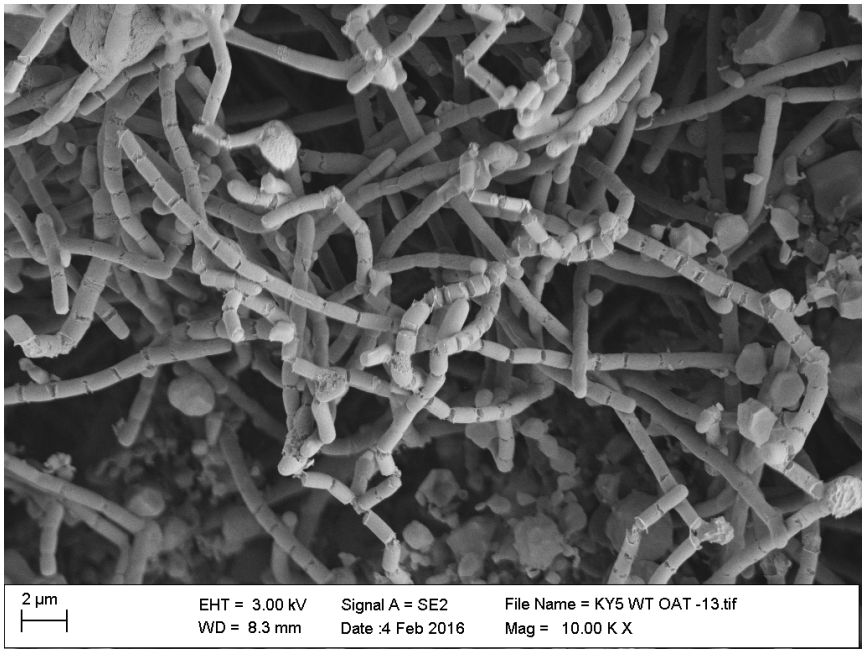


Figure 3.4: *Streptomyces formica* was grown on oat agar for 10 days at 30°C, flash frozen in liquid nitrogen and visualised by cryo-SEM at the John Innes Centre by Elaine Barclay. Samples were visually inspected for signs of sporulation at a variety of magnifications. Samples are pictured here at A: 6000x magnification and B: 10 000x magnification. Sporulation is evident although septum distribution appears uneven in places.

Spores were thus prepared from 10-day old cultures of *S. formicae* grown on MYM agar. A 20 nucleotide protospacer was designed to be specific to a region within the predicted formicamycin BGC. This oligonucleotide was assembled into the BbsI site of the pCRISPomyces-2 vector by golden gate assembly (**Chapter 2.16**) to confer target site specificity. Then, 1kb flanking arms were PCR amplified from either side of the putative formicamycin BGC and assembled together into the XbaI site using Gibson assembly (**Chapter 2.17**) to generate a vector that would encourage the complete, clean deletion of the 46 kb region predicted by AntiSMASH using the method described in **Figure 1.7.** Although a dense spore stock was used, only five colonies were obtained from conjugation reactions between wild-type *S. formicae* and *E. coli* ET12567 pUZ8002 containing the deletion vector. Nevertheless, three of the five were confirmed to contain the desired deletion and were cured of the plasmid resulting in three biological replicates of the unmarked, whole cluster deletion mutant (*S. formicae* Δ*for*). These mutants were confirmed by PCR and sequencing of the resulting PCR product as described in **Chapter 2.23**. To determine whether the deletion of the T2PKS cluster had any effect on formicamycin production, the three mutants were grown in parallel with the wild-type control under previously established formicamycin producing conditions (incubation at 30°C for 9 days on SFM agar). LCMS analysis of mutant extracts (conducted by Dr Zhiwei Qin, JIC) confirmed that no fasamycins or formicamycins were produced by the *S. formicae* Δ*for* strains (**Figure 3.5**) (Qin *et al.*, 2017).

To ensure that loss of fasamycin and formicamycin biosynthesis was due to deletion of the T2PKS gene cluster and not off target effects, genetic complementation of the mutant was required. BioS&T Co. (Montreal, Canada) generated a P1-derived artificial chromosome (PAC) library of the *S. formicae* genome made in pESAC13 and they screened the library using three primer pairs, one each for the beginning, middle and end of the predicted formicamycin BGC, to identify a single clone carrying the entire predicted BGC. This clone, named pESAC13-215-G, was used to re-introduce the whole BGC *in trans* to generate *S. formicae* Δ*for*: *ΦC31 for* 215-G. LCMS analysis of this complemented strain confirmed fasamycin and formicamycin biosynthesis had been restored in the mutant, although titres were lower in the complemented

strain than in the wild-type. This is probably because the cluster is expressed *in trans* and not in its native locus, resulting in changes in the levels of transcription. Nevertheless, these results indicate that the predicted BGC does indeed encode for the biosynthesis of the fasamycins and formicamycins in *S. formicae* (**Figure 3.5**) (Qin *et al.*, 2017).

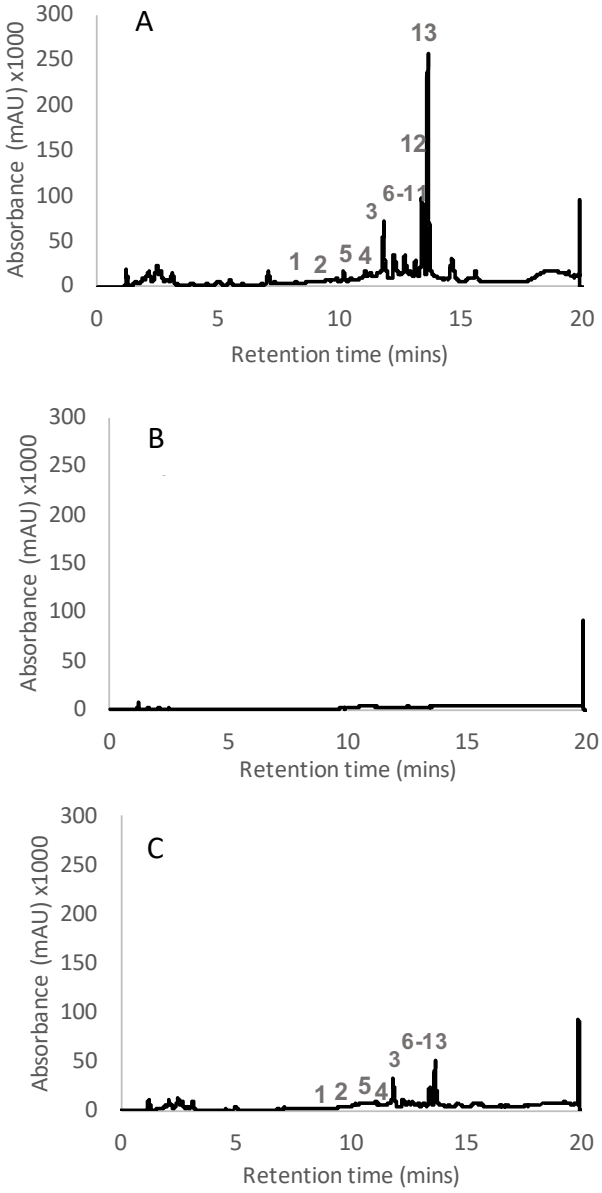


Figure 3.5: HPLC traces (UV 250 nm) showing extracts of: (A) *S. formicae* wild-type; (B) *S. formicae* Δ*for*; (C) *S. formicae* Δ*for*: *ΦC31 for* 215-G with any peaks from blank media removed. Deletion of the T2PKS BGC abolished fasamycin and formicamycin production. Re-introducing the cluster *in trans* restored production of all compounds, although at lower titres than in the wild-type (Qin *et al.*, 2017). Three biological replicates of each strain were generated for analysis. The presence of all compounds labelled was confirmed by LCMS analysis. HPLC(UV) and LCMS analysis was conducted by Dr Zhiwei Qin at the John Innes Centre.

3.2 Defining the edges of the formicamycin BGC

Having identified the BGC responsible for the biosynthesis of the fasamycins and formicamycins, we next wanted to identify the roles of the gene products encoded by this BGC. Often, BGCs predicted by AntiSMASH are larger than the minimal cluster of genes required for compound biosynthesis. In accordance with this, the predicted gene annotations in the formicamycin cluster suggest that several of the genes at the cluster edges may not be required for fasamycin or formicamycin biosynthesis (**Figure 3.1**).

As such, the ReDirect method was used to sequentially PCR target the edges of the gene cluster on the pESAC13-215-G DNA in *E. coli* (Gust *et al.*, 2004). The ReDirect method allows genes of interest to be deleted and replaced by a resistance cassette. In this case, the protocol was modified slightly so that the genes were replaced with just an apramycin resistance gene. The entire resistance cassette described in the original protocol was not used since there is already on *oriT* in the PAC and having one in the cassette as well would result in undesired recombination events. The altered PACs with truncated versions of the formicamycin BGC were then conjugated into *S. formicae* Δ*for*. Levels of formicamycin production were compared between all strains to determine the effect of removing either edge of the BGC on the progression of the biosynthetic pathway.

Initially genes 1-4 and 36-43 were deleted and replaced with the apramycin resistance gene. When these altered cosmids were introduced into *S. formicae* Δ*for*, all the fasamycins and formicamycins were produced at approximately wild-type levels, suggesting the deleted genes are not required for their biosynthesis (data not shown, HPLC(UV) LCMS conducted by Dr Zhiwei Qin, JIC). The experiment was then repeated to remove genes 1-7 and 34-43. Similarly, production of all fasamycins and formicamycins was evident in these strains (**Figure 3.6**).

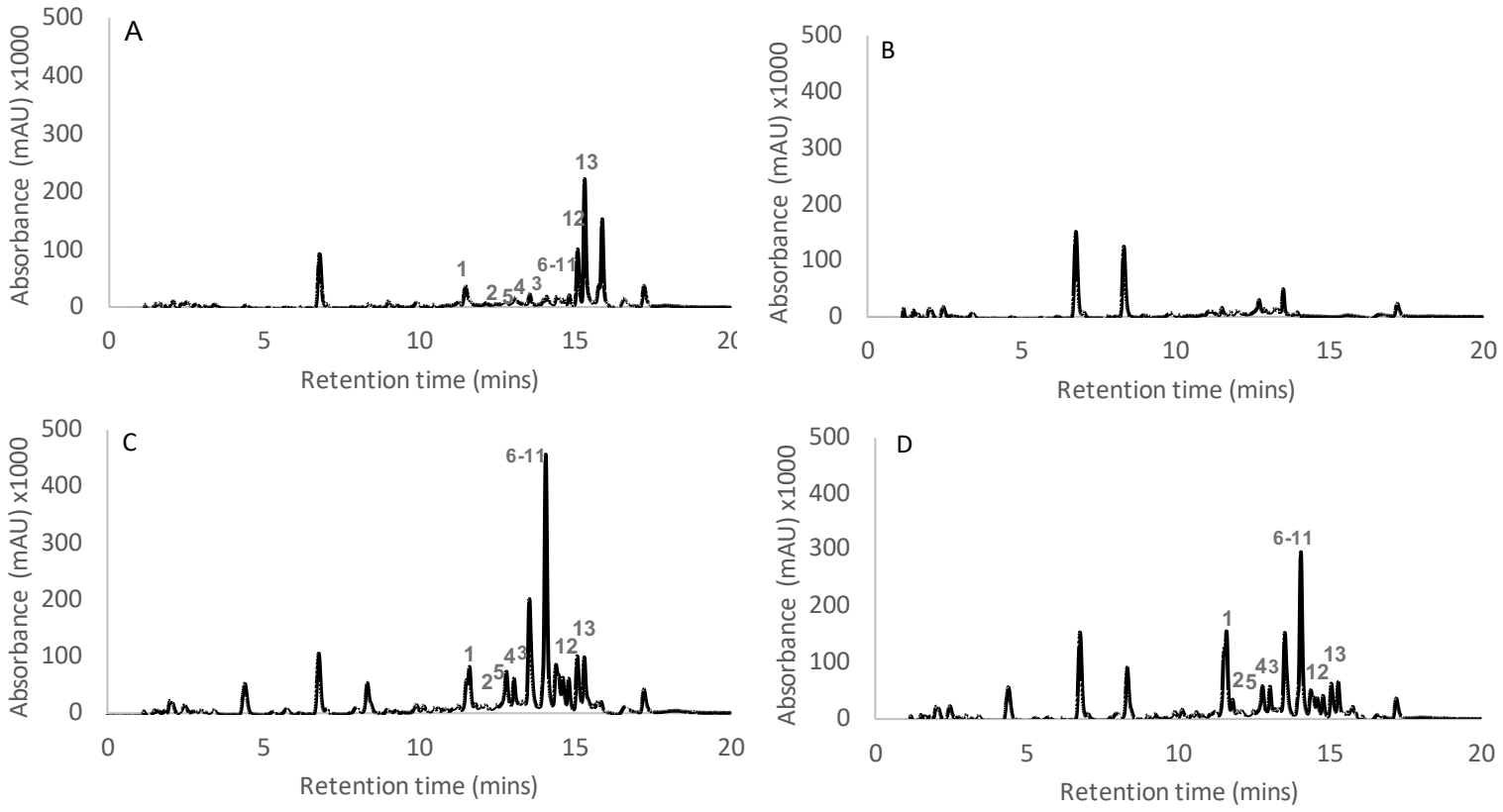


Figure 3.6: HPLC traces (UV 250 nm) showing extracts of: (A) *S. formicae* wild-type; (B) *S. formicae* Δ*for*; (C) *S. formicae* Δ*for*: φC31 *for* 1-7 *aac(3) IV*; (D) *S. formicae* Δ*for*: φC31 *for* 32-43 *aac(3) IV*. The edges of the formicamycin BGC were PCR targeted and replaced with an apramycin resistance gene to determine the minimal cluster required for biosynthesis. Expression of the edited, truncated BCGs in the *S. formicae* Δ*for* mutant resulted in production of all compounds, therefore genes 1-7 and 34-43 are not required for fasamycin and/or formicamycin biosynthesis. Three biological replicates of each strain were generated for analysis. HPLC(UV) LCMS analysis conducted by Dr Zhiwei Qin at the John Innes Centre.

This shows that genes 1-7 and 34-43 are not required for the biosynthesis of the formicamycins. Further experiments (described later) also indicate that ForBB and ForCC (genes 32 and 33) are not required for fasamycin or formicamycin biosynthesis (**Chapter 4.2**). Therefore, the minimal BGC was defined as the core 24 genes identified by AntiSMASH, from *forN* (gene 8) to *forAA* (gene 31) (**Figure 3.7**).

#	Gene	AA	Annotation	#	Gene	AA	Annotation
8	*forN*	590	Acylhydrolase	**20**	*forB*	415	KS_{β}
9	*forM*	261	Methyltransferase	**21**	*forA*	420	KS_{α}
10	*forL*	113	PKS cyclase	**22**	*forR*	131	Cupin (cyclase/monooxygenase)
11	*forK*	478	Na+/H+ exchanger	**23**	*forS*	106	Monooxygenase/cyclase
12	*forJ*	149	MarR family transcriptional regulator	**24**	*forT*	342	O-Methyltransferase
13	*forI*	455	ACC biotin carboxylase	**25**	*forU*	119	Monooxygenase/cyclase
14	*forH*	607	ACC carboxyl transferase	**26**	*forV*	430	Halogenase
15	*forG*	363	Sensor histidine kinase	**27**	*forW*	341	O-Methyltransferase
16	*forF*	219	LuxR family response regulator	**28**	*forX*	571	Monooxygenase
17	*forE*	171	ACC biotin carboxy carrier protein	**29**	*forY*	315	Oxidoreductase
18	*forD*	153	PKS cyclase/dehydratase	**30**	*forZ*	172	MarR family transcriptional regulator
19	*forC*	96	PKS ACP	**31**	*forAA*	513	Multidrug resistance protein

Figure 3.7: The defined, minimal formicamycin (*for*) BGC and predicted gene products.

3.3 Identifying the enzyme responsible for halogenation of the formicamycins

There are nearly 5000 reported halogenated natural products, many of which are antibiotics. In fact, around 20% of all reported active pharmaceutical agents are halogenated compounds. As stated previously, the formicamycins can be chlorinated at up to four positions on the carbon chain and supplementation of the *S. formicae* growth media with sodium bromide results in the production of additional brominated congeners (Qin *et al.*, 2017). Increased halogenation of the formicamycin backbone increases antimicrobial activity against Gram-positive bacteria. The beneficial effects of a carbon-halogen bond include increased stability against biodegradation and oxidation, as well as increased membrane permeability, both of which will contribute to increased bioactivity in most molecules (Cantillo and Kappe, 2017; Latham *et al.*, 2017). It is unsurprising therefore that incorporation of one or more halogen atoms into organic molecules is one of the most important and often required transformations in organic synthesis. Currently, synthetic halogenation events present a challenge industrially as they often require expensive reagents that are highly corrosive and result in the generation of toxic waste products (Pongkittiphan, Theodorakis and Chavasiri, 2009; Schnepel and Sewald, 2017). As such, there has been a lot of interest in biological halogenation in recent years.

Halogenase enzymes usually display high regioselectivity (van Pée *et al.*, 2016). It is surprising therefore that the formicamycin BGC only contains a single predicted halogenase, given that there are four different sites for potential halogenation. One possibility is that other halogenase enzymes encoded elsewhere in the genome are involved in halogenation of the formicamycins at R1-4. To investigate this, CRISPR/Cas9 was used to delete *forV*, which encodes the predicted halogenase within the *for* BGC, to generate the mutant strain *S. formicae* Δ*forV*. The extracts of three *forV* deletion mutants were analysed by HPLC(UV) LCMS by Dr Zhiwei Qin. Deletion of *forV* abolished the production of any halogenated fasamycins or formicamycins. Only fasamycin C, the non-halogenated fasamycin, along with a predicted structural isomer of the same compound with the same molecular formulae and UV spectrum, were present. The isomer has not been isolated and

therefore its structure remains unknown. The production level of fasamycin C (**1**) in *S. formicae* Δ*forV* was also much higher than is observed in the wild-type strain. Re-introduction of *forV in trans* under the predicted native promoter reinstated the production of all compounds at approximately wild-type levels (**Figure 3.8**). This suggests ForV acts non-specifically at all 4 different positions on the fasamycin/formicamycin backbone. Furthermore, the data indicate that ForV performs a gatekeeper role in controlling progression of the biosynthetic pathway as, when *forV* is deleted, only the non-chlorinated fasamycin accumulates, with no evidence of any non-halogenated formicamycins present in the extract of the mutant, implying halogenation must occur before formicamycin biosynthesis can be initiated (Qin *et al.*, 2017).

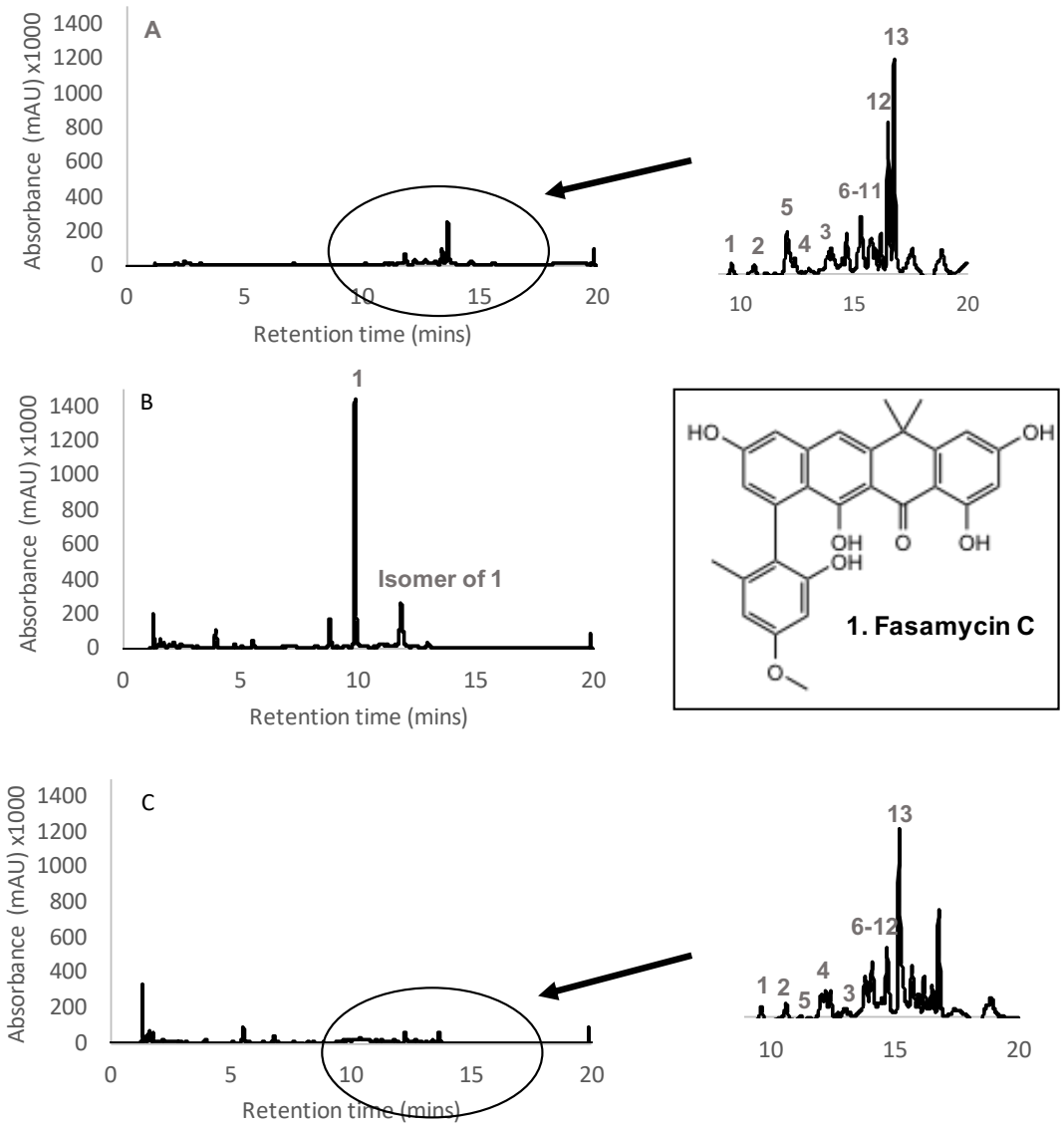


Figure 3.8: HPLC traces (UV 250 nm) showing extracts of: (A) *S. formicae* wild-type; (B) *S. formicae* Δ*forV*; (C) *S. formicae* Δ*forV*: φBT1 *for*V *pforU*. Deletion of *forV* results in the production of non-halogenated fasamycin C (**1**) only, the structure of which is shown in the insert. Re-introducing *forV in trans* restores production of all compounds (Qin *et al.*, 2017). Scales are increased to demonstrate the accumulation of fasamycin C (**1**) in *S. formicae* Δ*for*V. The presence or absence of all fasamycins and formicamycins was confirmed by LCMS and is indicated by UV traces with smaller scales in the inserts where necessary. Three biological replicates of each strain were generated for analysis. HPLC(UV) LCMS analysis conducted by Dr Zhiwei Qin at the John Innes Centre.

ForV is predicted by BlastP to be a FAD-dependent halogenase, a member of the flavin-dependent monooxygenase family, to which the extensively studied tryptophan halogenases also belong. Halogenation by these enzymes is usually carried out in the presence of O_2, Cl^- and $FADH_2$, which is formed by the reduction of the FAD cofactor by a reductase enzyme (Dorrestein *et al.*, 2005). Using this reduced

flavin, these enzymes can then go on to activate molecular oxygen to generate C4a-hydroperoxy flavin, which allows diverse downstream reactions such as hydroxylation and Baeyer-Villager oxidation to occur in many NRPS and PKS pathways. Although chlorination by these halogenases is more common, they are also able to brominate their substrate if appropriate concentrations of bromine are present in the surrounding environment (Latham *et al.*, 2017). Due to the fact that flavin-dependent halogenases display high regioselectivity and often halogenate at chemically unfavourable positions, much investigation into the structures and mechanisms of these enzymes has been conducted in recent years (Schnepel and Sewald, 2017).

All flavin-dependent halogenases contain a characteristic GxGxxG motif for flavin binding. In addition to this, all known FAD-dependent halogenase enzymes display high structural similarity. There are usually two distinct sites for substrate binding and cofactor binding, often separated by around 10 Å in distance. The flavin cofactor is usually located near the surface, to allow for rapid reduction of FAD even in the absence of the substrate, whereas the substrate binding site is buried deep within the protein (Schnepel and Sewald, 2017). To connect these sites, there is a tunnel of non-polar residues leading to the active site, often containing the WxWxIP motif (Latham *et al.*, 2017). At the end of this tunnel, there is a highly conserved lysine residue in the active site which interacts with the chlorinating agent either via hydrogen bonding or by forming a chloramine intermediate, to enable halogenation of the substrate (Andorfer and Lewis, 2018). Several studies have shown that mutating this residue in other flavin-dependent halogenases (K79 in RebH and PrnA, K74 in Rdc2 and RadH) completely abolishes halogenation activity in the enzyme (Dong et al. 2005; Yeh et al. 2006; Menon et al. 2017; Andorfer & Lewis 2018).

Modelling in Swiss-model reveals that ForV shows the greatest homology to CndH (34.99% identity), an FAD-dependent halogenase involved in the biosynthesis of the antimicrobial secondary metabolite chondrochloren by the myxobacterium, *Chondromyces crocatus Cm c5* (Jansen *et al.*, 2003; Buedenbender *et al.*, 2009; Rachid *et al.*, 2009; Waterhouse *et al.*, 2018). Similarly, analysis of ForV in I-TASSER

showed that the structural templates with the highest significant similarity to ForV were likely to be those of CndH and PltA, a flavin-dependent halogenase from the pyoluteorin biosynthetic gene cluster in *Pseudomonas fluorescens* Pf-5 (Roy, Kucukural and Zhang, 2010; Pang, Garneau-Tsodikova and Tsodikov, 2015). Pyoluteorin is an antifungal produced by a PKS-NRPS hybrid BGC halogenated by PltA at two sites on the scaffold (**Figure 3.9**). PltA was one of the first flavin dependent halogenases to be functionally characterised (Pang, Garneau-Tsodikova and Tsodikov, 2015). Alignment of ForV with CndH and PltA using T-Coffee multiple sequence alignment tools (Notredame, Higgins and Heringa, 2000) shows that all three protein sequences contain the characteristic lysine residue at position 76 (Lys-76), highlighted in green, and the WxWxIP fingerprints, highlighted in yellow, as well as the GxGxxG motif common to all flavin binding monooxygenases, highlighted in cyan (**Figure 3.10**). ForV and CndH display the highest sequence similarity to one another, although PltA also aligns well at the active site and the cofactor binding site. Interestingly, all three peptide sequences show unique C-terminal regions that do not align well to each other. This region may therefore serve an alternative function and be highly evolved and specific for each antibiotic biosynthesis pathway.

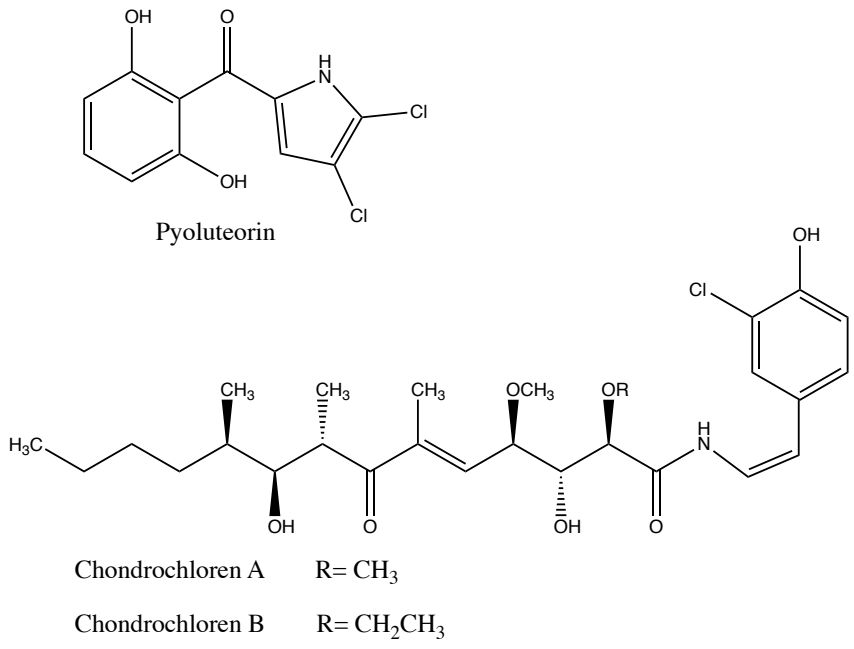


Figure 3.9: Structures of the halogenated natural products, chondrochloren and pyoluteorin. The halogenase enzymes responsible for the addition of the chlorine atoms on the rings of these molecules show similarity to ForV from the *for* BGC, therefore they may act via similar mechanisms.

```
CLUSTAL W (1.83) multiple sequence alignment

CndH            MSTR-PEVFDLIVIGGGPGGSTLASFVAMRGHRVLLLEREAFPRHQIGES
ForV            MSDQ-QRHYDVVVVGGGPAGASTAALLATEGRSVLVLEREKFPRYHIGES
PltA            GPHMSDHDYDVVIIGGGPAGSTMASYLAKAGVKCAVFEKELFEREHVGES
                 .    . :*::::****.*:: *: :*  *    ::*:* * * ::***
CndH            LLPATVHGICAMLGLTDEMKRAGFPIKRGGTFRWGKEPEPWTFGFTRH--
ForV            LIPGVW-PTLDRLGLRERLENMGLVRKYGGTLVWGRDLPQWTFSFA----
PltA            LVPATT-PVLLEIGVMEKIEKANFPKKFGAAWTSADSGPEDKMGFQGLDH
                *:*..       :*: :.::. .:  * *.:   . .    .:.*

CndH            --------------PDDPYGFAYQVERARFDDMLLRNSERKGVDVRERHE
ForV            --------------DGGPYPYAYQVRRADFDALLLTRARELGAHVVEDAT
PltA            DFRSAEILFNERKQEGVDRDFTFHVDRGKFDRILLEHAGSLGAKVFQGVE
                              .    ::::* *. ** :** .:   *..* :

CndH            VIDVLFEGERAVGVRYRNTEGVELMAHARFIVDASGNRTRVSQAVGERVY
ForV            VKEPLFDGERMTGVRYQPRGGDPVEAHADLVVDASGQQRWLGRHFDLIRW
PltA            IADVEFLSPGNVIVNAK-LGKRSVEIKAKMVVDASGRNVLLGRRLGLREK
                : :  * .   . *. :      :  :* ::*****..  :.: ..

CndH            SRFFQNVALYGYFENGKRLPAPRQGN----ILSAA--FQDGWFWYIPLSD
ForV            HEDLRNIAVWAYFQGCRRYEGQESGN----VLIEY--RPGGWLWFIPLGD
PltA            DPVFNQFAIHSWFDNFDRKSATQSPDKVDYIFIHFLPMTNTWVWQIPITE
                   :.:.*: .:*:.  *  . .. :    ::       . *.* **: :

CndH            TLTSVGAVVSREAAEAIKDGHEAALLRYIDRCPIIKEYLAPATRVTTGDY
ForV            GTTSIGYVTPTATLAESGLTAEKLWAEQTAESREVARMMEPATRVS----
PltA            TITSVGVVTQKQNYTNSDLTYEEFFWEAVKTRENLHDALKASEQVR----
                  **:* *.           *    .       :   : .: :*

CndH            GEIRIRKDYSYCNTSFWKNGMALVGDAACFVDPVFSSGVHLATYSALLVA
ForV            -GFRTIKDWSYNLERFHGPGWLAVGDAACFVDPLLSTGVTLALRGGATAA
PltA            -PFKKEADYSYGMKEVCGDSFVLIGDAARFVDPIFSSGVSVALNSARIAS
                  ::   *:**    .   .   :**** ****::*:** :*  ..  .:

CndH            RAINTCLAGEMSEQRCFEEFERRYRREYGNFYQFLVAFYDMNQDTDSYFW
ForV            SAAGMMLDTPATAHETGRRYEESYRRFLTGILDFVRAFYDQRKNRGEYYE
PltA            GDIIEAVKNNDFSKSSFTHYEGMIRNGIKNWYEFITLYYRLNILFTAFVQ
                      :     :    .:*   *.   .  :*:  :*  .     :

CndH            SARKIINTEER--ANEAFVRLIAGRSNLDEPVFQSVAKDFFTEREGFGAW
```

```
ForV            DAKKILDRGEGLPADVDFVTLVSGLSD---------GGDFFESP------
PltA            DPRYRLDILQLLQGDVY------------------------S--------
                ..:  ::  :   .:
CndH            FGGLVTSMAKGDGGGLMVGEGATDATESTGFAPENFMQGFTREITELQHL
ForV            -----------PEGGLELLEKLRSRVSAPPMTV-----------------
PltA            ------------GKRLEVLDKMREIIAAVESDPEHLWHKYLGDMQVPTAK
                             * : :   .   :
CndH            AMFGEDRGPETPLWSGGLVPSRDGLAWAVESGEDAAG
ForV            ----------------------------ERGR----
PltA            PAFEND------------------------------
```

Figure 3.10: Alignment of ForV with CndH and PltA, two known FAD-dependent halogenases. The key motifs present in all FAD-dependent halogenases are present; the GxGxxG motif at the site of FAD binding (cyan), the conserved lysine residue within the active site (Lys-76) (green) and the WxWxIP motif within the tunnel that leads to the active site (yellow). CndH and ForV have the highest sequence similarity. All three peptide sequences have highly variable C-terminal regions.

Previous studies have solved the crystal structures of CndH and PltA to 2.1 Å and 2.75 Å respectively (PDB 3e1t and 5dbj). Both models show the site of FAD binding as well as the site of the chlorine anion, although the C-terminal region is missing from the crystal structure of CndH as it is thought to have denatured during the crystallisation process (Buedenbender *et al.*, 2009). The predicted model of ForV generated in I-TASSER was therefore aligned in PyMol with the published models of both CndH and PltA. This showed that the structure of the FAD binding site and the substrate binding site were highly conserved between the proteins meaning that the ligand binding sites in ForV are likely to be the same. This implies that these enzymes likely function via similar mechanisms. By overlaying the models of CndH and ForV, the site of FAD binding and the hydrophobic tunnel leading to the active centre become clear, however, surprisingly, the active site seems to be open and exposed to the surface in both enzyme models (**Figure 3.11**). It should also be noted that the C-terminal regions of these proteins did not align, as the similarity between the regions is too low, so this region does not appear in the model below.

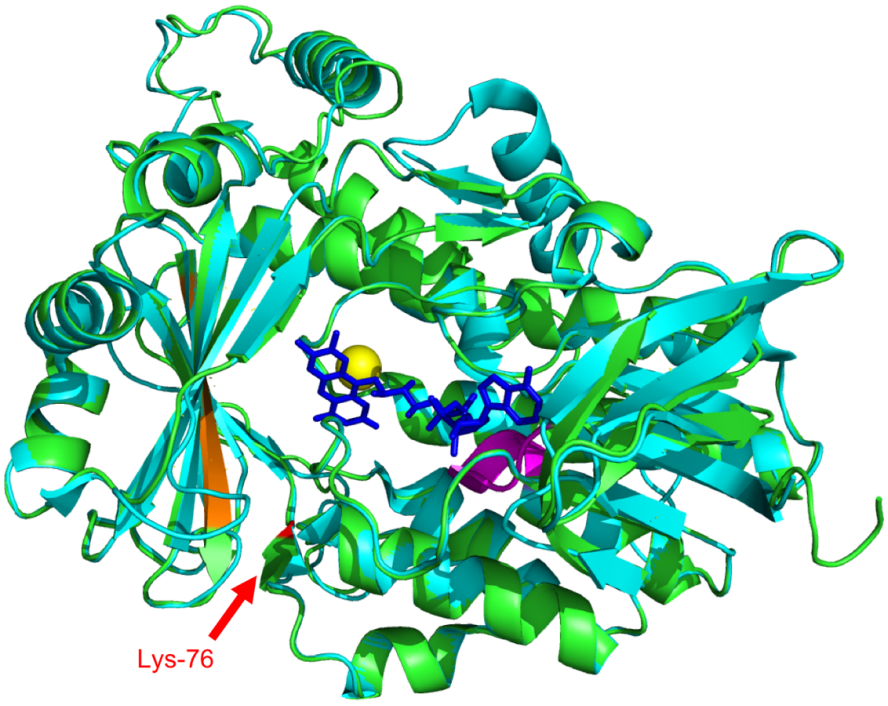


Figure 3.11: Alignment of the predicted model of ForV generated in I-TASSER (green) with the crystal structure of CndH (3e1t) (cyan) showing the binding sites of the FAD cofactor (dark blue) and the chlorine anion (yellow), generated in PyMol. The WxWxIP fingerprint is highlighted in orange, the FAD binding site GxGxxG motif is shown in magenta and the conserved lysine residue is shown in red (labelled). The structures align well in these key regions of the enzyme, suggesting that the mechanisms are likely to be similar. The C-terminal regions of these peptides do not align as the sequence similarity is too low and therefore this region does not appear in the model. In the absence of the C-terminal region, the active site is exposed to the surface of the protein.

Experiments were designed to generate *in vitro* activity data for ForV and produce crystals for accurate structural characterisation. To do this, *forV* was synthesised and codon optimised for expression in *E. coli* by GenScript®. This codon optimised gene sequence was sub-cloned into pGS-21a and expressed as a His-GST-ForV fusion protein in *E. coli* BL21 using 0.1 mM IPTG induction at 18°C for 16 hours. Even with the glutathione S-transferase (GST) tag, solubility of ForV was poor and as such, yields of pure protein were limited, however, MS-MS (conducted by Gerhard Saalbach at the proteomics facility at the JIC) confirmed that the small amount of protein obtained did align back to the sequence of ForV.

On incubation of purified ForV with substrate (either fasamycin C or formicamycin B), FAD reductase, FAD cofactor, NADH and a chlorine donor in phosphate buffered saline (PBS) at 30 °C for 1 hour, no halogenation activity was seen (data not shown).

Experiments were extensively optimised, including removing the GST tag from ForV in case this interfered with the enzyme activity, increasing the concentration of FAD in the reaction mixture, increasing the incubation time from 1 hour to overnight, and replacing the commercially available FAD reductase with a cloned and purified reductase enzyme that had previously been shown to activate a tryptophan halogenase (personal communication, Professor Changjiang Dong, 2016). In addition, a variety of different buffers for the reaction were tried including HEPES as well as changing the solvents for the substrates to remove any constituents that might inhibit the activity of ForV. However, no chlorination of substrates by ForV was witnessed in any of the experiments conducted in this project. Due to the low yields of ForV obtained and the absence of any observable *in vitro* activity, crystal screens were not conducted.

During the expression and purification of CndH, similar problems with poor solubility and low protein yield were encountered. Although the authors were successful in generating crystals after 3 weeks, no *in vitro* activity data was collected and this was predicted to be because either the reducing agent or the carrier protein was unknown (Buedenbender et al. 2009). During this investigation into CndH, the authors proposed that phenolic flavin-dependent halogenases could be broadly classified into two groups; type A halogenases that accept free substrates such as amino acids and type B enzymes that only accept substrates that are bound to a carrier protein (Buedenbender et al. 2009). PltA has previously been shown to specifically recognise the pyrrolyl substrate on the carrier protein scaffold, with no activity being present with the free substrate (Dorrestein *et al.*, 2005). Subsequently, when the crystal structure of PltA was solved in 2015, it showed the presence of a C-terminal region blocking the binding cleft that was not present on many previously studied flavin-dependent halogenases. The authors proposed that this region of the protein is responsible for binding to the carrier protein, inducing a conformational change allowing access to the substrate binding site for halogenation to occur (Pang, Garneau-Tsodikova and Tsodikov, 2015). They also noted that this C-terminal region is highly divergent between PltA and other halogenases, including CndH, in both sequence similarity and structural model (**Figure 3.12**). The absence of the C-

terminal domain might therefore explain why the active centre of CndH remains open in the crystal structure shown above. It may be that the highly variable C-terminal regions present on these flavin-dependent halogenases block the active site until they interact with the appropriate carrier protein, where a conformational change allows access for the required substrate. If this is the case, it is also likely that the C-terminal domain is unstable in the absence of the carrier-bound substrate and that is why previous studies have not enabled the crystallisation of this region, however a structure of the complex between the enzyme and the carrier-bound substrate would be necessary to confirm this hypothesis (Buedenbender et al. 2009).

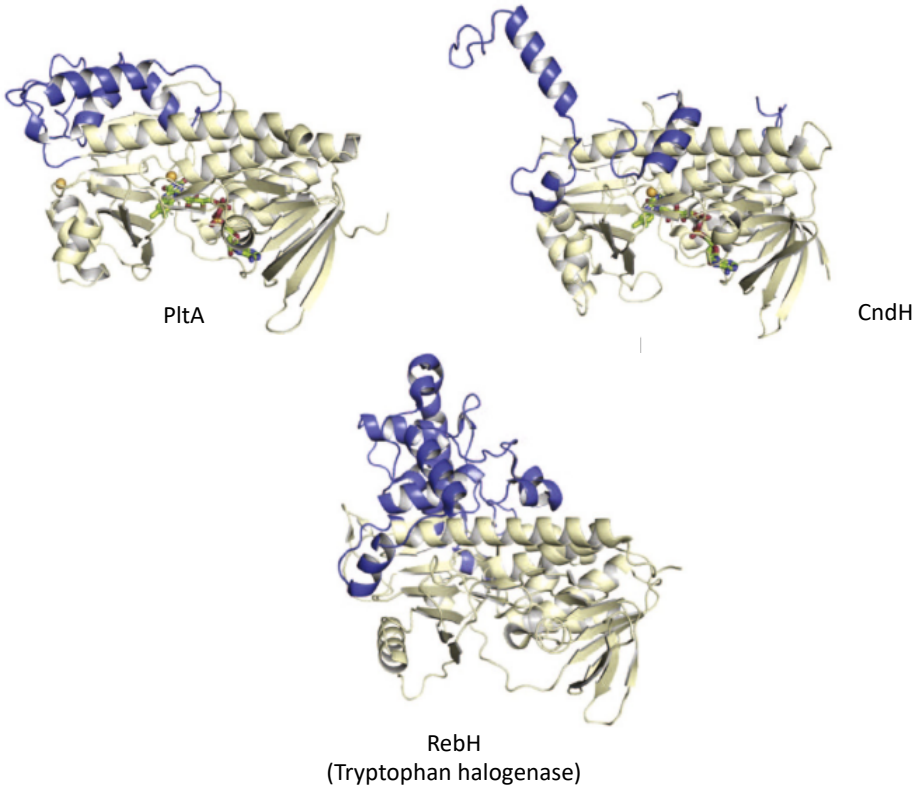


Figure 3.12: Models of the Type B halogenases PltA (crystal structure solved experimentally) and CndH (partial crystal structure and partial model) showing the highly variable C-terminal region in blue, compared with the Type A tryptophan halogenase RebH from the rebeccamycin biosynthetic gene cluster (crystal structure solved experimentally). It is predicted that the C-terminal covers the active site of Type B halogenases until the appropriate carrier is recognised, allowing the substrate into the active site. Figure adapted from (Pang, Garneau-Tsodikova and Tsodikov, 2015).

Using the model of ForV generated in I-TASSER, which predicts secondary structures based on the properties of the amino acids in the peptide chain, rather than aligning proteins to the most similar match as Swiss-model does, the C-terminal region of

ForV appears quite flexible and disordered, implying it may interact with other proteins (**Figure 3.13**). This model is highly speculative, as the C-terminal region does not align to either CndH or PltA, or any other proteins in the database, so no conclusions can be drawn from this model. However, it is possible that this region is specific to a carrier protein from within the formicamycin biosynthetic pathway and is responsible for recognition of the bound substrate for halogenation. Type B halogenases are probably more common in NRPS and PKS systems and these secondary metabolites are often synthesised by multi-enzyme complexes. Substrates are anchored to a carrier protein and passed from one module to another to ensure the pathway runs to completion with minimal side reactions occurring (Schnepel and Sewald, 2017). We therefore predict that ForV only recognises substrates that are bound to a carrier and this is why no activity was seen *in vitro* when purified fasamycins or formicamycins were used as the substrate.

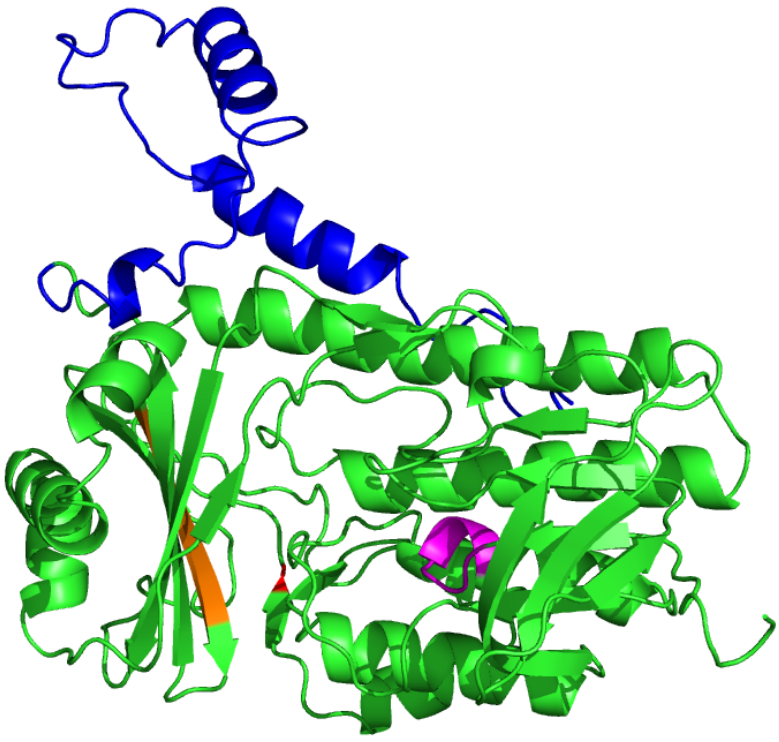


Figure 3.13: Predicted model of ForV generated in I-TASSER and visualised in PyMol. The WxWxIP fingerprint is highlighted in orange, the FAD binding site GxGxxG motif is shown in magenta and the conserved lysine residue is shown in red. The unique C-terminal region of ForV is shown in dark blue. The structure of this region is predicted at this moment but it appears highly flexible and may therefore be able to interact with other proteins from the formicamycin biosynthesis pathway.

Further *in vitro* work on ForV will be challenging until the carrier or enzyme complex can be identified or until a substrate-bound crystal structure of the enzyme can be solved. However, these type B halogenases are not thought to be effective catalysts for industrial halogenation at present due to the limitations discussed in this chapter (Schnepel and Sewald, 2017). For use in synthetic halogenation events, type A flavin dependent halogenases such as RadH, an enzyme involved in the biosynthesis of radicicol in *Chaetomium chiversi* which has recently been shown to halogenate a range of aromatic substrates *in vitro*, may be more appropriate due to their ability to accept free substrates (Menon *et al.*, 2017).

3.4 Predicted formicamycin biosynthetic pathway

Having identified the genes responsible for biosynthesis and halogenation of the fasamycins and formicamycins, we can tentatively put forward a predicted biosynthetic pathway for these compounds, based on previous knowledge of bacterial type 2 polyketide biosynthesis. Generally, this starts with a single acetyl-CoA starter unit loaded onto the acyl carrier protein (ACP). This is then transferred to the heterodimeric ketosynthase (KS) enzyme that is responsible for condensing multiple malonyl-CoA extender units to form a polyketide chain of a pre-determined length (Zhang, Pan and Tang, 2017). In the biosynthesis of the formicamycins, we predict that ForH, ForI and ForE generate the malonyl-CoA extender units by carboxylating acetyl-CoA and transferring them to the ACP. ForABC forms the minimal PKS, where ForA (KS_α) and ForB (KS_β) form the heterodimeric KS and ForC is the ACP. Feeding experiments with labelled acetate previously conducted by Dr Zhiwei Qin suggest that the condensation of multiple malonyl-CoA starter units by the PKS machinery forms a tridecaketide intermediate (**17**) which can then be cyclised, dehydrated, decarboxylated and methylated with the predicted enzymes as shown, to form fasamycin C (**Figure 3.14**) (Qin *et al.*, 2017). The functions of the latter enzymes are speculative at this point and are investigated experimentally later in this thesis.

Accumulation of fasamycin C in *S. formicae* Δ*forV* suggests that chlorination of the backbone by the ForV halogenase is the next step in the pathway and is required before formicamycin biosynthesis can be completed. This, together with the low levels of fasamycins in the extracts of wild-type *S. formicae* compared to the levels of formicamycins, suggests that fasamycins are biosynthetic precursors to the formicamycins (Qin *et al.*, 2017). The detailed mechanism underlying the conversion of a fasamycin into a formicamycin is investigated later in this thesis, although it is predicted at this stage to involve ForX and ForY, the monooxygenase and the oxidoreductase, respectively. ForN, the acylhydrolase, is likely to remove the cyclised intermediate from the ACP, at which point a decarboxylation reaction would also be required to remove the carboxyl group. This may be enzymatically catalysed by ForQ, the decarboxylase, in the wild-type strain, however, as seen in **Chapter 3.2**, deletion of ForQ does not affect formicamycin biosynthesis, therefore this reaction may also occur spontaneously. It is not yet clear at which point the intermediate is released from the ACP and whether that is the carrier that facilitates halogenation by ForV. This will also be discussed further later in this thesis.

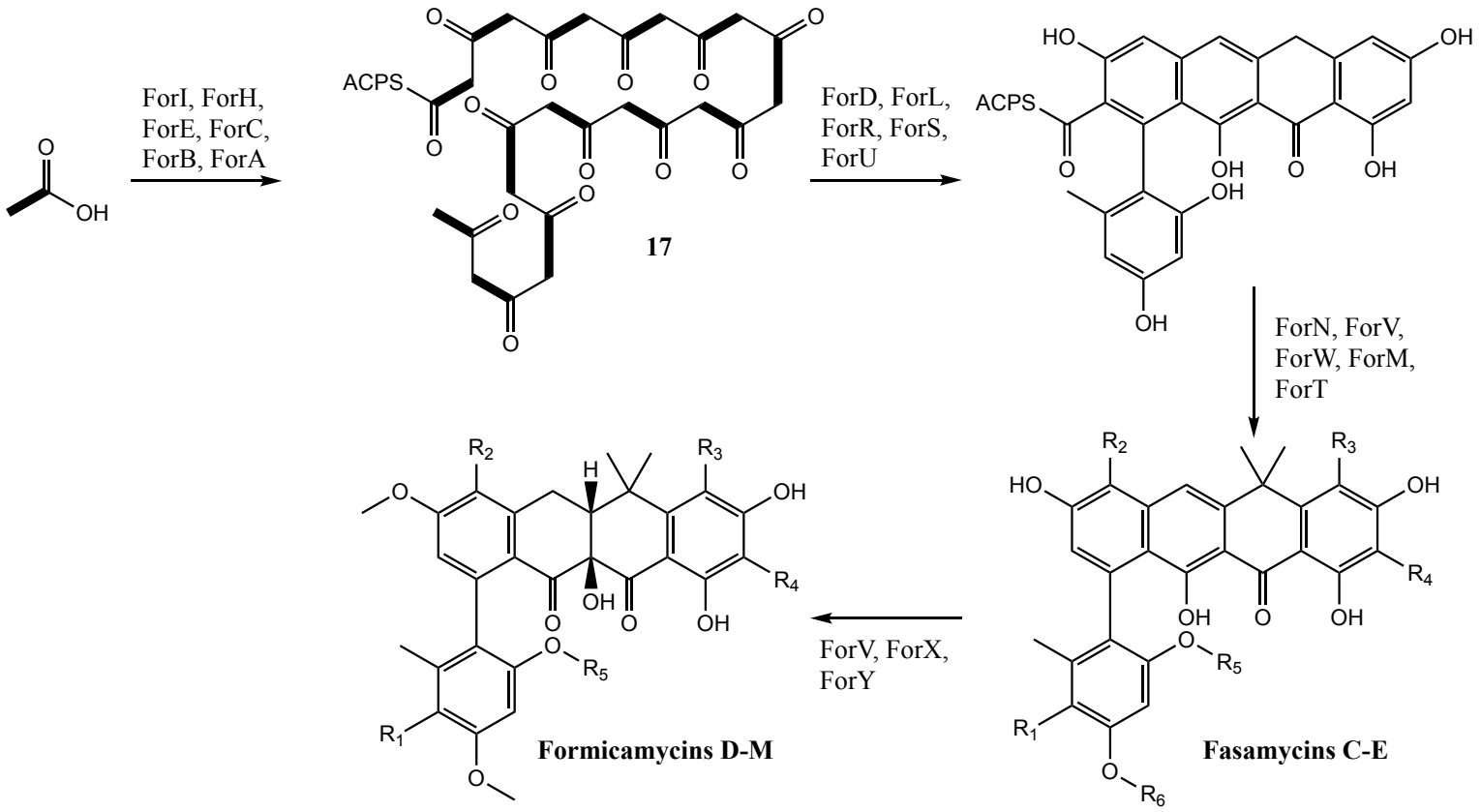


Figure 3.14: Proposed biosynthetic pathway for the fasamycins and formicamycins. Bold bonds in compound number 17 indicate the positions of the [1.2- $^{13}C_2$] sodium acetate units incorporated into the polyketide backbone during feeding experiments conducted by Zhewei Qin (Qin *et al.*, 2017).

3.5 Mode of action of the formicamycins

Having established that the formicamycins exhibit potent antibacterial activity, attention turned to identifying the mechanism of action of these compounds. As described above, the formicamycins are predicted to be derived from their biosynthetic precursors; the fasamycins. Fasamycins A and B were previously isolated through the heterologous expression of environmental DNA (eDNA) cosmid libraries that were constructed using DNA isolated from soil samples collected in Arizona (Feng, Kallifidas and Brady, 2011). Initial bioactivity studies showed that these compounds, like the fasamycins and formicamycins from *S. formicae*, are antibiotics with specific activity against Gram-positive bacteria including MRSA and VRE. A later study by the same researchers suggests fasamycins A and B target type II fatty acid biosynthesis, specifically by inhibiting FabF, with docking studies indicating the tricyclic chloro-*gem*-dimethyl-anthracenone substructure is the key pharmacophore, much like in other FabF inhibitors (Feng *et al.*, 2012). When comparing this region of the fasamycin molecule to the formicamycins, there are significant differences in the three dimensional structure, therefore it is possible that the formicamycins may have a distinct mode of action, or at least different affinities for their target (**Figure 3.15**) (Qin *et al.*, 2017).

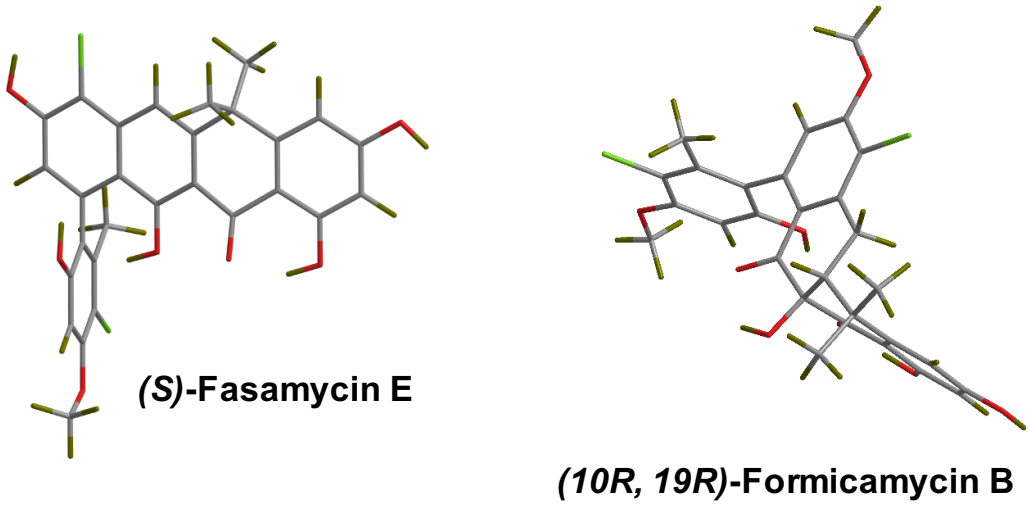


Figure 3.15: Comparison of the lowest energy conformers of fasamycin E and formicamycin B. The fasamycins appear to be much flatter than the formicamycins. The latter compounds have a bend in the tetracyclic backbone that makes them occupy a different 3-dimentional space (Qin *et al.*, 2017). Image generated by Dr Zhiwei Qin based on NOESY NMR data and reproduced with permission.

To investigate this further, multiple reporter strains were obtained from Professor Jeff Errington (Newcastle University). The panel of six *B. subtilis* reporter strains were generated by fusing a *lacZ* gene to various promoters to indicate responses to specific antimicrobial stresses. The reporter strains provided were sensitive to either inhibition of fatty acid biosynthesis (*fabHA*), cell wall damage (*ypuA*), DNA gyrase inhibition (*PL39*), DNA damage (*ϕ105*), RNA polymerase inhibition (*helD*) or cell envelope stress (*lial*) (Kepplinger *et al.*, 2018). Assays were conducted using fasamycin E and formicamycin J, due to the availability of these compounds in high titres from biosynthetic mutants created during this project (described later) and the fact that these were identified from previous bioassays to be among the most potent compounds from each class. Each assay was conducted at least three times under the same conditions, with at least two technical replicates being conducted in parallel on each attempt. The results reported here are a representative summary of all the assays conducted during this project (**Figure 3.16**).

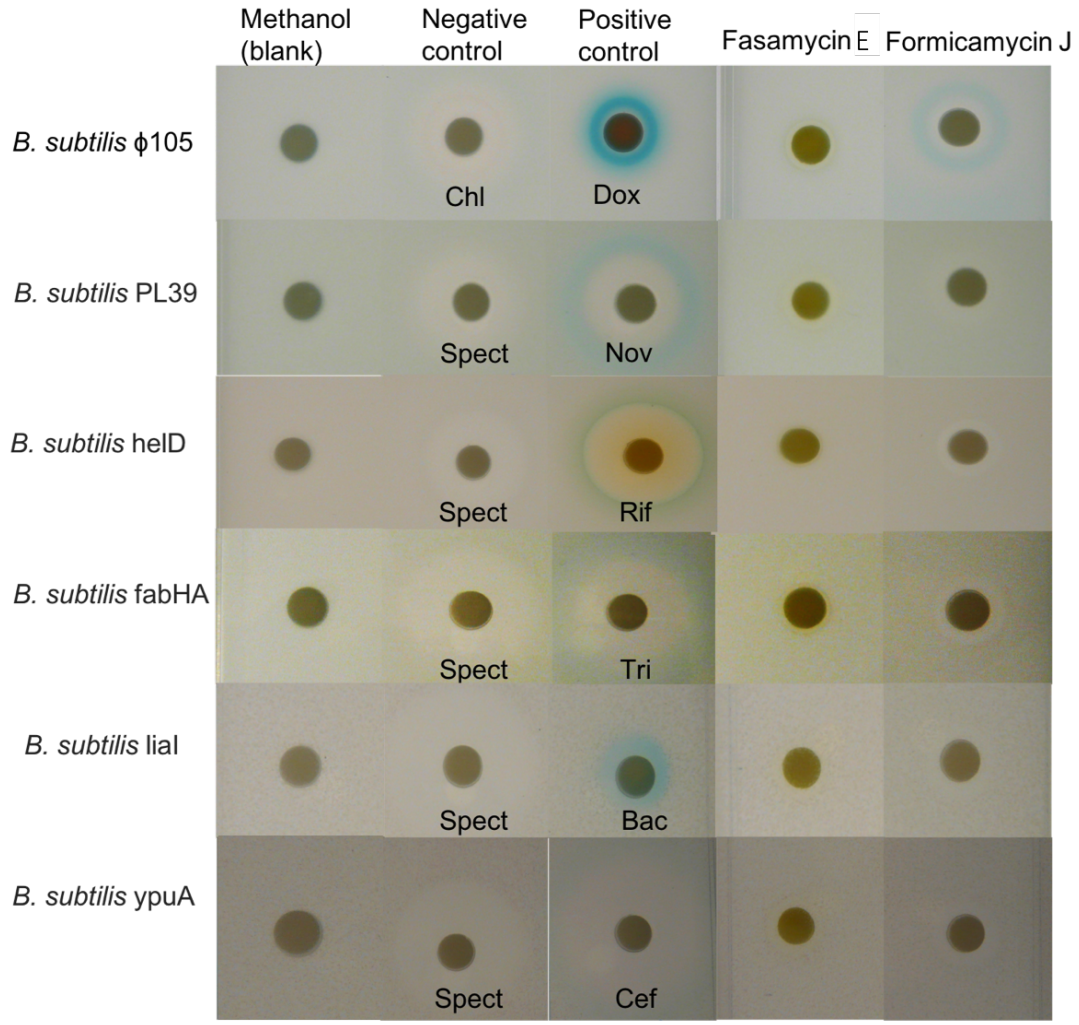


Figure 3.16: Paper discs with 40 μg fasamycin E or formicamycin J (or methanol control) were placed on soft LB (0.5% agar) inoculated with 1:10 *B. subtilis* at OD_{600} 0.4. *B. subtilis ϕ105* (*lacZ* fusion to a late promoter in a ϕ105 prophage) is the DNA damage reporter, *B. subtilis* PL39 *gyrA*::pMUTIN4 *ermC* gyrA'-lacZ PspacgyrA+ is the DNA Gyrase inhibition reporter, *B. subtilis helD*::pMUTIN4 *helD-lacZ ermC* is the RNA polymerase inhibition reporter, *B. subtilis fabHA*::pMUTIN4 *ermC gyrA'-lacZ* Pspac-*gyrA+* is the fatty acid synthesis inhibition reporter, *B. subtilis lial*::pMUTIN4 *lial-lacZ ermC* is the cell envelope reporter, *B. subtilis ypuA*::pMUTIN4 *ermC ypuA'-lacZ* is the cell wall damage reporter. Negative control antibiotics are those that will be active against the *B. subtilis* strains but will not trigger a reporter reaction as they act via a different mechanism of action. In contrast, positive control antibiotics are those that will trigger a reporter reaction as they act via the desired mechanism of action. Chl= chloramphenicol (125 μg/disc); Spect = spectinomycin (1 mg/disc); Dox = doxorubicin (10 μg/disc); Nov = novobiocin (10 μg/disc); Rif = rifampicin (10 μg/disc); Tri = triclosan (150 μg/disc); Bac = bacitracin (500 μg/disc); Cef = cefotaxime (125 μg/disc). Formicamycin J produced a positive reporter result when tested against the DNA damage reporter. Fasamycin E did not elicit a positive reporter response in any of the strains tested.

As expected, the methanol negative control produced no reporter activity or bioactivity in any of the experiments conducted. The negative control antibiotics display bioactivity with an absence of reporter activity and the positive control

antibiotics show positive bioactivity and reporter activity. For both fasamycin E and formicamycin J there is some level of bioactivity against each reporter strain. This is expected, as the compounds exhibit potent activity against Gram-positive bacteria, including *B. subtilis*. In every assay conducted with these compounds, the only positive reporter activity for formicamycin J was against the ϕ105 reporter indicative of DNA damage. There were no positive results for fasamycin E against any of the reporter strains tested. However, this does not necessarily mean that the fasamycins do not inhibit fatty acid biosynthesis as stated by Brady and colleagues in previous research. The reporter strain for fatty acid biosynthesis used here is based on the inhibition of FabH. Biochemical studies conducted by Feng *et al*. indicate that fasamycins A and B specifically bind to FabF and rule out any interaction with FabH. It is likely that fasamycins C-E behave in the same way and therefore no positive reaction would be expected with the reporter strain used here. Furthermore, it may be that higher concentrations of compounds are needed to get positive results for this type of reporter assay.

From this, we began investigating the possibility that formicamycins cause DNA damage. The positive control for this assay, doxorubicin, is a natural product of *Streptomyces peucetius* var. *casieus.* It is an anthracycline polyketide most widely used as an anticancer agent, however, it also displays antibiotic activity due to its ability to intercalate DNA (Westman *et al.*, 2012). The anthracyclines are a group of *Streptomyces* natural products produced by T2PKS clusters that are characterised by a planar anthraquinone chromophore that can intercalate between adjacent base pairs of DNA (Rayner and Cutts, 2014). Some examples of known anthracyclines are shown below (**Figure 3.17**).

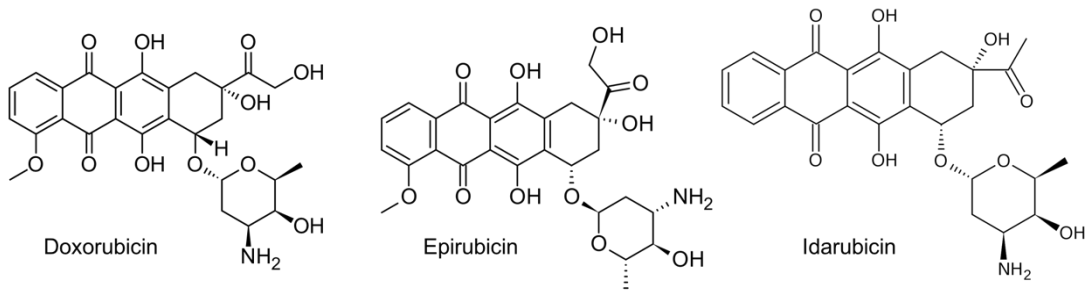


Figure 3.17: The structures of the fasamycins and formicamycins were compared to known anthracycline antibiotics, some of which are shown. There are some similarities in the planar rings present between fasamycins and the anthracycline antibiotics which may be indicative of a similar mode of action.

The fasamycins have similar planar ring structures compared to these other known DNA binding compounds and therefore may be able to interact with DNA. In contrast, the bend between the two central rings of the formicamycins send the structure off-plane, and therefore likely inhibits DNA binding. To confirm this, DNA was titrated onto a fixed concentration of Fasamycin E and Formicamycin J, and absorbance of the compound was measured between 300 and 500 nm using the Nanodrop UV Spectrophotometer. Although there should be no absorption of DNA in this region, DNA + solvent controls were run in parallel and used to calculate the absorbances of compound only (**Figure 3.18**). The data confirm that addition of DNA to fasamycin E causes a change in the absorbance of the compound in a concentration dependent manner. This is likely reflecting some interaction between the DNA and the compound as a similar shift in absorbance is seen when doxorubicin, a known DNA intercalator, is titrated onto ctDNA (Agudelo *et al.*, 2014). However, addition of DNA to formicamycin J caused no change in absorbance at any of the DNA concentrations tested, implying there is no interaction between them, at least at the tested concentrations.

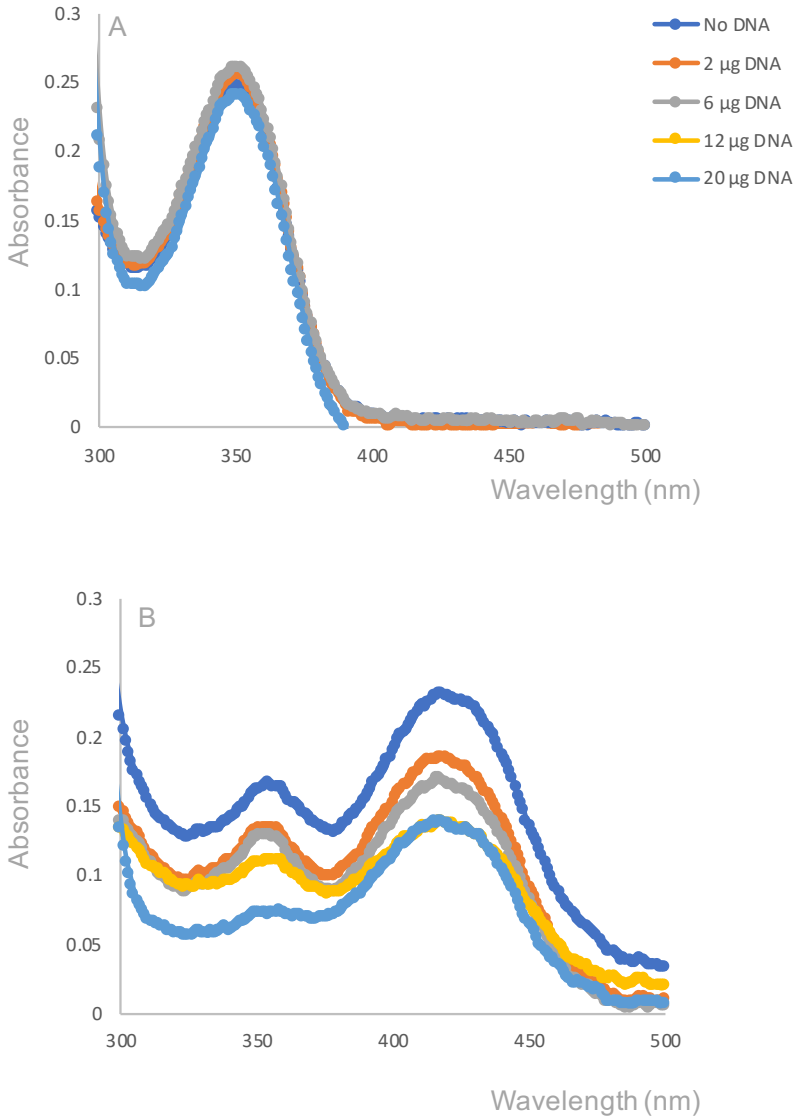


Figure 3.18: Increasing concentrations of calf-thymus DNA (ctDNA) was added to a fixed concentration (3 µg) of either formicamycin J (A) or fasamycin E (B), in PBS buffer in a total volume of 20 µl. Absorbance of the compound was measured between 300 and 500 nm. Appropriate controls were run to account for any effect of increasing DNA concentration, the buffer and the solvent the ligand was in (DMSO). No significant change in absorbance of formicamycin J is seen on addition of increasing concentrations of DNA. However, increasing the concentration of DNA does appear to change the absorbance of fasamycin C in a concentration dependent manner.

This DNA binding activity of the fasamycins does not explain the observations made in the reporter assay experiment, where formicamycins appeared to cause DNA damage but fasamycins did not. Anthracycline antibiotics are also known to target bacterial Type II DNA topoisomerase enzymes, DNA gyrase and topoisomerase IV. These enzymes function together to dictate the topological organisation of DNA

(Mitscher, 2005). DNA Gyrase catalyses the super-coiling of double-stranded DNA while Topoisomerase IV relaxes supercoils in order to remove knots and tangles in the bacterial chromosome (Reece and Maxwell, 1991). As part of this function, these enzymes generate transient double-stranded breaks in order to allow double helices to pass each other. As a result, they are major targets for several other classes of antibiotics, including the quinolones (Levine, Hiasa and Marians, 1998). Quinolone antibiotics bind non-covalently at the enzyme-DNA interface of the cleavage complex and physically block the re-ligation of DNA after cleavage. As a result these so called 'topoisomerase poisons' essentially convert gyrase and topoisomerase IV into cellular toxins that cause double strand breaks in DNA, triggering programmed cell death (Aldred, Kerns and Osheroff, 2014).

To investigate whether fasamycins and/or formicamycins are able to inhibit type II DNA Topoisomerase enzymes, extracts were provided to Inspiralis on the Norwich Research Park who provide high throughput topoisomerase inhibitor screening assays. The fasamycins and formicamycins were tested for both inhibition of topoisomerase IV relaxation and inhibition of gyrase supercoiling and compared to the quinolone antibiotic, ciprofloxacin, a known inhibitor of both processes. Enzymes from both Gram-positive (*S. aureus*) and Gram-negative (*E. coli*) organisms were also used, to determine whether the compounds have higher affinity for one over the other. Supercoiled plasmid substrate was incubated with Topoisomerase IV and DNA Gyrase was incubated with relaxed plasmid, either with or without the drug present, and the resulting DNA visualised using agarose gel electrophoresis. Supercoiled DNA runs further through a gel in a given time compared to relaxed DNA. Inhibition levels (% supercoiled or relaxed DNA) were measured by determining the relative fluorescence of the band using gel scanning software and used to calculate IC_{50} values for the most active compound from each group (personal communication with Inspiralis, 2018) (**Figure 3.19**).

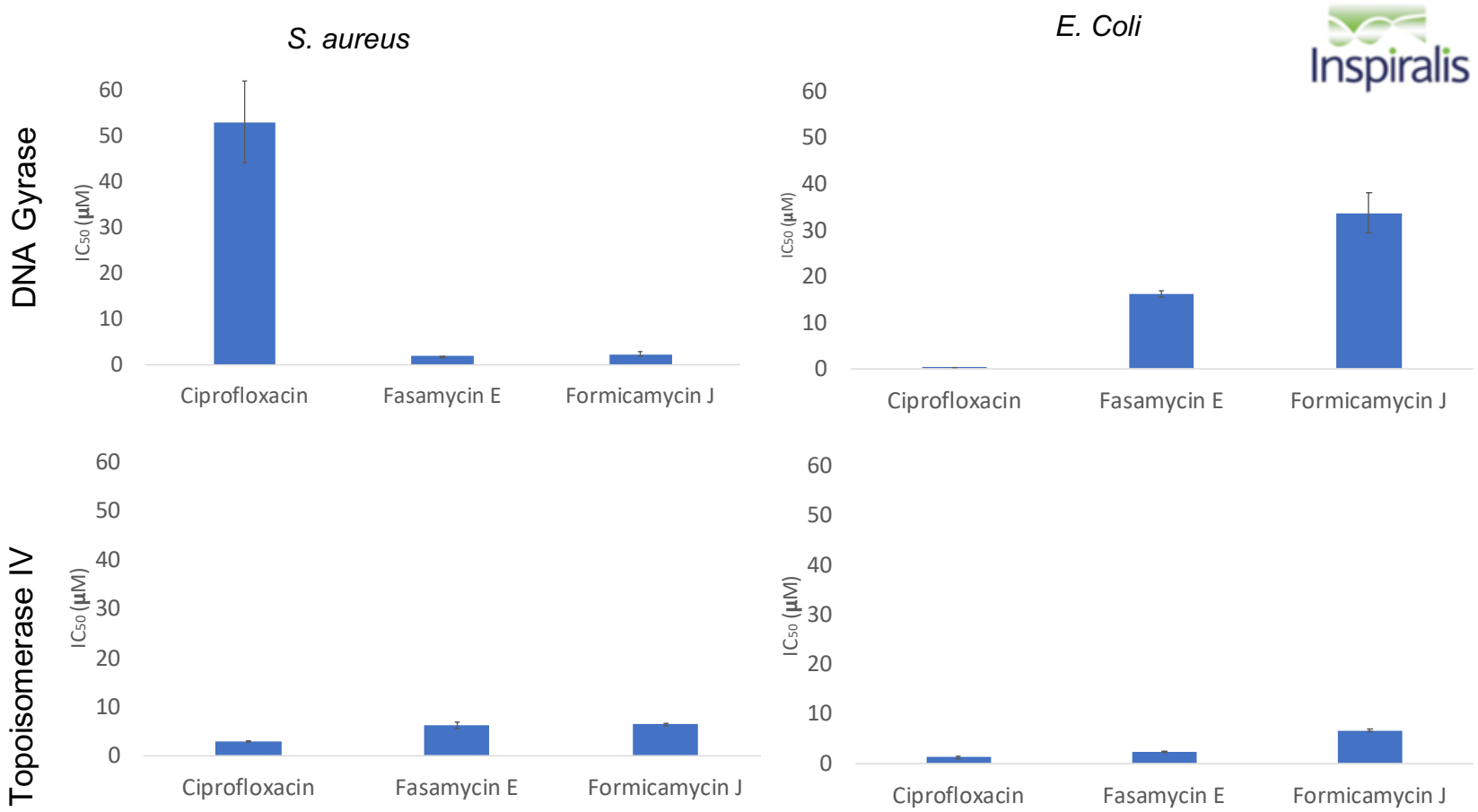


Figure 3.19: IC_{50} values were calculated for the most potent compounds in each group by Inspiralis. These values were plotted in Excel and compared with the known topoisomerase inhibitor, ciprofloxacin. The potency of the fasamycins and formicamycins compares well to ciprofloxacin against all enzymes tested.

IC_{50} is the concentration of an inhibitor required to reduce the activity of a target by 50% and is often used as a measure of potency of a compound. From these data, it appears that fasamycins and formicamycins inhibit DNA topoisomerase enzymes with a potency that is comparable to the known topoisomerase inhibitor, ciprofloxacin, under all conditions tested. The inhibition of the Gram-positive DNA Gyrase by the fasamycins and formicamycins is significantly stronger than the ciprofloxacin control. Inhibition of Gram-positive Topoisomerase IV is the same as the ciprofloxacin control. Interestingly, the fasamycins and formicamycins also appear to have relatively potent activity against Gram-negative enzymes, particularly Topoisomerase IV, when only Gram-positive activity has been seen *in vivo*.

Based on these data, we hypothesise that fasamycins and formicamycins are able to inhibit bacterial type II DNA topoisomerase enzymes. This could explain the DNA damage detected in the reporter strains, as this mechanism of action might result in double stranded breaks in the bacterial chromosome or dysregulated topological organisation of the bacterial chromosome. Although there was no reporter activity in the PL39 reporter strain, designed to detect inhibition of DNA Gyrase, it is possible that the reporter strain generated was not sensitive to inhibition by fasamycins and/or formicamycins if these compounds act via a novel mechanism. The bacterial DNA Gyrase enzyme is made up of two distinct functional subunits, known as GyrA and GyrB. GyrA is the subunit responsible for the interactions with DNA whereas GyrB contains the ATPase active site (Collin, Karkare and Maxwell, 2011). The in vitro data from Inspiralis clearly shows these compounds are able to inhibit the activity of topoisomerase enzymes, however, we currently have no information about the mechanism by which this occurs and where on the enzyme these compounds bind. Novobiocin, the recommended positive control used in the *Bacillus* reporter assays, is known to target GyrB, but quinolones and anthracyclines are known to bind GyrA, which contains the active site that attaches to cleaved DNA (Aldred, Kerns and Osheroff, 2014). In future, it will be important to obtain ligand-bound crystal structures of these topoisomerase enzymes in the presence of fasamycins and formicamycins to investigate this interaction further.

All microorganisms that encode for and produce antimicrobial secondary metabolites must also encode a resistance mechanism in order to evade self-toxicity. These can either modify the target, and this can help identify the mechanism of action, or increase the efflux of the compounds from the cell via transport systems (D. A. Hopwood, 2007; Mak, Xu and Nodwell, 2014). Manual investigation of the formicamycin BGC and the surrounding genes does not indicate any obvious gene candidates for self-resistance to the formicamycins by *S. formicae.* We predict that the two transporters encoded in the *for* BGC provide some level of host-resistance via an export mechanism and this is investigated further later in this thesis (**Chapter 4.4**). However, compounds will still accumulate intracellularly, therefore the host must have some mechanism of resistance encoded within its genome to overcome self-toxicity. Analysis using Antibiotic Resistance Target Seeker (ARTS) does identify known resistance mutations in the genes encoding for both the DNA gyrase A subunit (KY5_4068) and the DNA gyrase/topoisomerase B subunit (KY5_4067). Furthermore, a search of the annotated *S. formicae* genome on StrepDB reveals two additional genes, KY5_5972c and KY5_5956, which are also predicted to encode for a DNA topoisomerase IV subunit A and a type IIA DNA topoisomerase subunit B with 42.44% identity (91% coverage) and 44% identify (96% coverage) to KY5_4068 and KY5_4067 respectively. This implies that, as well as having known resistance mutations in the topoisomerase machinery, there may have been gene duplication events over the course of evolution that might enable *S. formicae* to be resistant to fasamycins/formicamycins during biosynthesis and before export.

As discussed above, previous work has shown that fasamycins also inhibit fatty acid biosynthesis. It is possible that fasamycins and formicamycins are able to target both fatty acid synthases and DNA topoisomerases but with different affinities. If these compounds are able to target both fatty acid synthases and DNA topoisomerases, this would make these compounds unique in respect to other antibiotics, in that they can inhibit two distinct molecular targets. This dual-target mechanism of action might explain the lack of resistant mutants generated *in vitro* during earlier work (**Chapter 1.8**), but further investigation is needed to confirm this theory. Developing a comprehensive structure-activity relationship using *in vitro* gyrase/topoisomerase

inhibition assays, DNA cleavage assays and *in vitro* fatty acid synthase inhibition assays will be the priority going forward. If crystals of the inhibitor-target complex could be obtained, we could confirm the interaction with topoisomerase enzymes and therefore gain information about the potential toxicity of the compounds if they were ever to be considered for clinical use. Furthermore, the generation of resistant mutants to provide information on the mechanisms of resistance to these compounds and therefore the likely mechanism of action will be key. Although previous work in the Hutchings laboratory did not successfully obtain resistant mutants to these compounds, the experiment was only conducted using MRSA. Using a different susceptible strain, such as *Enterococcus faecalis* as in the previous study on fasamycin mechanism of action, may be more successful (Feng *et al.*, 2012). This will be the subject of future investigation in the Wilkinson laboratory.

Finally, while the compounds did not show any activity against Gram-negative bacteria *in vivo*, the results presented here show that there is some affinity towards enzymes from Gram-negative bacteria *in vitro*. Gram-negative bacterial pathogens are infamously challenging to treat with antibiotics due to the layer of lipopolysaccharide (LPS) in the outer membrane that forms an impermeable barrier to prevent compounds entering the cell. When this cell envelope is disrupted, large-scaffold antibiotics are taken up more efficiently (Muheim *et al.*, 2017). In accordance with this, previous work has shown that fasamycins A and B are active against membrane permeabilised *E. coli*, suggesting that the absence of activity in Gram-negative bacteria may be due to the inability of the compounds to access their target, rather than a decreased affinity (Feng *et al.*, 2012). As such, a strain of *E. coli* with an in-frame deletion of *imp*, a gene encoding for a protein involved in the assembly of LPS in the outer membrane (*E. coli* NR698), was obtained from the John Innes Centre and assayed against fasamycin C and formicamycin J, the most potent fasamycin and formicamycin against Gram-positive pathogens isolated from *S. formicae*. In contrast to all previously conducted assays of these compounds against Gram-negative organisms, there were significant zones of inhibition for both compounds when tested against *E. coli* NR698 (**Figure 3.20**).

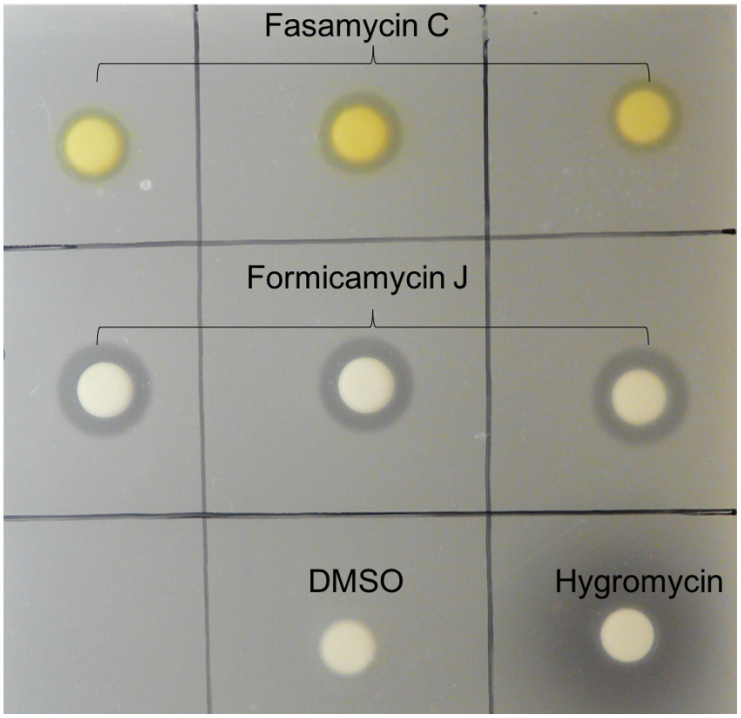


Figure 3.20: 20 uL of approximately 300 ng/µl fasamycin C and formicamycin J were dried on Whatman paper discs (approximately 5 mM of each compound per disc). Discs were placed on 0.5% agar LB plates inoculated with *E. coli* NR698 and incubated at 30°C overnight. The zones of inhibition against this membrane-permeabilised strain imply that the fasamycins and formicamycins do act on the intracellular target of Gram-negative cells, however, they cannot cross the normal LPS-rich Gram-negative cell membrane.

This further suggests that these compounds are in fact broadly bioactive against both Gram-positive and Gram-negative pathogens but cannot cross the Gram-negative outer membrane. In future, analogues of the formicamycins that are more membrane permeable could be chemically synthesised. Alternatively, the formicamycins could be co-administered with small molecules that increase the permeability of Gram-negative cell membranes. For example, the small molecule MAC13243 is an inhibitor of LolA, a chaperone that traffics lipoproteins from the inner membrane to the outer membrane. It has antimicrobial activity against Gram-negative pathogens but degrades in aqueous solution and therefore was concluded to not be a clinically relevant lead compound. However, sub-inhibitory concentrations of MAC13243 have been shown to increase the uptake and resultant bioactivity of large-scaffold antibiotics such as novobiocin and erythromycin (Muheim *et al.*, 2017). If this or another similar small molecule could be found to

increase the uptake of formicamycins by Gram-negative cells, their clinical potential could be significant.

3.6 Discussion

In this chapter, the genes responsible for fasamycin and formicamycin production were successfully identified using CRISPR/Cas9-mediated genome editing of the new species *Streptomyces formicae*. Since 2015, CRISPR/Cas9 genome editing has been applied to several streptomycetes and other rare actinomycetes, enabling the manipulation of environmental isolates with greater efficiency than ever before (Tao *et al.*, 2018). The CRISPR system does not rely on the availability of cosmid libraries, only a good quality genome sequence, something which is becoming more accessible as sequencing becomes cheaper and easier to do. That said, not all strains are genetically tractable using CRISPR and often, the use of these tools in new organisms requires extensive optimisation. The development of the protocols used in this chapter will allow further genetic work to be carried out in *S. formicae*, including additional work on the formicamycin BGC, such as identification and generation of pathway intermediates and novel bioactive analogues.

So far, we have used the pCRISPomyces-2 system to show that two related but distinct compound families, the fasamycins and the formicamycins, originate from a single BGC in *S. formicae,* by generating strains with the whole BGC deleted. The deletion of the formicamycin BGC (approximately 46 Kbp) represents the largest published single deletion conducted using the pCRISPomyces-2 system recorded to date (Tong, Weber and Lee, 2018). Furthermore, we were able to complement this large deletion using DNA from an artificial chromosome library, all in the native host. In addition, the work in this chapter identified the core genes required for the biosynthesis of these compounds and began to identify the mechanisms of some of the novel enzymes encoded by this genetic machinery. Later in this thesis, further work will investigate how these enzymes work together to synthesise these compounds and experiments will be conducted to determine the biosynthetic link between the fasamycin precursors and the more potent formicamycins. The fact that

these novel compounds originate from a previously uncharacterised environmental isolate may mean that we uncover novel mechanisms of biosynthesis that can be used to inform future drug discovery efforts.

Antimicrobial resistance is a major threat to modern healthcare and the discovery of new antimicrobials will be vital in combating this threat. Anything we can learn about the biosynthesis of natural products is valuable, however, priority should be given to compounds that display novel mechanisms of action as these represent more promising candidates for clinical development as they may be effective against drug-resistant infections. The work in this chapter has begun to assess the suitability of the fasamycins and formicamycins as clinically relevant antibiotics. We have shown that these novel compounds display potent antibacterial activity against Gram-positive pathogens including multi-drug resistant isolates like MRSA and VRE and have also shown that they have the potential to inhibit Gram-negative pathogens too. Although the fasamycins and formicamycins have distinct three-dimensional structures, the evidence so far suggests that they may act via the same mechanism, specifically by targeting DNA topoisomerases and possibly also fatty acid synthases. If we can prove that these compounds can indeed inhibit two distinct intracellular targets using the experiments described above, the formicamycins would be unique amongst antibiotics and potentially very attractive for further clinical development.

Overall, this work shows that by searching underexplored environments and using new genomic techniques, novel anti-infectives can be identified that may bypass current antimicrobial resistance mechanisms. In future, the development of these antibiotics into clinically relevant medicines may rely on mass production on an industrial scale. So far, *S. formicae* only produces the fasamycins and formicamycins during growth on agar plates, making extraction protocols lengthy and complex. If we could identify a mutation that makes *S. formicae* produce these compounds during liquid culture, perhaps by changing the way the host regulates the biosynthesis of these compounds, there would be great potential for using *S. formicae* for the production of these clinically relevant antibiotics in the future.

4 Regulation of formicamycin production in *Streptomyces formicae* and mechanisms of host resistance

The formicamycins are polyketide antibiotics produced by *S. formicae* that represent promising new compounds for clinical development. Antibiotics are produced by microorganisms as part of their secondary metabolism and are not essential for growth. The production of these highly decorated molecules can be metabolically expensive, therefore expression of BGCs is very tightly regulated and controlled by the native host to ensure that transcription is only activated when the production of the secondary metabolite will provide a significant survival advantage (Chater, 2016; van der Meij *et al.*, 2017). As such, microbial genomes contain a large number of transcriptional activators and repressors that together form a complex regulatory network that controls the expression of biosynthetic genes (Busby, 2019). For most secondary metabolites, one or more regulatory proteins are encoded in the BGC and these are responsible for controlling the expression of all the genes within the cluster. These are known as cluster situated regulators (CSRs). Work presented in the previous chapter identified the genes required for formicamycin biosynthesis and defined the borders of the BGC. Several putative CSRs are encoded within the *for* BGC, including two MarR-family regulators and two two-component systems (TCSs). The aim of this chapter is to investigate the role of these CSRs and determine how production of the formicamycins is controlled in the native host.

In addition, work in the previous chapter began to elucidate the mechanism of action of the fasamycins and formicamycins. Previous attempts to raise mutants of MRSA with resistance to these potent antimicrobials was unsuccessful, therefore identifying the mechanism of action remains a challenge. As discussed in the previous chapter, all antibiotic producing microorganisms will encode a host-resistance mechanism to evade self-toxicity upon compound biosynthesis. Within the formicamycin BGC in *S. formicae*, there are two putative transporter genes. These are predicted to export fasamycins and formicamycins across the cell membrane and provide some host-resistance by preventing toxic intracellular accumulation. The roles of these gene products will be investigated during this

chapter. By identifying the host regulation and resistance mechanisms, it may be possible to generate mutant strains of *S. formicae* with increased formicamycin biosynthesis capabilities, which could be useful for the industrial production of these clinically relevant molecules.

4.1 Determining the transcriptional organization of the formicamycin BGC

The regulation of antibiotic production begins with the initiation of transcription within the BGC (Romero *et al.*, 2014). Transcription is the process by which the DNA encoding for a gene is transcribed into messenger RNA (mRNA) by RNA polymerase (RNAP). This mRNA will be translated into the functional protein by the ribosome. The first step of bacterial transcription initiation is promoter recognition by RNAP. Identifying the promoter of a putative gene or operon can thus inform how the transcription is controlled by the binding of transcription factors. For example, binding of a transcriptional activator or the release of a transcriptional repressor, can increase RNAP binding to a promoter sequence within the DNA. This results in the DNA strands around the transcription start site unwinding and the template strand entering the catalytic site of the RNAP for transcription of the gene and generation of the mRNA (Busby, 2019). mRNA usually consists of a 5' untranslated region (5'UTR) upstream of the translation start site which is important for regulation of translation of the transcript via binding of the ribosome at the Shine-Dalgarno sequence. This promoter region is usually present within intergenic regions of DNA directly upstream of the start codon of the gene.

Identifying transcription start sites within secondary metabolite BGCs can be complex, because often they contain genes whose coding regions overlap with the gene either upstream or downstream, leading to the generation of multi-gene transcripts or polycistronic mRNAs. Genes with overlapping DNA sequences are thought to have evolved due to translational coupling (Dangel *et al.*, 2009). For example, the biosynthesis of secondary metabolites often requires multi-enzyme complexes like the alpha and beta subunits of the KS enzyme required for polyketide biosynthesis. Thus, it is evolutionarily advantageous to encode for both enzyme units on the same transcript, as the functionality of one enzyme directly relies on the

presence of the other. As such, these genes are usually also co-located in operons within BGCs. In the formicamycin BGC, many of the genes encoding for the major biosynthetic machinery are co-located and overlapping. The overlapping genes *forH* and *forI* are predicted to encode the enzymes that synthesise the malonyl-CoA starter unit and the *forA, forB* and *forC* genes encode the core PKS. The formicamycin BGC contains various length intergenic regions between the coding regions of DNA, from short regions of just a few base pairs up to large regions of several hundred base pairs (**Figure 4.1**). In general, longer intergenic regions are indicative of promoters that drive the transcription of one or multiple downstream genes, for example the 247 bp region upstream of *forM*, which is predicted to contain a promoter driving the transcription of the *forMLKJ* transcript. Adding to the complexity is the occurrence of genes in different orientations within a single BGC. Divergent promoters that control the transcription of adjacent genes in opposite directions are common and there are predicted to be several within the formicamycin BGC, for example the region between *forH* and *forG* and between *forT* and *forU*.

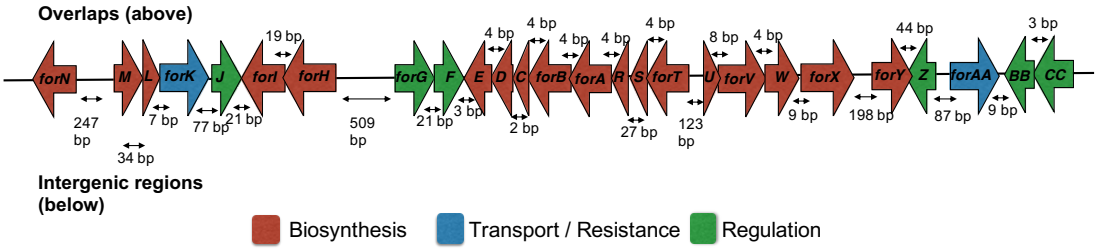


Figure 4.1: The formicamycin BGC with overlapping coding sequences of DNA labelled (above) and intergenic regions (below). In general, longer intergenic regions are indicative of promoter regions. Overlapping DNA sequences are probably transcribed on a single mRNA transcript. Several of the intergenic regions within the formicamycin BGC could contain divergent promoters that control the transcription of multiple adjacent genes in opposite directions.

RNA sequencing allows the highly dynamic bacterial transcriptome, or the complete set of expressed mRNAs within an organism, to be characterised under different conditions. This can be useful to learn more about which genes are expressed under certain conditions, but it does not map the transcript start sites (TSSs). More recently, methods of using RNA-sequencing to identify TSSs across bacterial genomes have been developed and these enable the promoter regions of genes to be

identified by pin-pointing the start of the 5'UTR. Cappable RNA sequencing is similar to differential RNA sequencing (dRNA-Seq) but instead of depleting 5' degraded RNA molecules, the 5'end of full length, primary transcripts are specifically targeted and enriched for sequencing by enzymatically modifying the 5' triphosphorylated end with a selectable tag. In this way, the primary transcripts present in a cell can be compared with the monophosphorylated mRNA molecules in order to identify the TSSs. This allows the promoters in multi-gene operons to be identified, as well as indicating which genes are co-transcribed on polycistronic mRNAs (Ettwiller *et al.*, 2016). In order to determine the transcripts expressed from the formicamycin BGC, RNA was extracted from 2-, 3- and 4-day old cultures of wild-type *S. formicae* grown under formicamycin producing conditions and quantified using formaldehyde gel electrophoresis and Qubit analysis. RNA capping and sequencing was conducted by Vertis Biotechnologies AG. The cappable RNA sequencing technique was used to enrich the 5' ends of all the expressed full-length mRNA transcripts and identify all the TSSs (EMBL-EBI Array Express accession number E-MTAB-7975). The data were listed in a table of all the TSSs using the Vertis Biotechnologies AG bioinformatics platform to a single nucleotide resolution with enrichment scores for each. Significant peaks were called as those with enrichment scores of 10 or more compared to the untreated samples and these were annotated as TSSs. This revealed 11 significant TSSs within the formicamycin BGC and shows that the 26 genes within the cluster are expressed on 10 transcripts as described below (**Table 4.1**).

Table 4.1: Cappable RNA sequencing was used to locate the transcription start sites across the *S. formicae* genome. In this table, a summary of transcription start site locations within the formicamycin BGC are described. Transcription start sites are listed in relation to the position of the start codon of the predicted genes.

Transcript	Gene	Function	Chromosome position				TSS relative to ATG of first gene
			Start	End	TSS	DNA strand	
1	*forN*		7756911	7755139	7757095	-	-184
2	*forM*	Methyl transferase	7757158	7757943	7757123	+	-35
	forL	Cyclase	7757978	7758319			

	forK	Resistance transporter	7758327	7759763			
3	*forJ*	Transcriptional regulator	7759841	7760290	7759783	+	-58
4	*forH*	ACC biotin carboxylase	7763484	7761661	7763527	-	-43
	forI	ACC carboxyl transferase	7761679	7760312			
5	*forG*	Sensor kinase	7763994	7765085	7763765	+	-229
	forF	Response regulator	7765093	7765752			
6	*forE*	ACC biotin carboxy carrier protein	7766271	7765756	7771343	-	-53
	forD	Cyclase	7766729	7766268			
	forC	PKS ACP	7767022	7766732			
	forB	KS_{β}	7768266	7767019			
	forA	KS_{α}	7769525	7768263			
	forR	Cyclase	7769917	7769522			
	forS	Cyclase	7770265	7769945			
	forT	O-Methyltransferase	7771290	7770262			
7	*forU*	Cyclase	7771411	7771770	7771370 7772595	+	-41 Middle of *forV*
	forV	Halogenase	7771763	7773055			
	forW	O-Methyltransferase	7773052	7774077			
	forX	Monooxygenase	7774087	7775802			
	forY	Oxidoreductase	7775981	7776928			
8	*forZ*	Transcription regulator	7777403	7776885	7777403	-	+1 (leaderless)
9	*forAA*	Resistance transporter	7777492	7779033	7777443	+	-49
10	*forBB*	Response regulator	7779705	7779043	7780997	-	+1 (leaderless)
	forCC	Sensor kinase	7780997	7779702			

As predicted, most of the intergenic regions within the BGC contain TSSs and therefore must act as promoters for the genes either up- or downstream (**Figure 4.2**).

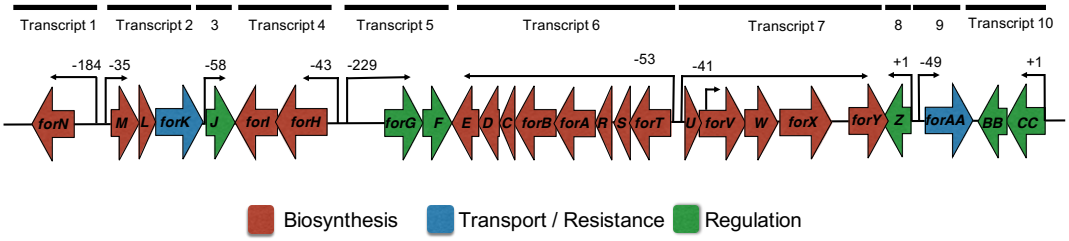


Figure 4.2: Visual representation of the TSSs identified within the formicamycin BGC in relation to the nearest start codon. The 26 gene formicamycin BGC is expressed on 10 transcripts as shown. The majority of the biosynthetic machinery is co-transcribed on two multi-gene transcripts, transcripts 6 and 7, which are under the control of a single, divergent promoter.

The acylhydrolase gene *forN* and the MFS family transporter gene *forAA* are expressed on their own as single transcripts. The expression of the acylhydrolase gene is driven by a divergent promoter that also contains the TSS for *forM*, *forL* and *forK*, encoding the putative methyltransferase, cyclase and sodium/hydrogen exchanger respectively. Somewhat surprisingly, the *forJ* gene, encoding a putative MarR-family regulator, is also under the control of its own promoter, even though the intergenic region upstream of the coding sequence is relatively small, at only 77 base pairs long. Most promoter sequences contain binding sites for transcriptional activators and repressors in addition to the RNAP, and therefore it is unlikely that this 77 base pairs alone represents the whole promoter region as this length of DNA is probably not long enough. We predict that the promoter region of *forJ* continues back into the coding region of the upstream *forK.* The genes encoding for the production of the starter units required for polyketide biosynthesis, *forH* and *forI,* are co-transcribed on a single transcript, likely resulting in translationally coupled proteins. Similarly, the TCS operons, *forGF* and *forBBCC* are expressed on single transcripts. This is common for TCSs, as the sensor kinase cannot function to control gene expression in the absence of the response regulator.

The majority of the formicamycin core biosynthetic machinery is expressed on two long transcripts, transcripts 6 and 7, which are under the control of the divergent promoters *pforU* and *pforT*. This suggests that co-transcription and translation is most efficient for these genes and may indicate that at least some of the biosynthetic

machinery from the formicamycin biosynthetic pathway functions as a multi-enzyme complex. Interestingly, although there is a 198 base pair intergenic region between *forX* and *forY*, there was no TSS present within this region in any of the sequenced samples. Thus, *forY* must be transcribed with *forUVWX* and the intergenic region between *forX* and *forY* must have an alternative role. It may still be some sort of regulatory element, but instead of acting as a promoter it could be an enhancer of downstream transcription or somehow affect translation, perhaps by changing the secondary structure of the mRNA. There is also a TSS within the coding region of *forV*. Internal promoters like this have been identified previously in *Streptomyces* genomes and are predicted to be important for maintaining the transcription levels in long operons where the distance of the terminal genes from the transcription start site can negatively affect the levels of transcription of downstream genes (Dangel *et al.*, 2009). In the formicamycin BGC, we predict that the internal promoter within *forV* has evolved to maintain transcription levels of *forY* at the end of the polycistronic mRNA transcript *forUVWXY*. Furthermore, transcription of the second putative MarR regulator gene in the formicamycin BGC, *forZ,* and the second TSS, *forBBCC,* are leaderless, with the TSS occurring on the start codon. This is a phenomenon in bacterial transcriptomics where an mRNA transcript does not contain the usual 5'UTR and Shine-Dalgarno sequence and therefore the start codon itself serves as the signal for initiation of translation. Although considered to be relatively uncommon in bacteria such as *E. coli*, leaderless transcripts have been found to be quite common in actinomycete genomes which have had their global TSSs mapped, accounting for approximately 20% of all the genes present in sequenced actinobacterial genomes deposited in the database (Zheng *et al.*, 2011; Romero *et al.*, 2014).

4.2 Characterising cluster situated regulators of formicamycin biosynthesis

Having identified the transcriptional organization of the formicamycin gene cluster and the locations of the promoters for the genes and operons within the BGC, it was important to identify the functions of the putative CSRs. CSRs can respond either directly to extracellular signals or indirectly via signals from global regulatory systems. They can either activate or repress transcription of secondary metabolite biosynthesis genes within the same gene cluster and/or, cross-regulate other BGCs in the genome. Often, CSRs encode for small DNA binding proteins that control transcription by binding to promoters or other regulatory regions within a BGC. The formicamycin BGC is predicted to encode two MarR-family regulators, ForJ and ForZ as well as two TCSs, ForGF and ForBBCC. It is likely that at least some of these putative gene products work together to activate or repress expression of the formicamycin biosynthesis genes.

ForBB is predicted to be a DNA-binding response regulator of the LuxR family and ForCC is predicted to be a sensor histidine kinase. Together, these genes likely form a TCS. Due to the location of these genes at the outer perimeter of the formicamycin BGC, deletion mutants of the whole TCS were generated using CRISPR/Cas9. *S. formicae* Δ*forBBCC* was grown under formicamycin producing conditions and the extracts were quantitatively analysed by HPLC(UV) MS by Dr Zhiwei Qin. Deletion of *forBBCC* had no effect on formicamycin or fasamycin production and was therefore judged to be outside the border of the minimal BGC (data not shown) (see **Figure 3.7**). The deletion mutant had no clear phenotype under the conditions tested, so function of the ForBBCC TCS is currently unknown but it is not predicted to be involved in the regulation of fasamycin/formicamycin biosynthesis.

4.2.1 The MarR-family regulator, ForJ

Bioinformatic analysis of ForJ shows that it is a putative DNA-binding transcriptional regulator of the MarR-family that contains a helix-turn-helix domain typical of these Multiple antibiotic resistance Regulators. MarR proteins often repress transcription by binding to palindromic sequences within the promoter regions of secondary metabolite BGCs. In order to determine the role of ForJ in regulating formicamycin biosynthesis, an in-frame unmarked deletion was generated in *forJ* using pCRISPomyces-2. The *S. formicae ΔforJ* strain was genetically complemented, but this experiment was conducted while waiting for the cappable RNA sequencing data, so the complementation construct had *forJ* under the control of *pforM* instead of its own native promoter as we assumed *forMLKJ* formed an operon. Three biological replicates of *S. formicae ΔforJ* and the complemented strain were grown under formicamycin producing conditions and extracts were analysed by HPLC/LMCS by Dr Zhiwei Qin to compare the levels of formicamycins in the samples to the wild-type (**Figure 4.3**). Deletion of *forJ* resulted in a large increase in the levels of all formicamycin congeners present in the extracts compared to the wild-type strain. Production levels decreased to wild-type levels on re-introduction of the *forJ in trans*.

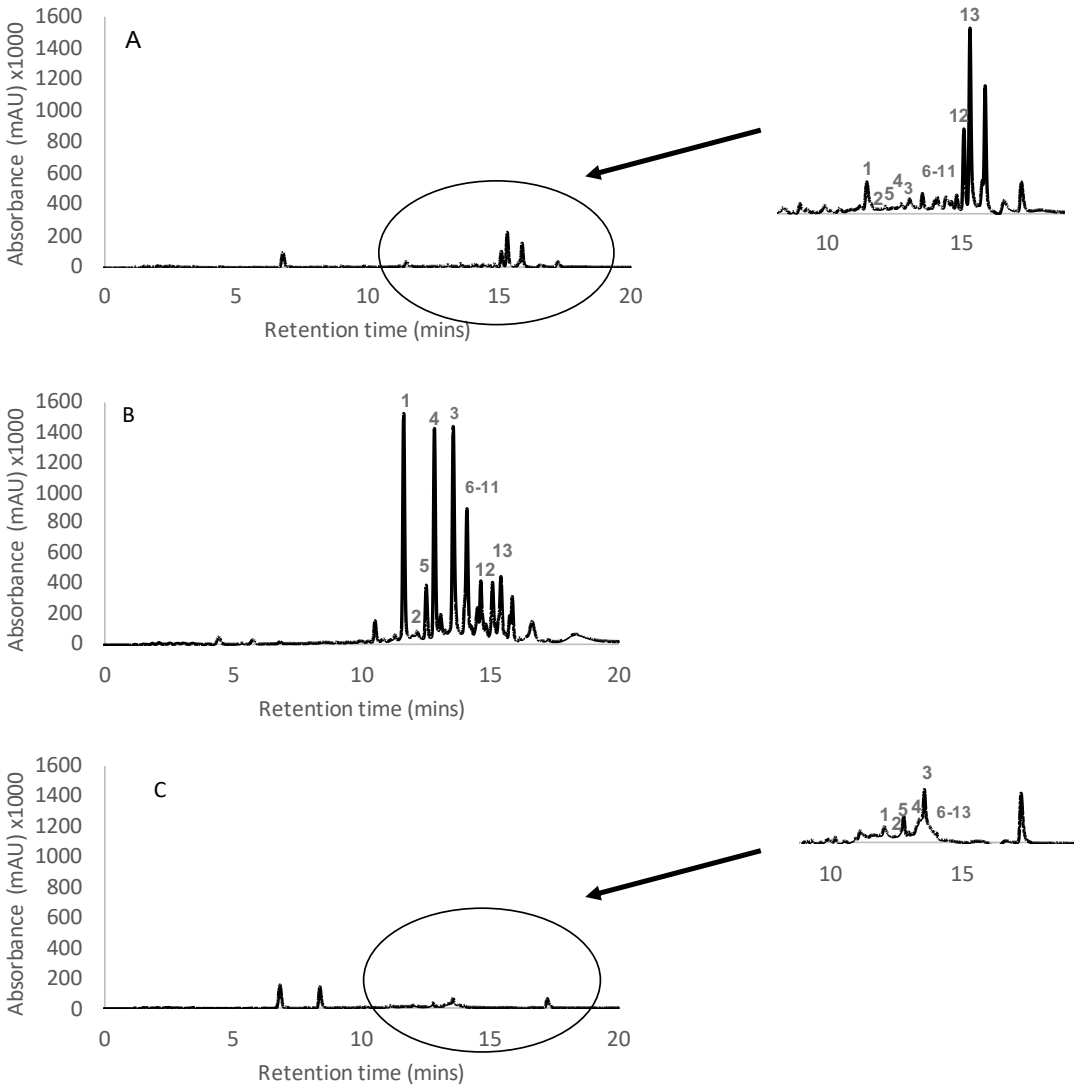


Figure 4.3: HPLC traces (UV 280 nm) showing extracts of: (A) *S. formicae* wild-type; (B) *S. formicae* Δ*forJ*; (C) *S. formicae* Δ*forJ*: φBT1 *forJ pforM*. Deletion of *forJ* results in increased accumulation of all fasamycins and formicamycins compared to the wild-type levels. Complementation of *forJ in trans* reduces formicamycin production to a similar level to the wild-type. Scales are decreased to demonstrate the accumulation of compounds in *S. formicae* Δ*for*J in a manner that is comparable across samples. The presence of formicamycin congeners was confirmed by LCMS and is indicated by UV traces with smaller scales in the inserts. Three biological replicates of each strain were generated for analysis. HPLC(UV) LCMS analysis conducted by Dr Zhiwei Qin at the John Innes Centre.

These data suggest ForJ is a repressor of the formicamycin BGC. When grown under formicamycin producing conditions next to a wild-type control, *S. formicae ΔforJ* also has a clear phenotype indicative of formicamycin over-production. The mycelia are very yellow compared to the wild-type and fasamycins are yellow when purified. Furthermore, *S. formicae ΔforJ* does not sporulate even after extended incubation on MYM agar at 30°C. In addition, the mutant only makes approximately 60% of the

biomass of the wild-type. This could be because over-production of these antibiotics is overwhelming the host-resistance mechanisms and resulting in self-toxicity that negatively affects the growth and development of the mutant. This phenotype is only partially rescued in the complementation mutant *S. formicae ΔforJ:ΦBT1 forJ pforM* (**Figure 4.4**). This is likely due to the fact that *pforM* is not the native promoter of *forJ* and it may have been more effective to use 200-300 base pairs upstream of the *forJ* gene to capture the TSS and the full promoter region. Alternatively, the change of location when the gene was introduced *in trans* may have affected the relative levels of expression. Nevertheless, the fact that reintroducing *forJ* under the control of *pforM* reduces the levels of formicamycin production compared to the *ΔforJ* mutant implies that over-production is a result of the *forJ* deletion and further suggests that *forJ* represses the transcription of some or all of the genes within the formicamycin BGC.

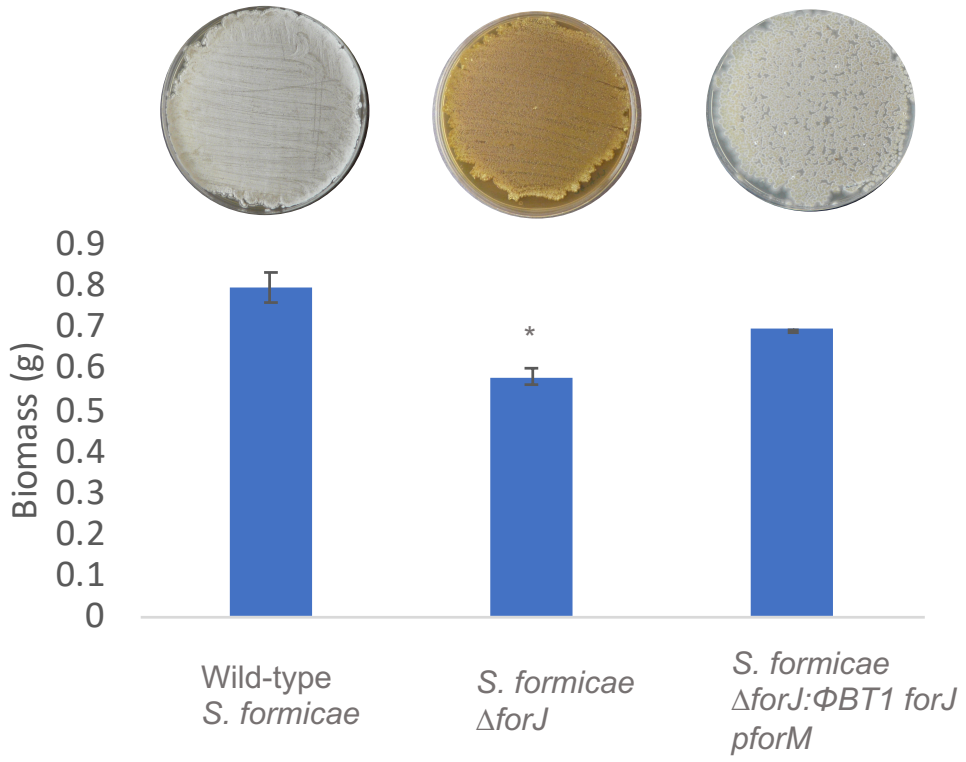


Figure 4.4: Deletion of *forJ* in *S. formicae* results in an extreme phenotype, with bright yellow mycelium forming. After 10 days incubation on MYM, the wild-type can be seen to be sporulating (white, fluffy appearance on a plate) whereas *S. formicae ΔforJ* does not sporulate. Similarly, the biomass of *S. formicae ΔforJ* is significantly lower than the wild-type (two-tailed T-test, p = 0.03). The complementation mutant shows a phenotype in-between that of the mutant and the wild-type. This is probably because *pforM* was used instead of the native promoter.

Previous attempts to heterologously express the formicamycin BGC were unsuccessful (data not shown). However, we predicted that we could use this knowledge of formicamycin BGC repression by ForJ to switch on the *for* BGC in a heterologous host. As such, the modified ReDirect protocol described in **Chapter 3.2** was used by Dr Neil Holmes (previous post-doctoral researcher in the Hutchings laboratory) to modify the PAC 215-G containing the entire formicamycin BGC and replace the coding region *forJ* with an apramycin resistance gene. This construct was then introduced into *S. coelicolor* M1146 which is optimised for the expression of secondary metabolite BGCs through the deletion of the BGCs encoding for actinorhodin, prodiginines, coelimycin and the calcium-dependent antibiotic (Gomez-Escribano and Bibb, 2014). The ex-conjugant colonies had a yellow appearance suggestive of fasamycin/formicamycin production that was not present in the no-PAC control or the strain containing the un-modified 215-G (**Figure 4.5**). To confirm these strains were producing fasamycins/formicamycins, cultures were grown on SFM agar and extracted for HPLC(UV) LCMS analysis, conducted by Hannah McDonald (JIC). This showed that only *S. coelicolor* M1146 containing the modified PAC with *forJ* deleted produced fasamycins and formicamycins, whereas the no-PAC control and the strain containing the unmodified PAC showed no evidence of any formicamycin congeners in the extracts. This demonstrates that deletion of *forJ* results in de-repression of the gene cluster in a heterologous host, further providing evidence for the role of ForJ as a major repressor of the formicamycin BGC.

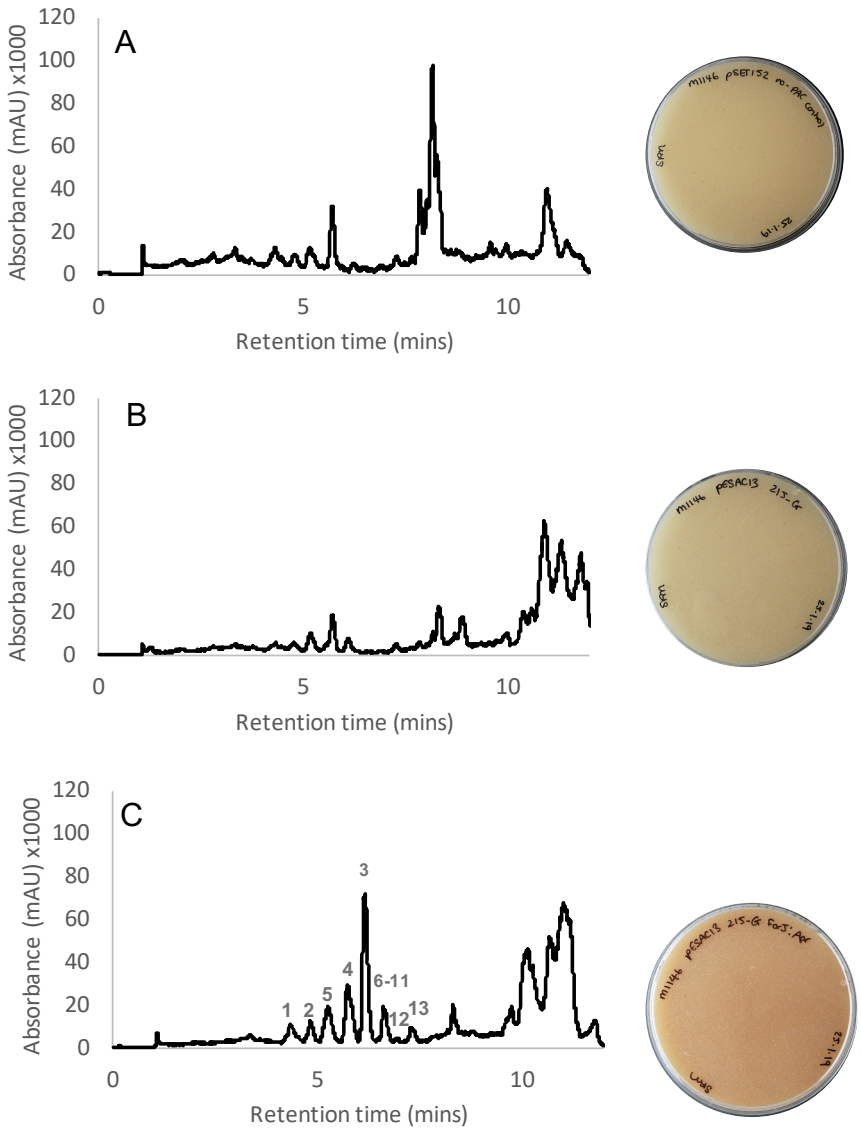


Figure 4.5: HPLC(UV) traces (250 nm) showing extracts of: (A) *S. coelicolor* M1146 no-PAC control (containing pSET152 in the same integrative site); (B) *S. coelicolor* M1146 215-G; (C) *S. coelicolor* M1146 215-G *forJ:apr*. Heterologous expression of the 215-G PAC containing the whole formicamycin BGC in *S. coelicolor* M1146 did not results in the production of any fasamycins or formicamycins. However, when the PAC is de-repressed by deleting *forJ*, transcription of the BGC is activated resulting in biosynthesis of the compounds in the heterologous host. This is evidenced by the yellow colouration of the media in the host containing the de-repressed PAC. HPLC(UV) LCMS conducted by Hannah McDonald, JIC.

4.2.2 The two-component system, ForGF

ForG is predicted to be a signal transduction histidine kinase and ForF is a DNA-binding response regulator with a conserved receiver domain. These genes are typical of a classical bacterial TCS. To determine the role of *forGF* in the regulation of formicamycin biosynthesis, the entire operon was deleted using CRISPR/Cas9 to generate the *S. formicae ΔforGF* strain. This mutant, along with a wild-type control and a complemented strain expressing *forGF in trans* under the control of its native promoter were grown under formicamycin producing conditions and the culture extracts analysed by HPLC(UV) LCMS by Dr Zhiwei Qin. Deletion of the TCS operon *forGF* completely abolished fasamycin and formicamycin biosynthesis, with no congeners present in the culture extracts. The complementation mutant *S. formicae ΔforGF: ΦBT1 forGF pforG* produces fasamycins and formicamycins although the levels of production are higher than the wild-type suggesting expression is higher when the genes are placed *in trans* (**Figure 4.6**).

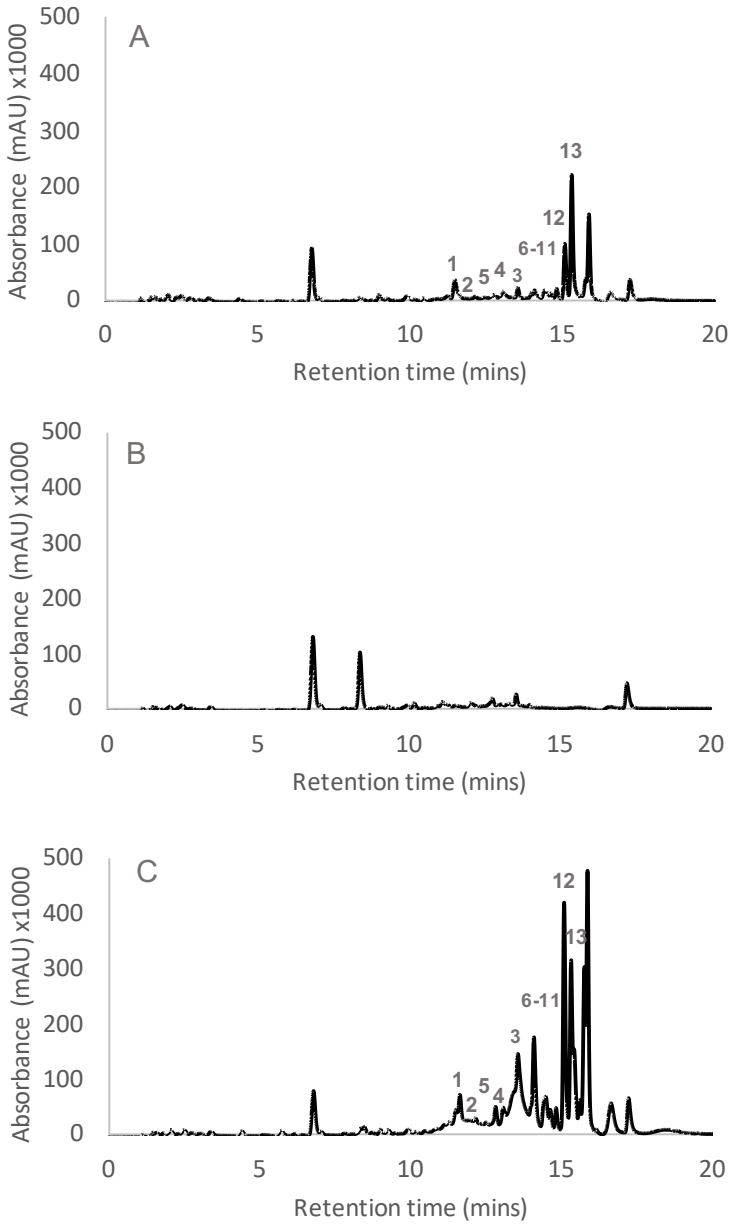


Figure 4.6: HPLC traces (UV 280 nm) showing extracts of: (A) *S. formicae* wild-type; (B) *S. formicae* Δ*for*GF; (C) *S. formicae* Δ*for*GF: φBT1 *for*GF *pforG*. Deletion of *forGF* abolished fasamycin and formicamycin biosynthesis. Complementation of *forGF in trans* restores production of all congeners to approximately wild-type levels. Three biological replicates of each strain were generated for analysis. HPLC(UV) LCMS analysis conducted by Dr Zhiwei Qin at the John Innes Centre.

These data strongly suggest that the TCS ForGF is required to activate the biosynthesis of the formicamycins. TCSs often control the expression of their target genes in response to extracellular signals. We hypothesise that ForG is a membrane associated sensor kinase that controls the activity of ForF in response to an extracellular signal. On receipt of the signal, ForF is phosphorylated and activated, causing DNA binding and the initiation of transcription of some or all of the

biosynthetic genes within the formicamycin pathway. At this time, the DNA target of ForF is unknown and therefore there is no way to know which genes within the formicamycin BGC are under the control of this two-component system. The activation signal of ForG is also unknown. Although many TCSs have been identified and characterised across *Streptomyces* genomes, very few have known activation signals and most are implicated in global regulation rather than cluster-specific regulation. Experimentally characterising the activation signal is challenging because these TCSs function as part of a diverse and highly complicated regulatory system that links the multi-cellular developmental stages of *Streptomyces* growth with both the primary and secondary metabolism (Hutchings *et al.*, 2004).

4.2.3 The MarR-family regulator, ForZ

ForZ is predicted to be another DNA-binding transcriptional regulator of the MarR-family. Domain analysis predicts the presence of a helix-turn-helix motif for DNA binding. BlastP analysis shows that ForZ is quite different to ForJ, the other MarR regulator encoded in the formicamycin BGC, with relatively low identity between the two sequences (44% coverage and 30.77% identity). Therefore, it is likely that they play distinct roles in the regulation of formicamycin biosynthesis. To determine the function of ForZ, gene deletion of *forZ* was conducted using CRISPR/Cas9 followed by complementation, *in trans*, with *forZ* under the control of its native promoter. *S. formicae ΔforZ* and *S. formicae ΔforZ: ΦBT1 forZ pforZ* were grown in parallel with a wild-type control under formicamycin producing conditions and extracts were analysed by HPLC(UV) LCMS by Dr Zhiwei Qin. All fasamycin and formicamycin congeners were present in the extracts of the *forZ* deletion mutant, however, they were at lower levels than in the wild-type extracts (**Figure 4.7**). This suggests that ForZ may be able to activate the transcription of some genes within the fasamycin/formicamycin biosynthesis pathway, however, it is not absolutely required for compound production.

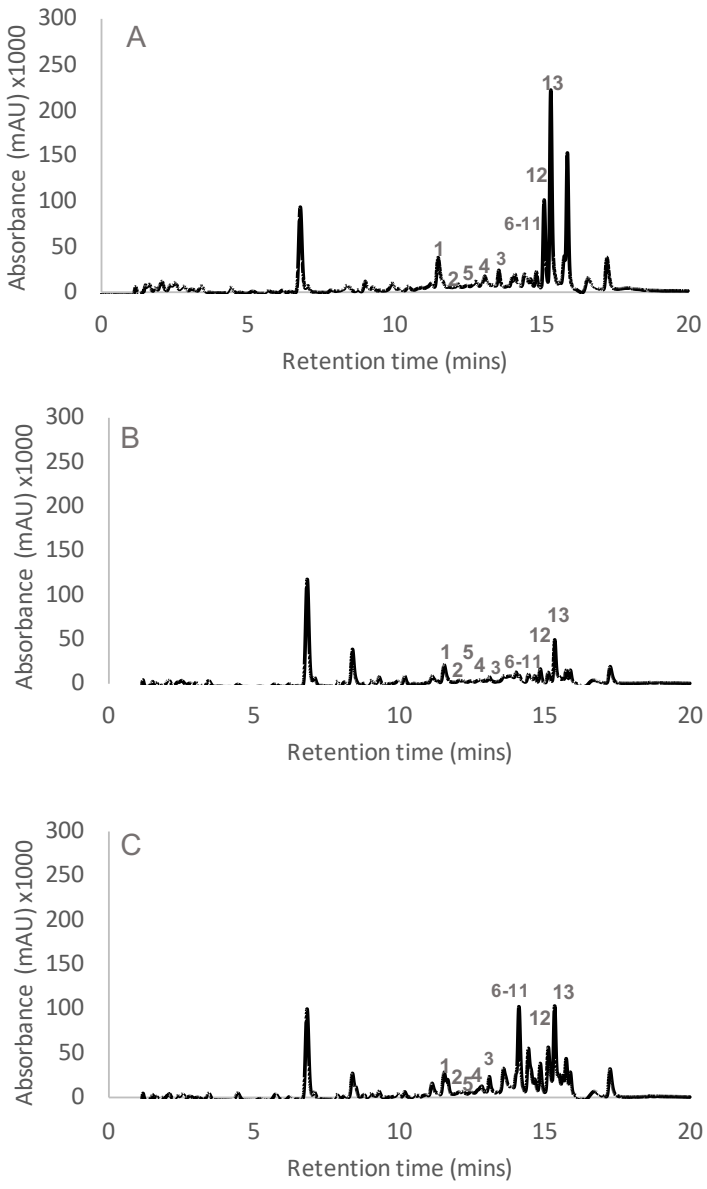


Figure 4.7: HPLC traces (UV 280 nm) showing extracts of: (A) *S. formicae* wild-type; (B) *S. formicae* Δ*forZ*; (C) *S. formicae* Δ*forZ*: φBT1 *forZ pforZ*. Deletion of *forZ* reduced the levels of fasamycin and formicamycin biosynthesis by around 50% compared to the wild-type. Complementation of *forZ in trans* restores production of all congeners to approximately wild-type levels. Three biological replicates of each strain were generated for analysis. HPLC(UV) LCMS analysis conducted by Dr Zhiwei Qin at the John Innes Centre.

Overall, this work shows that ForJ and ForZ, the MarR regulators, and the TCS ForGF, are all involved in the regulation of formicamycin biosynthesis. ForGF appears to be the main activator of transcription of the formicamycin BGC, whereas ForJ is the repressor. ForZ also appears to have a role in the activation of expression of genes within the formicamycin BGC.

4.3 Identifying targets of ForJ, ForF and ForZ within the formicamycin BGC

4.3.1 ChIP Sequencing

The previous section shows that *forJ, forGF* and *forZ* encode regulators of formicamycin biosynthesis in *S. formicae*. In order to learn more about the functions of the encoded CSRs, it was important to determine their binding sites to predict which transcripts were under the control of which regulators. Some regulators specifically control the transcription of a single gene within a BGC whereas others may bind to multiple DNA sites across the cluster to generally regulate the expression of secondary metabolite biosynthesis genes. Furthermore, some CSRs are known to bind to multiple BGCs across the genome. Chromatin Immunoprecipitation combined with sequencing (ChIP-Seq) is a widely employed technique used to identify the target sites of DNA-binding proteins (Bush *et al.*, 2016; Som *et al.*, 2017). Proteins of interest are crosslinked to DNA using formaldehyde, the DNA is then sheared into small fragments and the DNA-protein complex of interest is immunoprecipitated using a specific antibody against the regulator. The DNA that is 'pulled down' in this reaction can be sequenced and mapped back to the genome to identify the genetic targets of the specified DNA-binding protein. In order to identify the binding sites of ForJ, ForF and ForZ, mutant strains of *S. formicae* expressing 3xFlag-tagged versions of each protein were generated. As not much was known about each protein, a 15 amino-acid long linker was included between the coding sequence and the tag to increase the flexibility of the tag and reduce negative effects on the overall protein functionality. The linker-FLAG region was added to the C-terminus of each regulator and used to express 3x-Flag-tagged ForJ, ForF or ForZ *in trans* in each of the corresponding deletion mutants generated in the previous section. These tagged complementation strains were shown by LCMS analysis to have restored formicamycin biosynthesis, suggesting that the tags did not adversely affect the activities of the regulators (data not shown).

These strains were therefore used to perform ChIP against duplicate cultures of each strain after 2, 3 and 4 days under formicamycin producing conditions with wild-type

S. formicae as the negative control. These time points were chosen because the yellow colour indicative of fasamycin and formicamycin production appears around day 3 and it is expected that the regulatory and biosynthetic genes are expressed in advance of compound detection in order for synthesis to occur (Seipke, Patrick and Hutchings, 2014). Day 2 was the earliest time point that sufficient biomass could be harvested for ChIP-Seq. RT-PCR was used to confirm that all the regulatory genes were expressed at these time points (data not shown). After sampling for ChIP, small amounts of mycelium were also lysed and analysed by western-blot using an anti-FLAG antibody to confirm the presence of the flag-tagged proteins in the cell lysate (data not shown). The isolated DNA was quantified using Nanodrop UV-Vis and Qubit analysis before being sent to Genewiz (New Jersey) for sequencing using the Illumina HiSeq platform (EMBL-EBI Array Express accession number E-MTAB-8006). FASTQ files were processed and analysed by Dr Govind Chandra (JIC) to generate bedgraph files that could be visualized in Integrated Genome Browser.

As expected, some level of noise was present in all samples, with an average coverage depth of approximately 1500 background reads mapping across the reference genome. By comparing the reads present in the wild-type samples with those present in the ForJ, ForF and ForZ samples, the minimum limit of coverage was set at 2000 reads in order to view significantly enriched peaks. Significance was confirmed by analysis conducted by Dr Govind Chandra where every 25 nucleotides, the area of enrichment over 4000 nucleotides was divided by the reads of the central 50 nucleotides. These resulting enrichment values were then analysed according to a negative binomial distribution and only enrichment values with statistical significance of $p < 0.05$ were included. This ensures that peaks shown are significant compared to the surrounding background reads as well as compared to the wild-type. In general, only a few significant peaks were present in the day 2 samples, probably due to *S. formicae* still being in an early growth stage and therefore the activity of secondary metabolite BGCs is likely to be low at this time point. However, by day 4, there were multiple significant peaks present compared to the wild-type background, mostly present within the formicamycin BGC. All peaks that were present in the day 2 sample were still present in the day 4 sample, showing that even

these small peaks were significant and the limit of background reads set was appropriate for this dataset.

3xFlag-ForJ binds at multiple locations across the formicamycin BGC. Significant peaks were called upstream of the following transcripts; *forMLK, forHI, forTSRABCDE* and *forUVWXY* (**Figure 4.8**). These data, combined with the HPLC(UV) LCMS presented above, suggests that ForJ might repress the expression of most of the major transcripts within the formicamycin BGC by binding to these promoters and preventing recognition of the promoter sequence by the RNAP. By binding to *pforT/U,* ForJ is likely preventing transcription of the *forTSRABCDE* transcript, and these genes encode the core biosynthetic machinery required for formicamycin biosynthesis (see later chapters). Therefore, the repression of this transcript is important for the repression of formicamycin biosynthesis. Interestingly, ForJ also appears to bind within the coding sequences of itself and *forE*, which encodes the ACC biotin carboxylase. In addition to preventing the RNAP initiating transcription by binding to promoters, some DNA binding proteins prevent transcription via a roadblock mechanism where they bind within the coding region of a gene (Roy *et al.*, 2016). Thus, we predict that ForJ is auto-repressing transcription of *forJ* and *forE* through a roadblock mechanism. In addition, ForJ binds upstream of two genes elsewhere in the genome; KY5_3182, which encodes a putative MoxR-type ATPase and KY5_5812, which encodes a hypothetical protein. The significance of these binding sites is currently unknown.

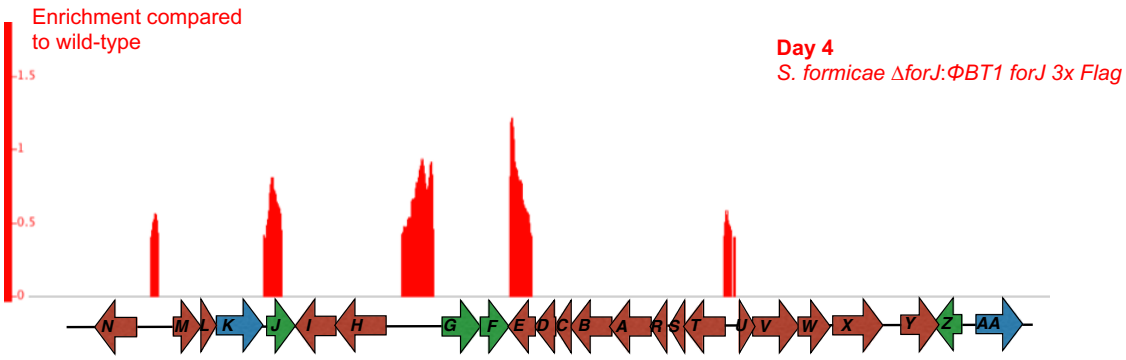


Figure 4.8: ChIP-seq results from *S. formicae ΔforJ:ΦBT1 forJ 3x Flag* strain at day 4. ForJ binds at five different positions within the formicamycin BGC, suggesting it is a major regulator of formicamycin biosynthesis. Elsewhere in the genome there were only two peaks present, suggesting ForJ may specifically be a regulator of formicamycin biosynthesis. Peaks represent enrichment relative to the wild-type background.

It should be noted that generally the enrichment values for the peaks in the 3xFlag-ForJ dataset were lower than for the other regulators tested. This is predicted to be because the ForJ deletion mutant grows more slowly compared to the wild-type and complementation with the Flag-tagged copy of ForJ was conducted using *pforM,* before we knew *forJ* had its own promoter, meaning the biomass obtained for this strain was lower compared to the other mutants and the wild-type control. This was reflected in the fact that the total amount of DNA obtained from the 3xFLAG-ForJ samples was also lower. This could have led to less coverage in the sequencing overall, bringing down the relative enrichment values compared to the wild-type. The peaks are still statistically significant and reflective of real protein-DNA interactions.

Analysis of the 3xFlag-ForF dataset revealed there was only a single binding site within the formicamycin BGC which was between the divergent *forHI* and *forGF* transcripts (**Figure 4.9**). As the LCMS data above suggests that ForGF is the activator of formicamycin biosynthesis, it is possible that initial binding of the response regulator to this promotor region allows autoregulation and transcription of *forHI*, forming the initial malonyl-CoA starter unit, which presumably initiates biosynthesis of the formicamycins.

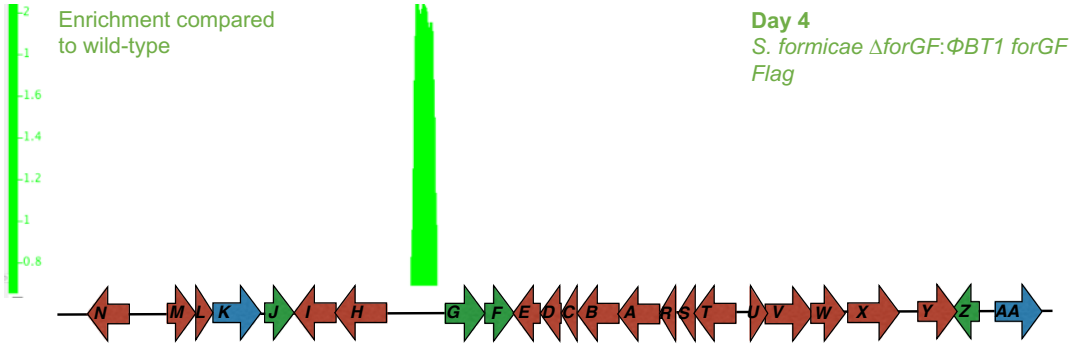


Figure 4.9: ChIP-seq results from *S. formicae ΔforGF:ΦBT1 forGF 3x Flag* at day 4. ForF appears to only bind to a single divergent promoter within the formicamycin BGC. Peaks represent enrichment relative to the wild-type background.

ForF also appears to bind upstream of KY5_0375. BlastP analysis suggests this gene may encode for an NLPc/P60 family protein. This protein family consists of cell-wall peptidases that are widespread across multiple bacterial lineages. Those that have been characterised in *Bacillus subtilis* have been shown to hydrolyse the D-glutamyl-meso-diaminopimelate linkage present in cell wall peptides (Anantharaman and Aravind, 2003). Perhaps more interestingly to this project, this gene sits within another predicted BGC from the AntiSMASH analysis of the *S. formicae* genome, predicted to encode for a RiPP due to the presence of two radical SAM enzyme encoding genes (Holmes *et al.*, 2018). This suggests ForGF may be responsible for regulating the production of more than one antimicrobial in *S. formicae*, a hypothesis which is investigated later in this thesis (**Chapter 7.2**).

Analysis of the 3xFlag-ForZ dataset revealed the presence of just one peak in the whole genome of *S. formicae*. ForZ binds to the divergent promoter region between *forZ* and *forAA* (**Figure 4.10**).

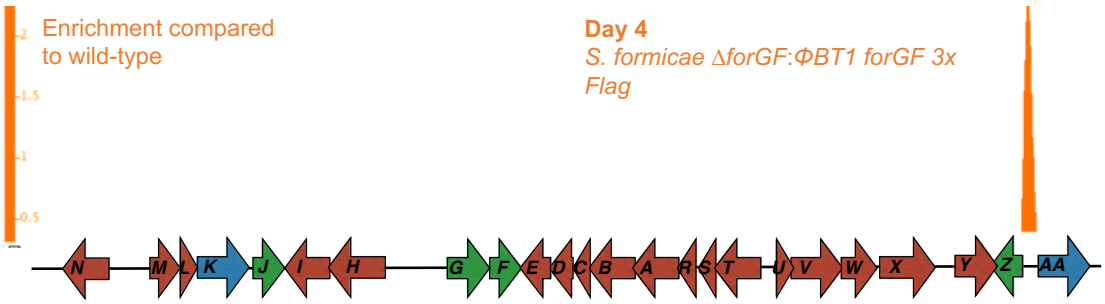


Figure 4.10: ChIP-seq results from *S. formicae ΔforZ*:*ΦBT1 forZ 3x Flag* at day 4. ForZ binds to a single divergent promoter within the formicamycin BGC. Peaks represent enrichment relative to the wild-type background.

MarR regulators are generally involved in regulating the activity of genes involved in stress responses and the degradation or export of toxic molecules such as phenolic compounds and antibiotics from the cell. The conventional mode of regulation via a MarR protein occurs via binding to an intergenic region between two divergent genes, one encoding for the MarR regulator itself and the other encoding for the gene under its regulation (Perera and Grove, 2010). ForZ is therefore predicted to autoregulate and regulate the expression of the putative multidrug resistance transporter gene *forAA*, which may be responsible for export of formicamycins. If this is the case, it is currently unclear why deletion of ForZ in the previous section resulted in a decrease in formicamycin production. It is possible that there is some sort of feedback loop that limits the levels of expression of the biosynthetic genes until transcription of the export mechanism is activated to prevent compounds accumulating to toxic levels.

4.3.2 Motif analysis of DNA binding sites identified by ChIP-seq

The DNA-binding sites of each CSR identified using ChIP-seq are only to approximately 100-300 base pair resolution (**Figure 4.11**). Normally, transcriptional regulators specifically recognize short inverted repeat sequences within their target DNA, usually between 16 and 20 bp long, that are either palindromic or inverted repeats. It is possible to use bioinformatic approaches to search for motif sequences within the identified target promoters for example by using motif-based sequence analysis tools such as MEME (Bailey *et al.*, 2015). As only a single binding site was identified for ForZ, MEME could not be used to search for motifs with any statistical

significance. Similarly, ForF only had two binding sites in the ChIP experiment, so any binding sites identified by MEME would likely not be significant. Furthermore, response regulators are hugely diverse and proteins within the family are able to bind a variety of different motif sequences, making it hard to define search criteria in this type of analysis. However, there were five binding sites for ForJ within the formicamycin BGC and MarR-family regulators are known to frequently bind inverse repeat sequences, therefore MEME analysis is more likely to identify significant motifs within the target promoters.

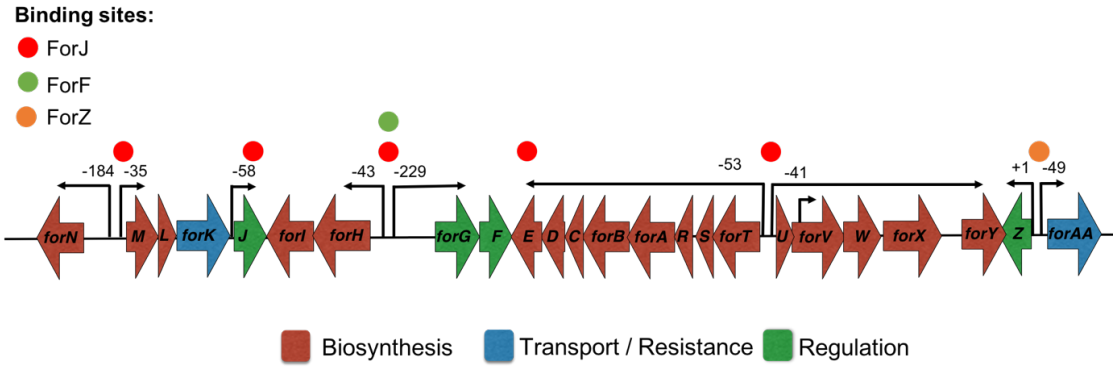


Figure 4.11: TSS (indicated by arrows) were identified using cappable RNA sequencing. Approximate binding sites for each cluster situated regulator were determined by ChIP-sequencing and are represented by the circles. These binding sites are only defined to approximately 100-300 bp resolution.

The DNA regions identified as ForJ binding sites were entered into the MEME tool to look for motifs that were present in all sequences and to preferentially search for palindromic motifs. The results gave three predicted motifs, the most significant of which is shown (**Figure 4.12**). The locations of all the identified motifs were examined in comparison with the TSSs identified from the cappable RNA sequencing experiment. It might be expected that the ForJ repressor would bind on or very near to the TSS to prevent binding of the RNA polymerase. The motif shown in **Figure 4.12** appears exactly across the TSS of *pforM* and *pforT* as well as being within just 21 base pairs of *pforU*. The motif also appears almost exactly half way through the intergenic region between *forH* and *forG*, sitting between the two TSSs of this divergent promoter. The motif is also present at the 3' end of the coding region of both *forJ* and at the edge of the peak identified in the coding region of *forE*, right at the end of the overlapping coding region with *forD* (**Table 4.2**). We therefore propose that

these 14 base pair imperfect inverted repeat sequences are the DNA binding sites of ForJ. The locations of the other two predicted motifs identified by MEME were not as striking in terms of proximity to the transcription start sites, therefore, due to their lower significance scores, it was concluded that these did not represent likely binding sites for ForJ. Many CSRs found in actinomycete genomes been shown to recognise heptameric repeat sequences within promoter regions and the motif presented here is a non-perfect inverted heptameric repeat sequence, further implying that it may represent the DNA binding site of ForJ.

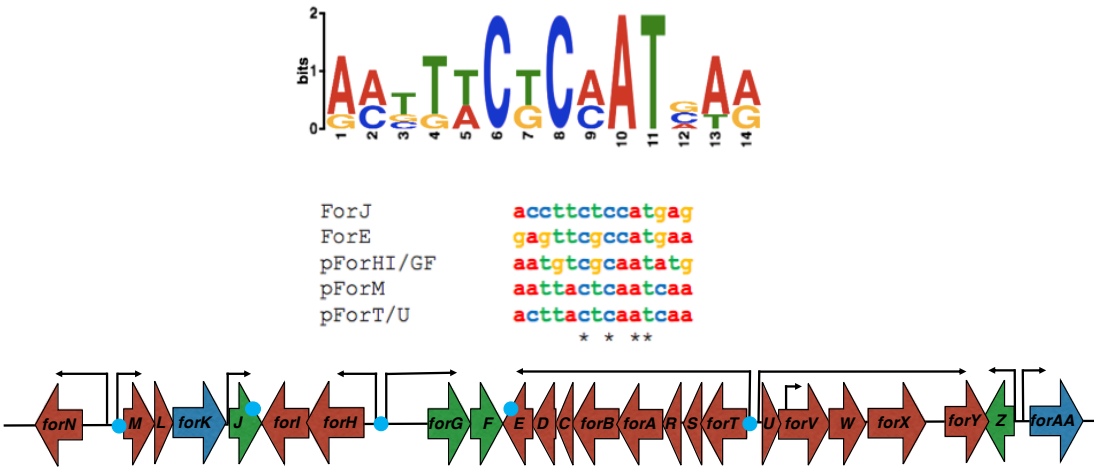


Figure 4.12: MEME analysis identified a 14 base pair inverted repeat sequence that is almost palindromic within the binding sites identified in the ChIP-seq data (E value 2.4e+000). The motif, identified by a light blue dot, is located at or near the transcription start site within each of the target promoters of ForJ and within all the binding sites identified in the ChIP experiment. This motif likely represents the DNA binding site of ForJ.

Table 4.2: Nucleotide positions of the promoter regions or coding regions within the *for* BGC that were identified as targets of ForJ during ChIP-Sequencing along with the positions of the nearest TSS and the possible ForJ binding site identified by MEME. In all the identified promoter region targets of ForJ, the motif identified by MEME sits across or very close to the TSS identified by capable RNA sequencing.

	Promoter or coding region position		**Nearest TSS**	**Site of ForJ binding site**	
	Start	**End**		**Start**	**End**
pforM	7756912	7757157	7757123	7757115	7757130
forJ	7759841	7760290	7759783	7760248	7760262
pforHI/pforGF	7763485	7763994	7763527 7763765	7763714	7763728
forE	7765756	7766271	7771343	7766385	7766399
pforT/U	7771291	7771410	7771343 7771370	7771335	7771349

4.3.3 Mechanistic insights into regulation of the formicamycin BGC

This work has begun to characterise the mechanisms of regulation of formicamycin biosynthesis by the MarR-family regulators ForJ and ForZ, and the TCS, ForGF. We predict that ForJ represses biosynthesis of the formicamycins by binding to multiple regions across the *for* BGC, through the recognition of a short 14 bp palindromic repeat sequence. In this way, as well as autoregulating, ForJ is likely able to block the transcription of much of the core biosynthetic machinery within the formicamycin BGC by a combination of preventing the recognition of promoter regions by the RNAP or preventing transcription through roadblock. The TCS ForGF likely controls the transcription of *forHI*, the genes encoding for production of the starter unit for polyketide biosynthesis, i.e. the first stage of formicamycin biosynthesis, as well as autoregulating. On receipt of an appropriate signal, the histidine kinase ForG presumably activates its response regulator ForF, which can then bind to the divergent promoter *pforG/H* allowing expression of the *forHI* transcript and autoregulation of *forGF*.

The majority of classical response regulators belong to the OmpR/PhoB subfamily and consist of a typical conserved α/β domain referred to as the receiver domain and a C-terminal effector domain that is responsible for DNA binding. The response regulator ForF aligns most closely with the vancomycin-resistance associated response regulator VraR from *Staphylococcus aureus,* with 30.14% identity and 95% coverage. VraR regulates the cells response to antibiotic-related cell wall stress in conjunction with its partner sensor histidine kinase. Experimental characterization of VraR has shown that upon phosphorylation and dimerization, the response regulator binds DNA in a sequence specific manner to control transcription of the genes in its regulon (Liu *et al.*, 2009). Modelling of ForF in Swiss-model shows the receiver domain likely consists of the highly conserved five-stranded parallel β sheets surrounded by five amphipathic helices found in most OmpR/PhoB α/β domains. Characterisation of other OmpR/PhoB family response regulators shows that the receiver domain contains conserved aspartate residues for phosphorylation by the sensor kinase. Phosphorylation of the receiver domain causes small conformational changes in the structure of the response regulator causing enhanced DNA binding affinity in the C-terminal effector domain (Gao, Mack and Stock, 2007). This is also highly likely to be the method of transcriptional activation by ForF. The C-terminal domain of ForF consists of a typical winged helix-turn-helix structure for optimal DNA binding and transcriptional control (**Figure 4.13**).

Figure 4.13: Modelling of the response regulator ForF shows that it forms a homodimer for DNA binding and activation of transcription of the formicamycin BGC. Like many classical response regulators, it consists of a conserved N-terminal α/β receiver domain that can be phosphorylated and a C-terminal effector domain that likely binds to DNA.

Modelling suggests that like most MarR regulators, ForJ and ForZ are also winged helix-turn-helix DNA binding proteins that form triangular homodimers (**Figure 4.14**). ForJ and ForZ share low sequence homology, suggesting that they each act via different mechanisms. ForJ is predicted to be most similar to the MarR transcriptional regulator YusO (26.81% identity, 90% coverage) which functions as a homodimer to repress expression of itself and the divergent multidrug resistance protein YusP in *Bacillus subtilis* (Kim *et al.*, 2009). ForZ aligns most closely to the transcriptional regulator RdhR-CbdbA1625 (16.08% identity, 83% coverage) that represses transcription of reductive dehalogenase homologues in the obligate organohalide-respiring *Dehalococcoides mccartyi* in response to dichlorophenol ligands (Wagner *et al.*, 2013). It has been shown for other MarR regulators that the recognition helix within the DNA binding pocket binds to the major groove of DNA while the wing region is essential for stabilising the binding by interacting with the adjacent minor grove. This mechanism reduces distortions caused to the DNA molecule (Perera and Grove, 2010).

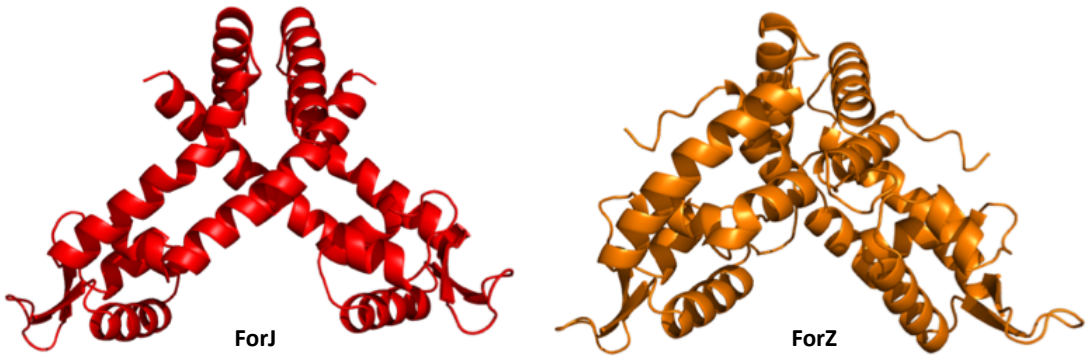


Figure 4.14: Modelling of ForJ (red) and ForZ (orange) in Swiss model and PyMol shows that both proteins likely function as homodimers that form a triangular shape with a DNA-binding pocket between each monomer unit.

MarR-family regulators have ligand-binding capabilities in addition to DNA-binding and, unlike other cluster situated regulators, they bind their ligand and the DNA in the same domain. Ligand binding changes the confirmation or flexibility of the MarR transcription factor and alters its DNA-binding activity, resulting in a change in target gene expression levels (Grove, 2017). MarR regulators are often encoded in pathways that synthesise aromatic compounds and the ligands for the regulators are often substrates of the biosynthetic enzymes under the control of the MarR regulator (Perera and Grove, 2010). It is possible that when ForF activates transcription of *forHI,* the malonyl-CoA starter unit binds to ForJ, causing it to de-repress transcription of the other biosynthetic genes in the *for* BGC, resulting in the production of formicamycins. This hypothesised mechanism means that transcription of the *for* BGC remains repressed by ForJ until there are enough starter units available for formicamycin biosynthesis.

In addition, many MarR regulators are known to be involved in the regulation of expression of host resistance mechanisms in response to accumulating antibiotics within the cell. For example, OtrR is a MarR-family regulator encoded in the oxytetracycline biosynthetic pathway in *Streptomyces rimosus* that has been shown to repress the expression of the divergent gene *otrB*, which encodes for a membrane associated MFS family transporter. The activity of OtrR is sensitive to the presence of oxytetracycline and other biosynthetic intermediates from the pathway which induce the expression of *otrB* (Pickens and Tang, 2010; Mak, Xu and Nodwell, 2014).

The ForAA protein is thought to be responsible for exporting formicamycins out of the cell and therefore contributing to the host-resistance. Generally, the expression of host resistance mechanisms is tightly controlled to ensure that it occurs either prior to compound biosynthesis or at the same time. Without this tight control, antimicrobial natural products would accumulate inside the cell causing self-toxicity. The expression of these host resistance mechanisms is often induced by the antibiotic products themselves or by biosynthetic intermediates produced by the pathway. In this way, the host can respond to increasing compound titres and prevent overwhelming its own capacity for resistance (Mak, Xu and Nodwell, 2014). Given this information, we predict that accumulating formicamycins or other biosynthetic intermediates from the pathway bind to ForZ, causing it to de-repress the expression of *forZ* and the divergent *forAA* transporter gene to ensure that the compounds are removed from the cell before they can cause toxicity to the host.

However, the data presented earlier in this chapter indicate that the role of ForZ goes beyond simply the control of *forZ/forAA* expression, as deletion of *forZ* reduced the production of formicamycins by around 50% (**Chapter 4.2.3**). This suggests that ForZ can also activate formicamycin biosynthesis or somehow increase levels of compound production. Recently, the cluster-situated MarR regulator SAV4189 from the avermectin BGC in *S. avermitilis* was shown to repress the transcription of itself and the adjacent transporter gene as well as also indirectly activating transcription of other avermectin biosynthesis genes by changing the expression levels of *aveR*, the cluster-situated activator. In a SAV4189 deletion mutant, levels of the *aveR* activator were lower and avermectin yields were reduced by approximately five fold (Guo *et al.*, 2018). The amino acid sequence of ForZ aligns to SAV4189 with 30.91% identity and 55% coverage. It could be that as well as autoregulating and controlling the transcription of *forAA*, the putative resistance transporter, ForZ is able to interact with another CSR from the formicamycin BGC to further increase the transcription of other biosynthetic genes within the *for* BGC, increasing the titres of formicamycins produced. This mechanism ensures that high-titres of formicamycins are only produced once expression of the host-resistance mechanism is switched on, to avoid

self-toxicity. It also explains why, in a *forZ* deletion mutant, where expression of *forAA* remains repressed, formicamycin production levels are lower (**Figure 4.15**).

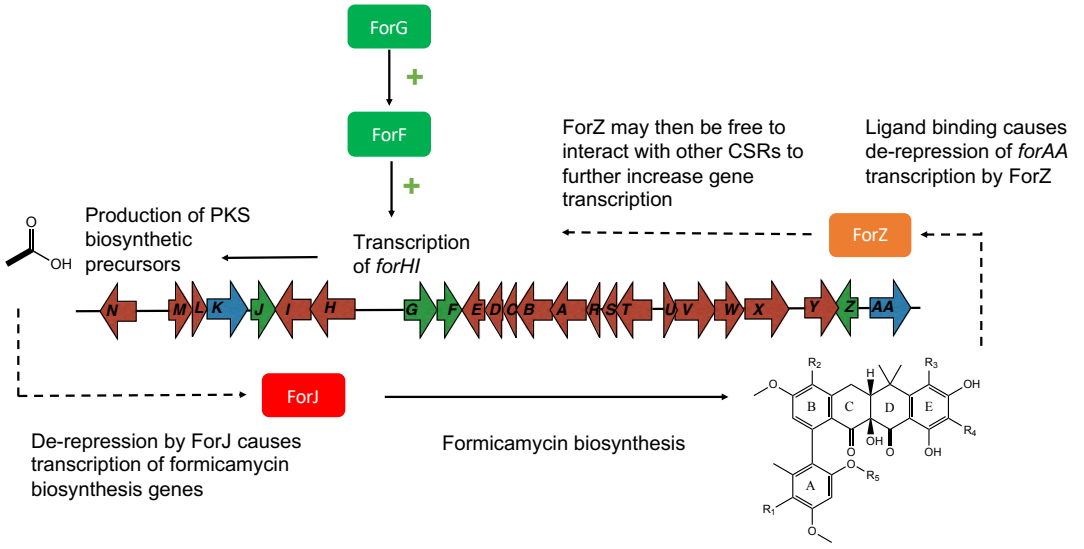


Figure 4.15: Regulation of the formicamycin BGC is controlled by the combined actions of the MarR regulators, ForJ and ForZ, as well as the TCS, ForGF. Transcription of *forHI* is predicted to be activated by the TCS, ForGF. The *forHI* transcript encodes the genes responsible for the biosynthesis of the malonyl-CoA starter unit. Transcription of the entire *for* BGC is thought to be repressed by ForJ until something, possibly the accumulation of malonyl-CoA, causes it to fall off the DNA and allow transcription of the biosynthetic genes. Once transcription of formicamycin biosynthesis genes is activated through this de-repression by ForJ, formicamycins accumulate. The binding of formicamycins or pathway intermediates to ForZ is predicted to de-repress *pforZ/AA*, inducing transcription of the host-resistance transport gene, *forAA*. ForZ may also be able to indirectly further increase transcription of the *for* BGC perhaps through interactions with another CSR. In this figure, solid arrows represent interactions that have been experimentally confirmed in this chapter and dashed lines represent predictions made based on the data available at this time.

4.4 Understanding the mechanisms of host resistance to formicamycins in *S. formicae*

Most secondary metabolite BGCs that encode for antibiotics also encode for at least one resistance mechanism to enable the host to produce the compounds without causing toxicity to itself. These host resistance mechanisms can be extremely diverse, ranging from target modification, antibiotic inactivation or transport via efflux pumps (Mak, Xu and Nodwell, 2014). Analysis of the formicamycin BGC presented previously (**Chapter 3.5**) did not identify any candidates for host-resistance within the cluster or near the cluster edge. However, there are two putative transport genes encoded within the formicamycin BGC. ForK is predicted to

be a cation/H+ antiporter and ForAA is a putative major facilitator superfamily (MFS) drug resistance transporter. Data described above suggest that ForAA may be involved in the export of formicamycins out of the cell under the control of ForZ. So far, the role of ForK is unknown. In order to determine the roles of the transporter genes in the formicamycin BGC, CRISPR/Cas9-mediated deletions of each gene were made in *S. formicae*. The initial hypothesis was that one of the transporters would be essential and that deletion would be lethal as a result of self-toxicity via formicamycin accumulation. Deletion mutants were therefore generated in both the wild-type *S. formicae*, which produces formicamycin congeners when grown under standard laboratory conditions and *S. formicae* Δ*forGF*, which has been shown previously to not produce any fasamycin or formicamycin congeners. We predicted that deletion of the host-resistance mechanism would be possible in *S. formicae* Δ*forGF* but not in the wild-type. Conversely to our predictions, confirmed replicates of *S. formicae* Δ*forK* and *S. formicae* Δ*forAA* were generated in both the wild-type and Δ*forGF* background. When grown under formicamycin producing conditions alongside a wild-type control, both the *S. formicae* Δ*forK* and *S. formicae* Δ*forAA* mutants appear to grow more weakly than the wild-type strain, although the phenotype is not extreme. This suggests that neither transporter on its own is essential for export of the fasamycins/formicamycins and perhaps that the two transporters work synergistically. To test this hypothesis, the pCRISPomyces-2 system was used to generate a mutant strain lacking both transporters, *S. formicae* Δ*forAA* Δ*forK.* This strain had quite a severe phenotype, much like *S. formicae* Δ*forJ* strain described previously; the mycelia were extremely yellow and the strain did not sporulate even after extended incubation under conditions that normally cause sporulation in wild-type *S. formicae* (**Figure 4.16**).

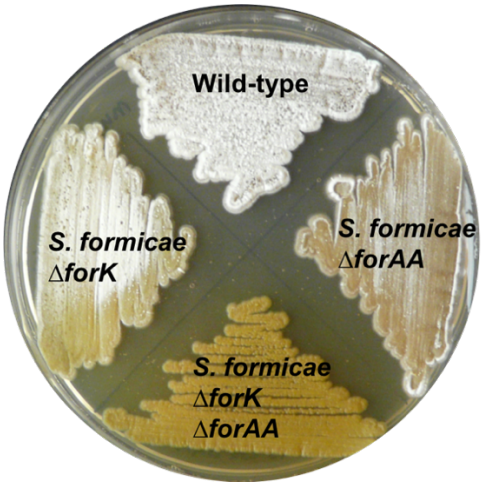


Figure 4.16: Each putative transport gene encoded in the formicamycin BGC (*forK* and *forAA*) was deleted using CRISPR/Cas9 both individually and in combination in the wild-type background. When grown under formicamycin producing conditions, the single deletion mutants (*S. formicae ΔforK* and *S. formicae ΔforAA*) do not show a severe phenotype compared to the wild-type strain, however, the double mutant (*S. formicae ΔforK ΔforAA*) grows poorly and does not sporulate, suggesting toxic compounds are accumulating within the cell.

Although the strain appears sick, deletion of both transporter genes is not lethal in *S. formicae*. The most likely explanation for this is that similar transporters encoded elsewhere in the genome can partially compensate for their loss and prevent lethality. Indeed, BlastP analysis of the *S. formicae* genome identifies multiple gene products with significant homology to both *forK* and *forAA* that might have the potential to function via a similar mechanism and transport these molecules out of the cell. There are 158 putative MFS family drug-resistance transporters encoded in the *S. formicae* genome, 64 of which show significant similarity to ForAA and 4 other cation/H+ antiporters that show homology to ForK (homology in this case is measured as >25% identity and/or >80% coverage to the query sequence). It is highly likely that one or more of these homologous transporters encoded elsewhere in the genome can export low levels of formicamycins out of the cell. In addition, *S. formicae* likely possesses its own host-resistance mechanism, for example additional copies of DNA gyrase as discussed in **chapter 3.5,** and can therefore tolerate low levels of compound accumulation. However, the obvious phenotype of the double transporter mutant, *S. formicae ΔforK ΔforAA,* suggests that the strain is struggling to cope with the accumulation of these compounds. This suggests that the other transporters are much less efficient at removing formicamycins from the cell and

shows that the presence of *forAA* and *forK* together is very important for host resistance. In the future, in order to determine which compounds from the *for* BGC are exported by each of the transporters, the transporter deletion strains will be grown on solid agar with cellophane discs so that the biomass can be removed from the growth media after fermentation to isolate the intracellular compounds from the ones that have been exported. These samples can then be quantitatively analysed by HPLC(UV)/LCMS to understand which compounds are exported via each transporter and at what levels. We also have complementation strains for the single deletion mutants, *S. formicae* Δ*forK: ΦC31 forK pforM* and *S. formicae* Δ*forAA: ΦC31 forAA pforAA* with each gene placed *in trans* under the control of its native promoter. Phenotypically, these strains appear and grow like the wild-type strain (data not shown). The complementation strains will also be analysed by HPLC(UV)LCMS alongside the mutants as described above.

MFS transporters like ForAA are widespread in nature, ranging in functions from resistance-conferring efflux pumps in bacteria to glucose transporters in large mammals. Most MFS transporters consist of 12 or 14 trans-membrane helices and contain two domains that together form a channel across the cell membrane. They are usually specifically evolved for the transport of structurally related substrates. Some drug-efflux systems rely on energy from the hydrolysis of ATP to move large molecules across the membrane, however, many, including some of the MFS transporters, are driven by the proton motive force (PMF) generated by the electrochemical gradient across the membrane (Paulsen, Brown and Skurray, 1996). These 'secondary-active' multidrug transporters function as antiporters that couple efflux of their substrate(s) to the influx of protons or sodium ions down the electrochemical gradient. The PMF provides the energy for the transporter to switch between inward and outward oriented states via a series of consecutive conformational changes between the two halves of the protein (Du *et al.*, 2015). ForK is predicted to be a putative cation/H+ antiporter that is responsible for moving protons across the cell membrane. We therefore hypothesise that ForK functions to increase the electrochemical gradient across the cell membrane due to increased ion exchange, providing more energy for efflux via ForAA and more efficient transport.

If this hypothesis is true, *S. formicae* Δ*forK* is able to survive because ForAA will export formicamycins, albeit to a lower efficiency than in the wild-type strain. In the *S. formicae* Δ*forAA* strain, we might expect levels of overall transcription of the *for* BGC to be lower through the indirect actions of ForZ with another CSR when the major host-resistance gene is removed or inactivated, meaning that compounds do not accumulate intracellularly to toxic levels. To test this in the future, in addition to HPLC(UV) and LCMS analysis, transcriptomic analysis such as qRT-PCR or RNA Sequencing of the regulator and transporter mutants could be conducted to compare the expression levels of the transcripts within the *for* BGC in the different backgrounds.

So far, this work shows that both transporters work synergistically to export fasamycins and formicamycins out of the cell. As well as providing more information about the risk of resistance to these compounds arising in pathogenic bacteria if they are ever used clinically, this knowledge of host resistance to formicamycins in *S. formicae* can be used to generate mutant strains with increased biosynthetic capabilities. Earlier in this chapter, we showed that deletion of *forJ* in *S. formicae* resulted in higher titres of fasamycins/formicamycins, however, it also resulted in severe phenotypic effects and the mutant grew poorly compared to the wild-type *S. formicae* (**Figure 4.4**). We hypothesised that over-production of fasamycins and formicamycins in *S. formicae* Δ*forJ* overwhelms the natural resistance mechanisms and transporters, resulting in a toxic accumulation of the compounds. We predicted we could generate strains that would both over-produce fasamycins/formicamycins and grow rapidly to a high biomass by increasing the expression of the host-resistance mechanism in *S. formicae* Δ*forJ*. Initially, strains of *S. formicae* Δ*forJ* overexpressing either ForK or ForAA *in trans* under the control of the constitutive promoter p*ermE** were generated as they were predicted to give the highest level of host resistance via compound efflux. However, ForAA and ForK are membrane proteins and their overexpression caused the resulting strains to grow extremely poorly, even compared to *S. formicae* Δ*forJ*, probably due to disruptions to the cell membrane (data not shown). Therefore, the constructs generated previously for

complementation of the deletion mutants with each transporter expressed under the control of its native promoter were conjugated into S. *formicae* Δ*forJ* to generate strains with an extra copy of either *forK* or *forAA*. The resulting strains showed greatly increased growth and sporulation compared to *S. formicae* Δ*forJ*. The extra copy of *forAA* seemed to increase both sporulation and biomass to a greater degree than the additional copy of *forK,* resulting in the strain *S. formicae ΔforJ*: ΦBT1 f*orAA pforAA* that seems to grow and develop the same as the wild-type (**Figure 4.17**). This further suggests that ForAA might be the major transporter of formicamycin congeners, whereas ForK plays a role in increasing the efficiency of transport by ForAA.

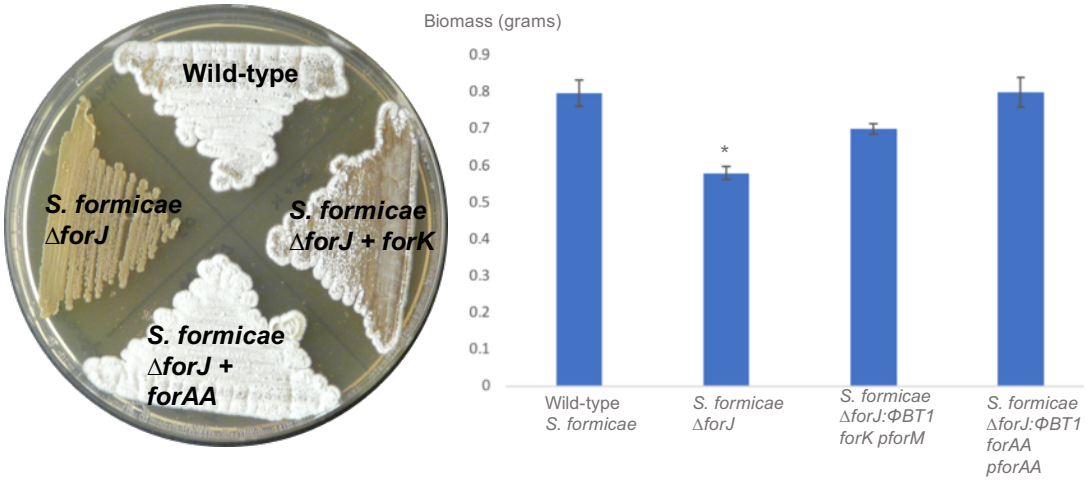


Figure 4.17: Previous work showed that *S. formicae ΔforJ* overproduces fasamycin and formicamycins but the strain is sick, presumably due to accumulation of toxic compounds inside the cell, resulting in significantly less biomass in the mutant compared to the wild-type (two-tailed T-test, p = 0.03). To overcome this, extra copies of the transporters *fork* and *forAA* were expressed *in trans* in the *forJ* mutant to increase the capacity of the host-resistance mechanism. The extra copy of *forK* resulted in a moderate increase in biomass and sporulation, however, the addition of an extra copy of *forAA* fully complemented the growth and development to that of the wild-type strain.

These strains will also be fully and quantitatively analysed by HPLC/LCMS in the future to characterise their levels of antibiotic production and export, however, it is assumed that they represent strains that overproduce fasamycins and formicamycins without the detrimental effect of self-toxicity on the host. These strains of *S. formicae* may be useful in future for the isolation of high titres of compounds for use in *in vitro* studies. It would also be interesting to determine whether these strains are capable of overproducing compounds during liquid culture as the majority of

industrial antibiotic production is performed in flasks and bioreactors (Manteca and Yagüe, 2018). For these clinically relevant antibiotics to be produced at an industrial scale, it will be important to continue improving strains of *S. formicae* for this purpose in the future.

4.5 Discussion

During this chapter, we have begun to understand the transcriptional organisation and regulatory control of the formicamycin BGC. Cappable RNA sequencing allowed us to characterise the transcriptional organisation of the BGC and enabled the identification of the promoters and transcripts required for expression of the formicamycin biosynthesis genes. This showed that many of the genes within the formicamycin BGC are translationally coupled, suggesting their functions are tightly linked. The roles of the biosynthetic enzymes encoded on these transcripts is fully investigated in the following chapters. The identification of the promoters within the formicamycin BGC is also used in genetic experiments characterising the roles of the biosynthetic genes within the cluster as it allows complementation constructs to be generated using the native promoter. We have also identified the CSRs that bind the gene promoters required for formicamycin biosynthesis and predicted which transcripts are likely to be under their control using ChIP-sequencing. This work begins to provide a detailed understanding of the complex systems that regulate biosynthesis of the formicamycins and host-resistance to this important new group of antibiotics.

Formicamycin biosynthesis appears to be activated by the ForGF TCS. ChIP-seq of a strain producing a 3xFlag-tagged ForF shows that it binds upstream of the *forHI* transcript, which encodes the proteins responsible for the first step of formicamycin biosynthesis, the generation of the malonyl-CoA starter unit. The majority of characterised TCSs are from the genomes of model actinomycetes such as *S. coelicolor* and *S. venezuelae*, due to the availability of genome data and protocols for genetic manipulation in these organisms (Rodríguez *et al.*, 2013). Of the TCSs studied in these organisms, many have been suggested to play a role in the higher-level regulation of both antibiotic biosynthesis and morphological differentiation, with

fewer cluster-situated TCSs having been characterised so far. Loss of ForGF does not appear to result in a phenotype in *S. formicae*, at least under the conditions tested during this work, therefore it is unlikely that ForGF plays a role in morphological differentiation in addition to its function as an activator of formicamycin biosynthesis. Some CSRs are also able to alter the expression of BGCs elsewhere in the genome through cross-cluster regulation. For example, JadR1 is an OmpR-type regulator that sits within the jadomycin BGC in *S. venezuelae*. JadR1 has been shown to activate the expression of genes within the jadomycin BGC as well as repress genes involved in chloramphenicol biosynthesis (Chater, 2016). ForF is the only CSR encoded in the formicamycin BGC that appears to interact with DNA in another BGC, as it binds to a promoter in cluster 4 (**Table 1.1**). This could imply that ForGF, like many other TCSs that have been characterised in *Streptomyces* genomes, can regulate the expression of genes from multiple BGCs across the genome of *S. formicae*. As cluster 4 is predicted to be silent under formicamycin producing conditions, it is currently unclear what effect deletion of ForGF has on the expression of this BGC. The potential cross-cluster regulation by the ForGF TCS is investigated later in this thesis.

TCSs are known to respond to extracellular signals but the specific signals that most TCSs respond to remain unknown (Hutchings *et al.*, 2004). Similarly, we do not know what signal(s) the ForGF TCS responds to. We do know that formicamycins are produced by *S. formicae* on both MYM and SFM, both of which are relatively rich and complex media, and production is evident relatively early in the growth of the organism, usually by day 5 and the CSRs are all expressed by day 2. This suggests that nutrient depletion is not the signal activating the transcription of the formicamycin biosynthetic genes, as is the case for some other BGCs. In the future, characterising the signals that activate cluster-situated activators of antibiotic biosynthesis like ForGF will be useful as the knowledge could be used to activate those BGCs which remain silent under normal laboratory conditions (Rutledge and Challis, 2015).

The work in this chapter also shows that expression of all the major transcripts in the *for* BGC are likely repressed by binding of the MarR regulator ForJ to their promoter regions. This is consistent with the fact that a *ΔforJ* mutant has significantly increased levels of formicamycin production compared to the wild-type. The vast majority of MarR-type regulators reported in bacterial and archaeal genomes are repressors of transcription (Wilkinson and Grove, 2006). The ligand binding capabilities of these MarR proteins implies that the accumulation of a ligand, perhaps the early biosynthetic starter unit generated by ForHI, causes ForJ to release from the DNA, de-repressing the expression of the formicamycin transcripts. The other MarR regulator encoded in the formicamycin BGC, ForZ, appears to be involved in the regulation of the host-resistance mechanism by controlling the expression of a major transporter, ForAA, in response to accumulating levels of formicamycins or other biosynthetic intermediates. CSRs of this type such as MarR and TetR regulators that can alter the levels of transcription in response to signals from one or more molecules within the cell are referred to as one component systems (Cuthbertson and Nodwell, 2013). In the future, it will be important to experimentally characterize the DNA binding sites identified in this chapter and the ligand binding capabilities of both the MarR regulators encoded in the formicamycin BGC. This could be done using an electrophoretic mobility shift assay (EMSA) where the purified protein is incubated with fluorescently tagged DNA probes and then run on a polyacrylamide gel. Protein-DNA complexes run more slowly through a gel compared to free protein and therefore protein-DNA interactions are represented by an upward 'shift'. In order to characterise ligand binding, formicamycin precursors and intermediates could be included in the EMSA reactions to determine whether the presence of these compounds abolishes the DNA-protein interaction. Furthermore, with the purified proteins in hand, crystal structures for both MarR regulators could also be generated. If possible, crystals of the regulator bound to both the ligand and/or DNA could be generated, however, this may be challenging as the optimal conditions for protein purification are often not optimal for stable DNA. This will be the subject of further investigations in the Hutchings laboratory.

Whilst the results from the cappable RNA sequencing experiment were highly informative and allowed the locations of all the TSSs to be identified, this type of transcriptomic analysis provides no information about the strength of the promoters or the levels of transcription of each gene/transcript in relation to others in the genome. In the future, transcription levels could be quantified using traditional RNA sequencing methods or qRT-PCR in both the wild-type and the regulator mutant backgrounds (*S. formicae ΔforJ, S. formicae ΔforGF* and *S. formicae ΔforZ*) to determine how the levels of transcription change in the absence of each CSR. For example, we might expect levels of the major transcripts from the *for* BGC to be higher in *S. formicae ΔforJ*, as this strain produces higher titres of formicamycin congeners and does not have the major repressor, however, this may not be the case for every transcript within the BGC. It would also be interesting to identify whether ForZ has any indirect roles in controlling levels of expression of any *for* BGC transcripts or other CSRs in addition to the de-repression of *pforZ/AA*.

Promoter activity can be measured by fusing the promoter to a reporter gene and then measuring the activity of the reporter gene product. In this way, the strengths of each promoter in the BGC can be compared in different genetic backgrounds. A stronger promoter will result in higher levels of reporter activity and vice versa. A reporter assay using the β-glucuronidase (*gusA*) gene has been developed for the characterisation of promoter strength and transcription levels in *Streptomyces* (Myronovskyi *et al.*, 2011). When incubated with β-glucuronide substrates such as 4-nitrophenyl β-D-glucopyranoside (PNPG), β-glucuronidase produces coloured breakdown products by hydrolysis, specifically 4-nitrophenol (PNP) which has a peak absorbance at 415 nm, the presence of which can be measured spectrophotometrically. The protocol for conducting these GUS reporter assays in *S. formicae* grown on solid culture was recently optimised by Katie Noble (former undergraduate student in the Hutchings Laboratory). Each of the promoters identified in the formicamycin BGC have been fused to *gusA* and conjugated into both the wild-type and *S. formicae* Δ*forJ* strains. Initial results show that the activities of *pforM, pforG* and *pforU* are significantly higher in *S. formicae* Δ*forJ* compared to the wild-type strain, whereas the activities of *pforH*, *pforT, pforZ* and *pforAA* do not

change. This suggests that ForJ represses transcription of *forMLK, forGF* and *forUVWX* but does not affect expression of *forHI* (which is likely activated by ForF), *forTSRABCDE*, for*Z* or *forAA*. Quantitative RT-PCR analysis is required to confirm these data but it leaves the question of what controls production of the core PKS through expression of *forTSRABCDE*. Currently, the *gusA* fusion to *pforJ* has not been tested as the strain was not available at the time of this undergraduate project, however, a full library of *S. formicae* GUS strains has since been constructed and will be the subject of future investigation to thoroughly compare the activities of each promoter in either the wild-type or formicamycin overproducing background, as well as in *S. formicae* Δ*forZ* and *S. formicae* Δ*forGF* mutants. This will help to confirm our hypotheses about the regulation of formicamycin biosynthesis, as it will highlight which CSRs control the activity of which promoters through the DNA binding activities identified in the ChIP-Sequencing experiment.

The knowledge of formicamycin regulation and host-resistance gained during this chapter has also been used to generate mutant strains of *S. formicae* with enhanced formicamycin production. In the future, it may be possible to generate even more enhanced production strains. For example, the stambomycin BGC in *S. amboficiens* is poorly expressed under normal laboratory conditions, however, activation of the cluster-situated LAL regulator, samR0484, by constitutive overexpression induced the expression of biosynthetic genes leading to the discovery of these novel macrolide antibiotics (Laureti *et al.*, 2011). Using the same approach, we could use CRISPR to knock-in a constitutive promoter such as *permE** in front of the genes encoding for ForGF in the *S. formicae ΔforJ*: ΦBT1 f*orAA pforAA* mutant background. This strain should have enhanced formicamycin biosynthesis as a result of deletion of the repressor ForJ and the overexpression of the activator ForGF, as well as enhanced host-resistance through the extra copy of ForAA. Furthermore, if ForGF does indeed regulate other genes outside of the formicamycin BGC, it is possible that constitutive activation of ForGF may affect the transcription of other secondary metabolite genes in *S. formicae*. In addition, an extra copy of ForK could also be added to the overproducing strain at a different integrative site to further help with

the host-resistance mechanism. A strain like this could be extremely valuable for the industrial production of these antibiotics.

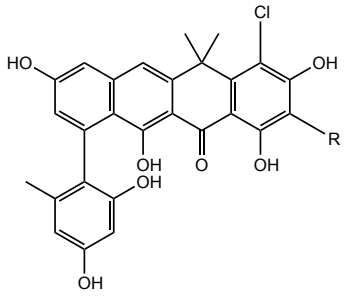
Overall, the work in this chapter demonstrates that *S. formicae* has evolved a tightly regulated system for controlling the expression of formicamycin biosynthesis genes and the coordination of the host-resistance mechanism by compound efflux. This ensures that biosynthesis of these antibiotics is not deleterious to the host, but rather provides an evolutionary advantage in its natural environment. The results from these experiments show how the transcription of the biosynthetic genes encoded in the formicamycin BGC is initiated. The following chapters will investigate the functions of each of these genes in more detail and begin to describe how the biosynthetic pathway progresses to produce the complex mixture of fasamycins and formicamycins seen in culture extracts of *S. formicae.*

5 The biosynthetic and evolutionary link between fasamycins and formicamycins

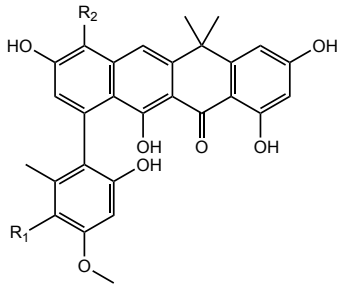
Previous work in this thesis has shown that the formicamycin BGC in *S. formicae* also produces pentacyclic polyketides called fasamycins. Despite the common biosynthetic origin, the carbon backbone and 3-dimensional structure of the fasamycins and formicamycins is significantly different. Previously, other fasamycins have been isolated from environmental DNA (eDNA) heterologously expressed in an *S. albus* host, but there were no co-produced formicamycins reported (Feng, Kallifidas and Brady, 2011). Experiments presented earlier in this thesis indicate that the fasamycins are the biosynthetic precursors of the formicamycins, as they are present in lower titres in the extracts of *S. formicae*, display lower bioactivity and accumulate in the *ΔforV* mutant. Although we know that ForV has a gatekeeper function, little else is known about the biosynthetic steps that link the fasamycins and formicamycins or how *S. formicae* came to evolve a single BGC that was capable of synthesising both molecules; it may be that they originate from two distinct BGCs or that the fasamycin BGC requires the addition of further biosynthetic genes to convert these precursors into formicamycins. The aim of this chapter is to identify the evolutionary development of the fasamycin and formicamycin BGC in *S. formicae* by searching for related BGCs in other organisms and identifying the biosynthetic steps that link these related but distinct compound families. By identifying additional strains that contain the genes for fasamycin and/or formicamycin biosynthesis, we may be able to find a strain that can produce these compounds to higher titres. By understanding the biosynthetic steps that are involved in the conversion of a fasamycin to a formicamycin, the BGC could be rewired so that only the more potent formicamycins are produced, instead of a mixture of formicamycins and fasamycins. Furthermore, knowledge of the biosynthesis of these antibacterial compounds can be used to inform future drug discovery efforts, especially if novel mechanisms are discovered.

5.1 Searching for BGCs related to the formicamycin BGC

The soil is a highly competitive environment and the transfer of genetic material between microorganisms is common. Although the producing organism was never identified, eDNA encoding a fasamycin BGC was isolated from soil samples collected in Arizona (Feng, Kallifidas and Brady, 2011; Feng *et al.*, 2012). The core carbon backbone of the compounds produced by this cloned BGC is the same as the fasamycins purified from *S. formicae*, with the only differences being the chlorination sites and the additional O-methylation at the base of ring A on the molecules isolated from *S. formicae* (**Figure 5.1**).



Fasamycin A R=H
Fasamycin B R=Cl
(Feng, Kallifidas and Brady,2011)



Fasamycin C R_1=H, R_2=H
Fasamycin D R_1=H, R_2=Cl
Fasamycin E R_1=Cl, R_2=Cl
(Qin *et al.*, 2017)

Figure 5.1: Fasamycins A and B were isolated through the heterologous expression of cloned eDNA from soil samples collected in Arizona (Feng, Kallifidas and Brady, 2011). The related compounds, fasamycins C, D and E, are produced by the formicamycin BGC in *S. formicae* (Qin *et al.*, 2017).

Although *S. formicae* was isolated from the *Tetraponera* system, we have no evidence that it is primarily an ant-associated strain, as the symbiosis between plant-ants and actinomycetes has not been studied to the same extent as the attine ant system. It may be that *S. formicae* is a soil-dwelling actinomycete that happened to be in proximity of the *Tetraponera* nest at the time of sampling, or that it is a strain the ants acquired during normal foraging behaviour. It is possible that *S. formicae*

might somehow be related to the soil-dwelling organism whose DNA was isolated and cloned in Arizona, or they may have exchanged the genes for fasamycin biosynthesis when they encountered each other over the course of evolution. BlastP was used to compare the proteins encoded in the BGC isolated in Arizona, expressed on cosAZ154 (GenBank accession number HQ828985), to the gene products of the formicamycin BGC in *S. formicae.* The analysis showed that although there has been significant rearrangement of the coding regions between these BGCs, many of the same putative gene products are present (**Table 5.1**).

Table 5.1: BlastP analysis results comparing putative gene products from cosAZ154 to proteins from the formicamycin BGC. Annotations of the gene products from cosAZ154 are displayed as reported in (Feng, Kallifidas and Brady, 2011) with protein IDs for each putative gene product and % similarities to proteins encoded within the formicamycin BGC in *S. formicae*.

		Similarity to gene products from *S. formicae* BGC		
Putative gene product cosAZ154	**Protein ID**	**% coverage**	**% amino acid identity**	**Annotation**
Pyridoxamine 5'-phosphate oxidase-related FMN-binding protein	AEM44246.1	30	30	ForM
Luciferase family protein	AEM44247.1	58	33	ForY
Acyltransferase	AEM44248.1	-	-	-
Hypothetical protein	AEM44249.1	-	-	-
Glycoside hydrolase family 48 protein	AEM44250.1	-	-	-
Beta-galactosidase	AEM44251.1	-	-	-
Hypothetical protein	AEM44252.1	-	-	-
Serine/threonine kinase	AEM44253.1	-	-	-
Histidine kinase	AEM44254.1	57	27	ForG
Two-component transcriptional regulator (LuxR family)	AEM44255.1	98	36	ForF

MMPL-domain containing protein	AEM44256.1	-	-	-
SARP family transcriptional regulator	AEM44257.1	-	-	-
TetR family transcriptional regulator	AEM44258.1	-	-	-
FAD dependent monooxygenase	AEM44259.1	99	51	ForX
Antibiotic biosynthesis monooxygenase	AEM44260.1	94	72	ForU
whiE protein I	AEM44261.1	87	53	ForS
O-methyltransferase	AEM44262.1	99	65	ForT
Halogenase	AEM44263.1	99	60	ForV
Cyclase	AEM44264.1	97	59	ForL
Cyclase	AEM44265.1	98	54	ForD
Acyl carrier protein	AEM44266.1	96	28	ForC
Keto synthase beta subunit	AEM44267.1	97	53	ForB
Keto synthase alpha subunit	AEM44268.1	96	61	ForA
Cyclase	AEM44269.1	89	50	ForR
EmrB/QacA family drug resistance transporter	AEM44270.1	93	50	ForAA
O-methyltransferase	AEM44271.1	88	39	ForW
Cytochrome P450	AEM44272.1	-	-	-
Lankamycin synthase, modules 5 and 6	AEM44273.1	-	-	-
Sodium/hydrogen exchanger	AEM44274.1	90	34	ForK
Putative polyketide cyclase	AEM44275.1	96	44	ForD

Carboxylase	AEM44276.1	97	57	ForI
Acetyl-CoA carboxylase biotin carboxyl carrier protein	AEM44277.1	92	40	ForE
Acetyl-CoA carboxylase alpha subunit-like protein	AEM44278.1	98	61	ForH

Many of the gene products within the cloned fasamycin BGC on cosAZ154 show homology to proteins from the formicamycin BGC in *S. formicae*, especially the core biosynthetic proteins such as the KS α and β subunits (ForA and ForB), which show high percentage coverage and identity. This makes sense as the core carbon skeletons of the molecules are so similar that they likely require very similar enzymes for biosynthesis. There are homologs of ForH, ForE and ForI that are likely responsible for making the starter units for chain extension and a carrier protein for loading onto the KS. Homologs to all five putative cyclase enzymes (ForD, ForL, ForR, ForS and ForU) are also present within the eDNA-derived cluster and these are predicted to aromatise the polyketide intermediate to form the pentacyclic fasamycin backbone. Interestingly, two separate gene products on cosAZ154 show homology to ForD. It may be that one of the copies of *forD* has become redundant over the course of evolution, explaining its absence in the *S. formicae* cluster. Similarities between cosAZ154 and the formicamycin BGC continue with respect to the transporters present in the system, with homologs to both ForAA and ForK present. This further supports the hypothesis made in the previous chapter that these transporters work synergistically to provide host resistance via compound efflux.

It is also worth noting that only one halogenase is encoded in cosAZ154 but fasamycin B is halogenated at two positions on the carbon chain when heterologously expressed in *S. albus*. The halogenase shows high homology to ForV, which we have shown can act at four positions on the fasamycin and formicamycin backbone (**Chapter 3.3**). It is therefore likely that the enzyme encoded on cosAZ154 is also able to act non-specifically on multiple positions on the molecule. In fact, when the

halogenase from cosAZ154 is modelled in I-TASSER and overlaid onto the predicted model of ForV in PyMol, the proteins align almost exactly, especially at the previously determined active site and the site of FAD binding (**Figure 5.2**). Furthermore, the highly variable C-terminal region of ForV that could not be aligned to other known flavin-dependent halogenases in the database, does partially align to the predicted model of the halogenase from cosAZ154, implying that these enzymes may bind to the same carrier and function by the same mechanism. In the future, it would be interesting to genetically complement *S. formicae ⊿forV* with the halogenase gene from cosAZ154 to see whether the new complementation strain is able to synthesise the full catalogue of halogenated formicamycin congeners. If so, it may be that these proteins do indeed act via the same mechanism and both halogenases can chlorinate at multiple positions on the fasamycin and formicamycin backbones.

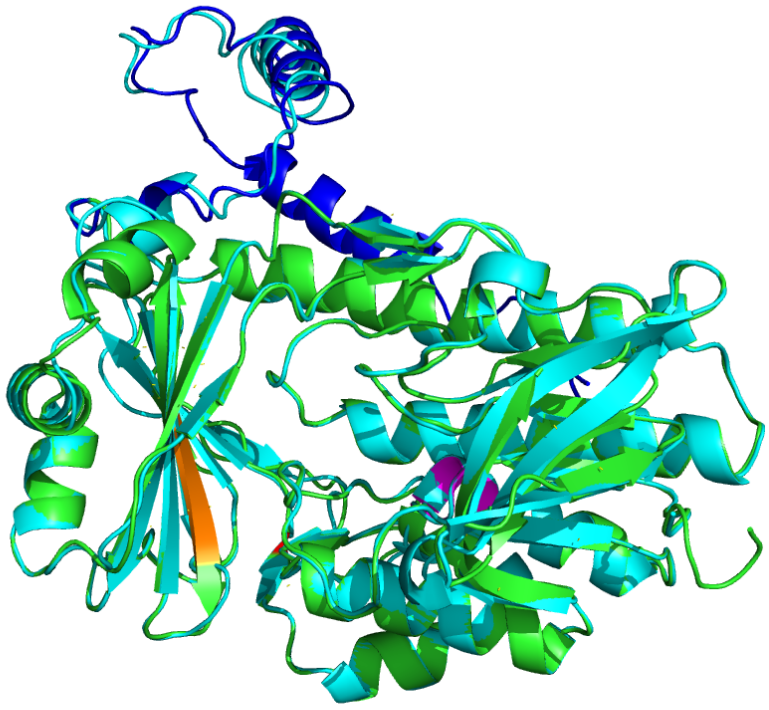


Figure 5.2: A predicted model of the halogenase from cosAZ154 was generated in I-TASSER and overlaid with the I-TASSER model of ForV in PyMol (generated in **chapter 3.3**). ForV is represented in green with a dark blue C-terminal tail and the halogenase from cosAZ154 is represented in cyan. The WxWxIP fingerprint is highlighted in orange, the FAD binding site GxGxxG motif is shown in magenta and the conserved lysine residue is in red. The models overlay well, showing high similarity between both peptides, including at the highly variable C-terminal region that could not be aligned to other flavin-dependent halogenases in the database (**Chapter 3.3**).

By comparing the gene homologs present in these two clusters, it is clear to see how cosAZ154 encodes the production of fasamycins A and B, given what we know about fasamycin biosynthesis in *S. formicae*. Although there is significant rearrangement of genes between the cluster on cosAZ154 and the *for* BGC in *S. formicae*, the percentage coverage and identities described here are generally much higher than the figures reported during the original study when the authors searched for protein homologs in the database at that time. This implies that there is a close evolutionary relationship between the BGC in *S. formicae* and the eDNA cloned into cosAZ154. The main difference between the two BGCs appears to be in the regulatory mechanisms that control production in the native host, which was never identified for the fasamycin cluster isolated in Arizona. Although there is a two-component system encoded on cosAZ154, the homology to ForGF is low, 27% identity between the response regulators and 36% identity between the sensor kinases. This is surprising, especially for the sensor kinase, considering these enzymes are known to all contain certain conserved domains in order to function. In addition, no MarR regulators are encoded on cosAZ154 but we have shown the importance of both ForJ and ForZ in the regulation of formicamycin production in *S. formicae* (**Chapter 4**). Instead, putative SARP and TetR regulatory proteins are present. This implies the regulation of production varies significantly between the host organisms, perhaps suggesting that the organisms themselves may not be especially closely related but they may have exchanged DNA encoding for the biosynthetic machinery when they encountered each other in the environment, something that is intriguing given the geographic separation of the strains when isolated.

On initial inspection of the biosynthetic gene products present encoded by the cosAZ154 BGC, there are no homologs missing that might account for the lack of formicamycin production when this cluster is heterologously expressed. However, it is possible that some of the genes are miss-annotated, as Blast analysis conducted in this way creates a bias towards aligning sequences even when only small areas of similarity are present. For example, the ForM and ForY homologs annotated in **table 5.1** are somewhat removed from the other biosynthesis genes, implying that they might actually lie outside of the border of the cluster. Furthermore, the percentage

similarities and identities to the proteins from the formicamycin biosynthetic pathway are relatively low for these gene products. The similarity highlighted during this analysis may just be in regions of cofactor binding, for example the flavin binding site of the oxidoreductase and the *S*-adenosyl methionine (SAM) binding site that is usually present in methyltransferase enzymes. Indeed, structural analysis implies that conversion of a fasamycin to a formicamycin requires reduction of ring C, which could plausibly involve an oxidoreductase enzyme like ForY*,* so the lack of a true ForY homolog on cosAZ154 may account for the lack of formicamycin production by this BGC. Similarly, the lack of a significant ForM homolog may account for the lack of O-methylation at the base of ring A on the fasamycins produced by cosAZ154 compared to those produced by *S. formicae*. The functions of ForY and ForM in the biosynthesis of the fasamycins and formicamycins from *S. formicae* are investigated later in this chapter.

Having identified similar genes to those in the formicamycin BGC on cosAZ154, we investigated the possibility that similar genes may also be found in multiple related microorganisms whose sequences are in the database. AutoMLST analysis conducted by Dr Neil Holmes shows that the closest relatives of *S. formicae* are *S. luteocolor* and *Streptomyces* sp NRRL_S-920 with an estimated average nucleotide identity (ANI) of 90.1% and 89.4% respectively (Alanjary, Steinke and Ziemert, 2019). The genomes for both *S. luteocolor* (accession no. NZ_BDGW00000000) and *Streptomyces* sp NRRL_S-290 (accession no. NZ_JODF00000000) were therefore obtained from NCBI and analysed using AntiSMASH. There was no evidence of any T2PKS clusters in either genome or any biosynthetic genes that looked similar to genes from the formicamycin BGC. This indicates that the formicamycin BGC is not present in these close relatives of *S. formicae*. However, the nucleotide sequence of the *for* BGC from *S. formicae* displays high similarity to a region of a contig from the genome of *Streptomyces kanamyceticus* NRRL B-2535 (GenBank ID LIQU00000000.1, contig LIQU01000034). The draft genome sequence of *S. kanamyceticus* currently available on NCBI is in 727 contigs (Labeda *et al.*, 2017). There is no annotation available for the sequence of contig LIQ01000034, however BlastN analysis comparing the genes present in the formicamycin BGC with the *S. kanamyceticus*

sequence revealed extremely high percentage identities, even at the nucleotide level (**Table 5.2**).

Table 5.2: Nucleotide Blasts comparing the *S. kanamyceticus* contig LIQU01000034 sequence with the formicamycin BGC shows high percentage identity between the genes present.

	Similarity to *S. kanamyceticus* contig	
Gene	**% coverage**	**% nucleotide identity**
forN	100	94
forM	99	95
forL	100	98
forK	99	94
forJ	100	96
forI	100	95
forH	99	94
forG	99	96
forF	99	96
forE	100	90
forD	100	96
forC	100	94
forB	100	96
forA	100	96
forR	100	95
forS	100	93
forT	100	96
forU	87	98
forV	100	95
forW	100	90

This suggests that most of the genes from the formicamycin BGC in *S. formicae* are also present in *S. kanamyceticus* NRRL B-2535. In addition, the putative genes in *S. kanamyceticus* are in the same order and orientation as in the formicamycin BGC, with a nucleotide Blast of the entire formicamycin BGC showing 99% coverage and 94% identity to the *S. kanamyceticus* DNA sequence. These striking similarities suggest there is a significant evolutionary relationship between the *S. kanamyceticus* DNA and the formicamycin BGC in *S. formicae*, even more so than the cosAZ154 BGC, which has been shown experimentally to encode fasamycins. The only difference between the cluster in *S. formicae* and the region of DNA in *S. kanamyceticus* is the lack of four genes from the 3' end of the formicamycin BGC; *forX, forY*, *forZ* and *forAA.* Apart from these four genes, the synteny between the formicamycin BGC and the *S. kanamyceticus* genome is significant and continues approximately 5 kb upstream and at least 40 kb downstream of the formicamycin BGC, which is as far as contig LIQU01000034 extends (**Figure 5.3**).

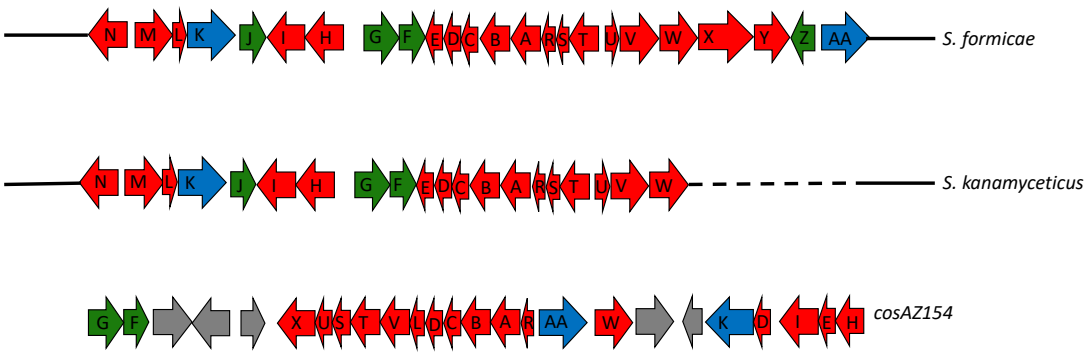


Figure 5.3: Representation of the putative cluster identified in *S. kanamyceticus* contig LIQU01000034 and the fasamycin BGC on cosAZ154 compared to the formicamycin BGC in *S. formicae*. Significant rearrangement of genes has occurred between cosAZ154 and the formicamycin BGC in *S. formicae*. The cluster in *S. kanamyceticus* looks much more similar to the BGC in *S. formicae*, however, four genes are missing; *forX, forY, forZ* and *forAA.* Continued synteny is represented by the solid black lines and missing regions are represented by the dashed line.

The gene product ForX is a putative flavin-dependent monooxygenase and ForY is a putative flavin-dependent oxidoreductase, the same oxidoreductase that is missing from cosAZ154. The genes *forZ* and *forAA* encode the MarR family transcriptional regulator and the MFS family transporter respectively (**Chapter 4**). It was predicted that production of the formicamycins requires the reduction of the fasamycin ring C and the introduction of a hydroxyl at C10, which could plausibly require a reductive

enzyme like ForY and a monooxygenase like ForX. Given the significant homology between the formicamycin BGC and the putative cluster on contig LIQU01000034 of the *S. kanamyceticus* genome, as well as the lack of *forX* and *forY* in the sequence of the latter, we hypothesized that *S. kanamyceticus* NRRL B-2535 may encode for a fasamycin BGC and may also be a close relative of *S. formicae*, even though it does not appear in the autoMLST results. *S. kanamyceticus* was therefore obtained from Professor Barrie Wilkinson at the John Innes Centre for further investigation. The strain is known as the producer of kanamycin, an antibiotic first isolated in 1957 which inhibits protein synthesis by binding to the 30S ribosomal subunit (Umezawa *et al.*, 1957; Misumi and Tanaka, 1980). PCR and sequencing analysis showed that *S. kanamyceticus* and *S. formicae* share 94% nucleotide similarity at the 16S level. *S. kanamyceticus* has not been the subject of extensive genetic work, although the BGC responsible for kanamycin production has been characterised. During these experiments, conjugation efficiency has commonly been reported as a limiting factor (Yanai and Murakami, 2004; Zhang *et al.*, 2018).

The optimal growth conditions for *S. kanamyceticus* proved to be very different to *S. formicae*, with significant sporulation only observed on oat agar after incubation at 28°C for 3-4 weeks. Growth under these conditions resulted in the production of yellow mycelium with white spores, similar to the phenotype of *S. formicae* when grown under formicamycin producing conditions. The fasamycins are yellow and fasamycin production is known to colour the mycelium of *S. formicae,* as overexpression of the pathway by deletion of *forJ* resulted in a yellow phenotype (**Figure 4.4**). As such, plates of *S. kanamyceticus* on both SFM and oat agar were extracted and analysed by HPLC(UV) LCMS (conducted by Dr Zhiwei Qin and Hannah McDonald, JIC) to determine whether any fasamycins were present in the culture extract. Under all conditions tested, no fasamycins or formicamycins were evident in the culture extracts of *S. kanamyceticus*, implying that the BGC may be silent under normal laboratory conditions and the yellow colour seen in cultures represents the production of another compound (data not shown).

As such, we attempted to use our knowledge of the regulation of formicamycin biosynthesis in *S. formicae* gained in the previous chapter to induce the expression

of this cluster in *S. kanamyceticus*. We hypothesised that by deleting the homolog to the major repressor, *forJ,* we could de-repress the pathway in *S. kanamyceticus* and induce expression of the BGC as we had done previously in the heterologous host, *S. coelicolor* M1146 (**Chapter 4.2.1**). Initially, a pCRISPomyces-2 vector was created to generate an unmarked *forJ* deletion in *S. kanamyceticus*, however, repeated conjugations with this plasmid did not result in any colonies. There are currently no reports of the CRISPR/Cas9 system being used in *S. kanamyceticus* for genetic manipulation. The poor quality, draft genome sequence of *S. kanamyceticus* makes designing unique and specific synthetic guide RNAs a challenge. As such, it could be that the lack of ex-conjugants during this experiment was due to the guide RNA non-specifically targeting areas of the genome and causing fatal double stranded breaks by the Cas9. In this case, improving the quality of the genome sequence should be a priority for improving the use of CRISPR/Cas9 in this strain. Alternatively, it has been suggested that the expression levels of the Cas9 nuclease used in many CRISPR systems is too strong for some organisms, also resulting in lethal double stranded DNA breaks occurring non-specifically across the genome (Tong, Weber and Lee, 2018; Alberti and Corre, 2019). It may therefore be that the pCRISPomyces-2 system is not suitable for use in *S. kanamyceticus*. In fact, when an empty pCRISPomyces-2 plasmid was conjugated into *S. kanamyceticus* there were no exconjugants isolated, even when significant numbers of spores and donor *E. coli* cells were used. Therefore, a CRISPR system that uses a weaker promoter, or even an inducible promoter like the pCRISPR-Cas9 system, may be more suitable for the genetic editing of *S. kanamyceticus* (Mo *et al.*, 2019).

Due to the lack of mutants obtained using the pCRISPomyces-2 system, the experiment for de-repression of the putative fasamycin BGC in *S. kanamyceticus* was re-designed and a suicide vector with an apramycin resistance gene was created that would disrupt *forJ* via homologous recombination. Multiple attempts to conjugate this new plasmid into *S. kanamyceticus* were trialled but all were unsuccessful. To confirm that induction of the putative fasamycin BGC in *S. kanamyceticus* was not fatal in the absence of ForX, ForY, ForZ and/or ForAA, ReDirect was used as previously described to genetically modify the 215-G PAC clone containing the whole

formicamycin BGC from *S. formicae.* The entire BGC was deleted and replaced with an apramycin resistance gene with the exception of the last four genes; *forX, forY, forZ* and *forAA.* This altered cosmid containing just *forX, forY, forZ* and *forAA* was then conjugated into *S. kanamyceticus* with the view of subsequently deleting *forJ* in any resulting ex-conjugants, however, none were isolated. Conjugation protocols were extensively optimised during this work, including increasing spore density, changing pre-germination times and temperatures, increasing the concentration of $MgCl_2$ or $CaCl_2$ in the conjugation plates and trialling methods of liquid mycelial conjugation described by other groups (Zhang *et al.*, 2018). However, none of these methods resulted in any ex-conjugants of *S. kanamyceticus*. This suggests that conjugation of large plasmids into *S. kanamyceticus* is highly inefficient and new protocols need to be developed to enable genetic manipulation in this organism.

The genetic manipulation of *S. kanamyceticus* is limited by the poor-quality genome sequence and the lack of available selection markers due to the organism's natural resistance to both kanamycin and hygromycin. This work shows that some non-model streptomycetes still remain genetically intractable or extremely challenging to manipulate even with the development of CRISPR and other molecular techniques. In future, we plan to capture the putative fasamycin BGC from *S. kanamyceticus* for heterologous expression in *S. coelicolor* M1146. This could be done via TAR cloning or through the generation of a cosmid library. As shown previously, de-repression of the 215-G PAC through the deletion of *forJ* does induce expression of the formicamycin BGC in *S. coelicolor* M1146 and therefore a similar approach could be taken with the fasamycin BGC from *S. kanamyceticus* in a host that is more genetically tractable. Conducting these future experiments would allow us to experimentally prove our hypothesis that the putative fasamycin BGC in *S. kanamyceticus* does encode biosynthesis of fasamycins and may be a close evolutionary relative of the formicamycin BGC in *S. formicae*. It does not appear that these strains share any other BGCs or other extensive regions of similarity between their chromosomes, however, a high-quality genome sequence of *S. kanamyceticus* is needed to confirm this.

Even without experimental characterisation of the putative fasamycin BGC in *S. kanamyceticus*, the extensive continued synteny either side of the four missing genes suggests an evolutionary relationship between these BGCs. We hypothesise that the fasamycin BGC evolved first, with the genes encoding ForX, ForY, ForZ and ForAA being acquired later by *S. formicae*. Blast analysis of the 4 gene region together does not result in any significant matches and there is no evidence of this region of DNA being on a transposable element. However, BlastP analysis shows that the closest homologs to ForY, ForZ and ForAA are all from strains of the *Actinomadura* genus, implying that the genes encoding these three may have been acquired together in a single insertion event, with *forX* being acquired on its own from a different source at another point in time (**Table 5.3**). It is conceivable that the addition of genes that enable a fasamycin producer to synthesise the more potent formicamycins would represent an advantageous adaptation and one that would likely become permanent over the course of evolution.

Table 5.3: Amino acid Blast results of ForX, ForY, ForZ and ForAA. ForY, ForZ and ForAA all originate from Actinomadura species. In fact, it is not just the top hit that is from this species, but the top 6, 7 and 10 results for ForY, ForZ and ForAA respectively.

Gene product	First hit in the database	% amino acid identity	% coverage	Accession number
ForX	Hypothetical protein from *Streptomyces caatingaensis*	50	94	WP_049717083.1
ForY	LLM class flavin-dependent oxidoreductase from *Actinomadura meyerae*	60	87	WP_089326038.1
ForZ	MarR transcriptional regulator from *Actinomadura pelletieri*	42	84	WP_121435158.1
ForAA	DHA2 family efflux MFS transporter permease subunit *Actinomadura* sp. NEAU-Ht49	57	96	WP_122199503.1

BlastP searches and Multigene Blast of the core peptides from the formicamycin BGC has not revealed any other related BGCs within the database. Therefore, if our assumptions about the functions of the biosynthetic genes are correct, the formicamycin BGC in *S. formicae* represents the only currently characterized pathway that produces formicamycins and one of only 3 sequenced BGCs that appear to contain genes for fasamycin biosynthesis.

5.2 The mechanism of conversion of a fasamycin to a formicamycin

As discussed in the previous section, it appears that the fasamycins are the evolutionary and biosynthetic precursors of the formicamycins. It is predicted that the formicamycin BGC evolved from a fasamycin BGC as a results of multiple horizontal gene transfer events, specifically the acquisition of four genes, *forX, forY*, *forZ* and *forAA.* The functions of ForZ and ForAA were investigated in the previous chapter; they are involved in regulating host-resistance to the formicamycins. ForX and ForY are predicted to be the biosynthetic enzymes responsible for converting a fasamycin into a formicamycin in *S. formicae*. To confirm this hypothesis and assign functions to these gene products, in frame CRISPR/Cas9 deletion mutants were generated. Three biological replicates of the Δ*forX* and Δ*forY* mutants were made and incubated under formicamycin producing conditions and LCMS analysis was conducted by Dr Zhiwei Qin at the JIC. Both *S. formicae* Δ*forX* and *S. formicae* Δ*forY* lost the ability to produce any formicamycins (**Figure 5.4**). *S. formicae* Δ*forX* only produces fasamycins C-E (**1-3**) and accumulates fasamycin E (**3**), the di-chlorinated fasamycin, to approximately 25 times the level of the wild-type strain. This suggests that ForX plays an important role in the conversion of a fasamycin to a formicamycin as without it, biosynthesis of the formicamycins does not occur and the most bioactive fasamycin precursor accumulates. *In-trans* complementation of *forX* under the control of the native promoter restored biosynthesis of the formicamycins. In the *S. formicae* Δ*forY* culture extracts, there were no recognisable fasamycins or formicamycins present. Instead six new peaks (**a-f**) were present that had not been seen in previous experiments. Complementation of Δ*forY* under the control of the native promoter restored biosynthesis of the formicamycins to approximately wild-

type levels, however, some of the new peaks from the Δ*forY* mutant still remained in the culture extract after complementation, suggesting expression levels of *forY* may have been lower in the complemented strain than in the wild-type.

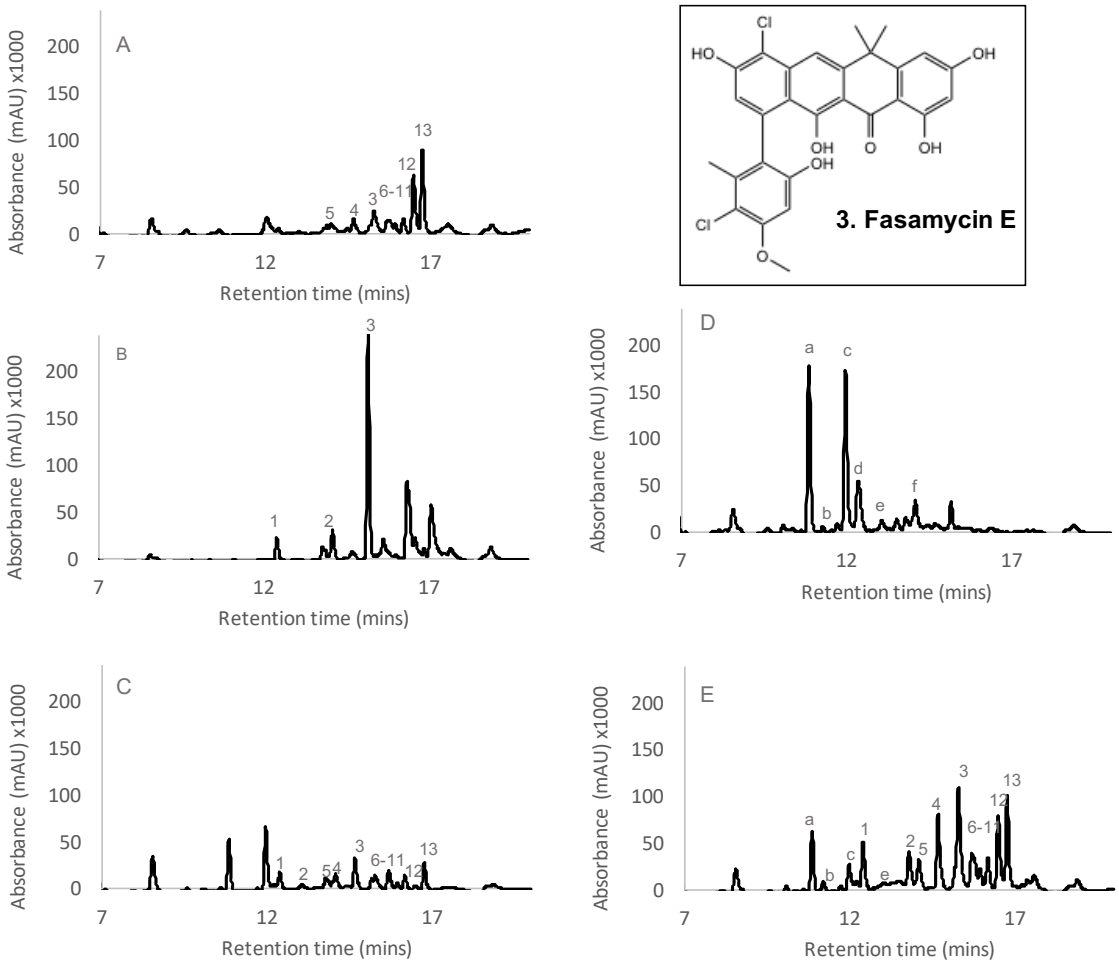
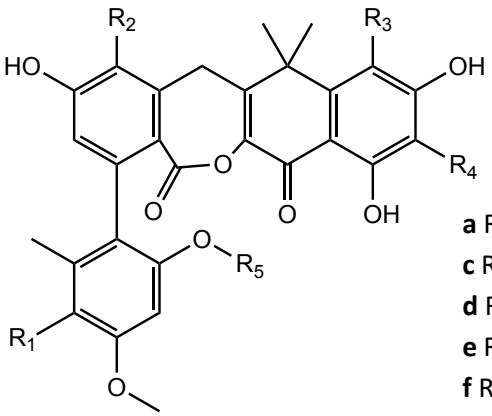


Figure 5.4: HPLC traces (UV 285 nm) of extracts from: (A) *S. formicae* wild-type; (B) *S. formicae* Δ*forX*; (C) *S. formicae* Δ*forX:forX* φC31; (D) *S. formicae* Δ*forY*; (E) *S. formicae* Δ*forY:forY* φC31. Deletion of *forX* results in the production of the fasamycins only with the di-chlorinated fasamycin E (**3**) accumulating (see insert for structure). Deletion of *forY* abolishes the production of fasamycins and formicamycins, with six new peaks (**a-f**) present instead. Wild-type production returns on complementation of each gene under the native promoter. HPLC(UV)/LCMS analysis conducted by Dr Zhiwei Qin at the John Innes Centre.

Dr Zhiwei Qin at the John Innes Centre scaled up the fermentations of *S. formicae* Δ*forY* in order to isolate each of the new compounds present. Small amounts (between 5 and 16 mg) of five of the compounds (**a, c, d, e** and **f**) were isolated and their structures solved by MS and NMR (**Table 5.4** and **Figure 5.5**). Production levels of **b** were too low to allow sufficient yields for isolation and structural elucidation.

Table 5.4: Large scale cultures of *S. formicae* Δ*forY* were grown up for compound isolation and purification by Dr Zhiwei Qin. Yields of compound b were too low for compound isolation. LCMS analysis of each compound by Dr Zhiwei Qin allowed molecular mass and molecular formula of each compound to be identified.

	Molecular mass	Molecular formula	Yield from 4L agar
a	522	$C_{28}H_{23}O_8Cl$	13 mg
b	536	$C_{29}H_{25}O_8Cl$	-
c	556	$C_{28}H_{22}O_8Cl_2$	10 mg
d	570	$C_{29}H_{24}O_8Cl_2$	5 mg
e	590	$C_{28}H_{21}O_8Cl_3$	13 mg
f	604	$C_{29}H_{23}O_8Cl_3$	6.3 mg



a R_1=H, R_2=Cl, R_3=H, R_4=H, R_4=H
c R_1=Cl, R_2=Cl, R_3=H, R_4=H, R_4=H
d R_1=Cl, R_2=Cl, R_3=H, R_4=H, R_4=CH_3
e R_1=Cl, R_2=Cl, R_3=H, R_4=Cl, R_4=H
f R_1=Cl, R_2=Cl, R_3=H, R_4=Cl, R_4=CH_3

Figure 5.5: The structures of five of the compounds from the extract of *S. formicae* Δ*forY* were isolated and their structure solved by NMR and MS by Dr Zhiwei Qin, JIC. Production levels of **b** were too low to allow sufficient yields for isolation and structural elucidation. All five of the isolated compounds contain a lactone at ring C, indicative of an intermediate between a fasamycin and a formicamycin.

All five of the isolated compounds have neither a fasamycin nor a formicamycin core carbon skeleton and instead are ring C lactone intermediates with various levels of additional post-PKS tailoring such as halogenation and O-methylation. Although yields of compound **b** were too low to allow for isolation and structural elucidation, we estimate from the molecular mass that it is the same as compound **a** with the addition of methyl group at R_4. The identification of these lactone intermediates allows us to predict how ForX and ForY function together to convert a fasamycin into a formicamycin. Monooxygenase enzymes like ForX are responsible for introducing

one atom of molecular oxygen into organic substrates, generating water as a bi-product. Baeyer-Villiger monooxygenases (BVMOs) are often present in catabolic pathways and are known to convert cyclic ketones into lactones. We therefore hypothesise that ForX is a BVMO that converts the halogenated fasamycin precursor into the 7-membered ring C lactone intermediate. The lactone intermediate generated by ForX can then be reduced by ForY, the flavin-dependent oxidoreductase, forming the formicamycin scaffold.

There are three classes of BVMOs; type 1 enzymes are encoded by a single gene product and are characterised by a tightly bound FAD cofactor and the presence of the FXGXXXHXXXWP fingerprint sequence. For type 1 BVMOs, the NADPH/$NADP^+$ cofactor is bound during catalysis. In contrast, type 2 BVMOs are encoded by two separate gene products; the reductase that generates the reduced FMN using NAD(P)H and the distinct oxygenase. The third type of BVMO was classified more recently and is known as the type 'O' or atypical class. These enzymes are encoded by a single gene product, have FAD tightly bound to the surface are NAD(P)H-dependent, but do not contain the FXGXXXHXXXWP fingerprint sequence and are therefore neither type 1 or type 2 BVMOs (Leisch, Morley and Lau, 2011). BlastP analysis of ForX shows a conserved FAD binding domain but no fingerprint sequence, implying it is probably the latter type of BVMO.

We hypothesise that this two-step ring-expansion, ring-contraction mechanism is responsible for the conversion of a flat fasamycin to the twisted L-shaped structure of the formicamycins. To the best of our knowledge, this proposed mechanism is unique in polyketide biosynthesis. Efforts in the Wilkinson laboratory to express and purify both ForX and ForY for biochemical analysis have so far been unsuccessful, however, feeding experiments conducted by Dr Zhiwei Qin confirm that when *S. formicae* Δ*forX* is incubated with a lactone intermediate from *S. formicae* Δ*forY*, production of formicamycins B, E, F and J is reinstated, showing that the generation of the lactone intermediates by ForX is a vital step in the conversion of a fasamycin to a formicamycin. The accumulation of fasamycin E in *S. formicae* Δ*forX*, is consistent with the earlier observation that *S. formicae* Δ*forV* accumulates fasamycin C, the non-halogenated fasamycin and confirms that the fasamycins are the

biosynthetic precursors to the formicamycins. These results also suggest that the halogenase, ForV, acts first and performs a gatekeeper function within this pathway, controlling the conversion of fasamycins to formicamycins. A key feature of many flavin-dependent halogenases like ForV is their ability to activate molecular oxygen using reduced flavin to generate flavin hydroperoxide, which allows for diverse downstream reactions including hydroxylation and Baeyer-Villiger oxidation (Latham *et al.*, 2017). The identification of halogenated lactone intermediates confirms that the monooxygenase ForX accepts substrates that have already been halogenated by ForV. We therefore predict that the activity of ForX is dependent on the generation of hydroperoxy-flavin by the halogenase ForV. The formicamycin backbone generated after the two-step ring-expansion, ring-contraction reaction catalysed by ForX and ForY can also be further halogenated by ForV, showing that these three flavin-dependent enzymes work closely together to convert the fasamycin precursors into the more potent formicamycins (**Figure 5.6**).

Figure 5.6 (see following page for figure legend)

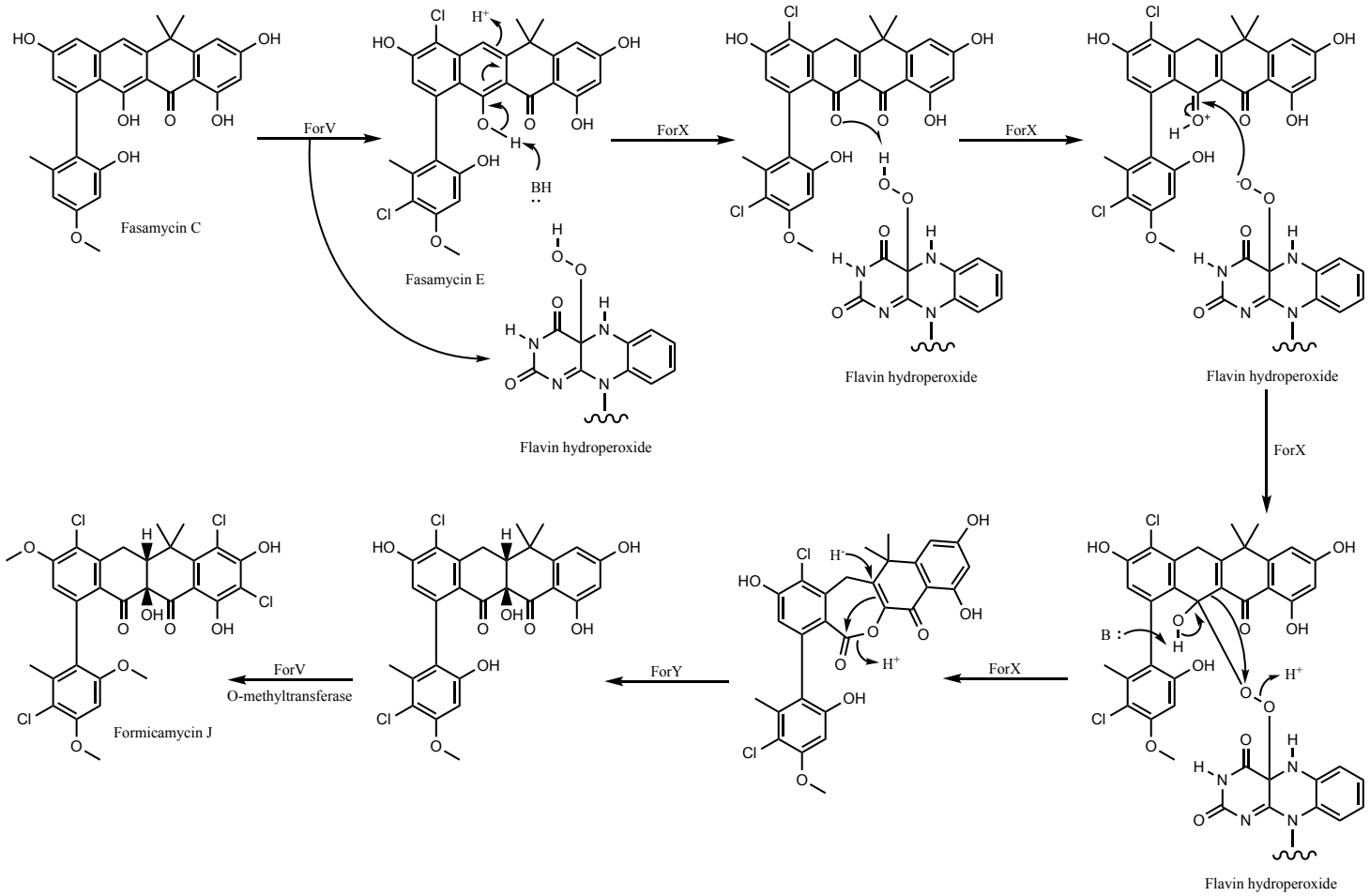


Figure 5.6 (previous): Formicamycins are converted to fasamycins by the activity of three flavin dependent enzymes; ForV, ForX and ForY in a novel two-step ring-expansion, ring-contraction mechanism. The activity of the halogenase, ForV generates hydroperoxy flavin. The Baeyer-Villiger monooxygenase, ForX, then uses this flavin to generate the 7-membered ring C lactone intermediate. The oxidoreductase, ForY, then reduces the lactone intermediate to form the formicamycin scaffold. This formicamycin scaffold can then be further halogenated by ForV and methylated by a so far unidentified O-methyltransferase. Adapted from Qin et al (in prep) with permission.

In order to determine whether the expansion of ring C affects bioactivity, bioassays against *B. subtilis* were set up using the isolated lactone intermediates (**Figure 5.7**).

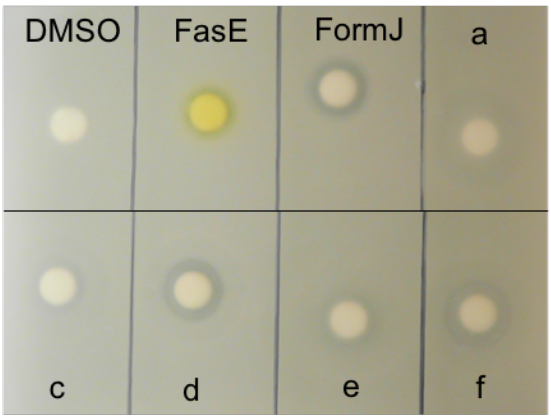


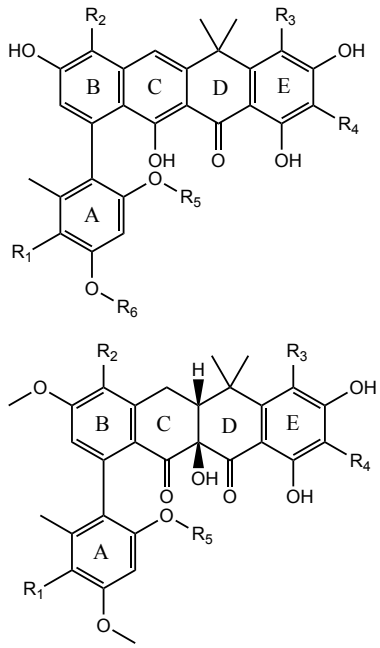
Figure 5.7: The bioactivity of the lactone ring intermediates against *B. subtilis* was examined by disc diffusion bioassay. Solutions of each compounds were made to 1 mM in a 20 ul volume. The whole volume was dried onto Whatman paper discs and placed on soft LB agar inoculated with *B. subtilis*. After overnight incubation at 30°C, zones of inhibition were observed. DMSO solvent controls were included as well as fasamycin E (FasE) and formicamycin J (FormJ) for comparison. The assays showed that the lactone intermediates contain bioactivity that is comparable to the formicamycins, implying that the expansion of ring C has no effect of the mechanism of action of these compounds.

The bioassays show that expansion of ring C to a 7-membered ring does not seem to significantly affect bioactivity as the lactone intermediates display comparable inhibitory activity against *B. subtilis* to the fasamycins and formicamycins. In addition, increased chlorination of the lactone backbone seems to increase the bioactivity in the same way as the fasamycins and formicamycins, with the di- and tri-chlorinated compounds giving the largest zones of inhibition. This shows that expansion of ring C does not change the active pharmacophore of the molecules in a way that causes decreased bioactivity, suggesting this area of the molecule is not essential for target interactions. This work also expands the family of bioactive compounds to originate from *S. formicae* and the formicamycin BGC. Furthermore, the knowledge of this two-step ring-expansion, ring-contraction mechanism in a

polyketide pathway may allow for further bioactive analogues of both formicamycins and other cyclic polyketides to be produced in the future.

5.3 Methylation of the fasamycins and formicamycins

Previous work in this thesis has identified the minimal genes required for the biosynthesis of the fasamycins and formicamycins and has begun to characterise the functions of the tailoring enzymes that modify the cyclised polyketide to generate the mixture of bioactive metabolites seen in extracts of *S. formicae*. As well as studying the conversion of a fasamycin to a formicamycin, we have shown that a single halogenase, ForV, is responsible for the halogenation of the molecules, generating congeners with increased bioactivity compared to the non-halogenated precursors. All fasamycin and formicamycin congeners also display various methyl groups that affect the structure and possibly also the bioactivity of the molecules (**Figure 5.8**). For example, all fasamycins, formicamycins and ring C lactone intermediates contain a di-methylation on ring D. All fasamycins and formicamycins from *S. formicae* show methylation at the base of ring A, although fasamycins A and B from cosAZ154 do not (**Figure 5.1**). Some of the formicamycin congeners and some of the lactone intermediates, but none of the fasamycins, are also O-methylated at a second site on ring A (R5). Finally, all the formicamycin congeners are O-methylated on ring B, but the fasamycins and lactone intermediates are not, implying this reaction may be an essential part of the conversion of a fasamycin to a formicamycin that occurs after Baeyer-Villiger oxidation by ForX. In addition, the lactone intermediates described in the previous section do not contain this methylation on ring B, therefore this reaction must occur after the action of ForX and may therefore be involved in the conversion of a fasamycin to a formicamycin.



Fasamycin A R_1=H, R_2=H, R_3=Cl, R_4=H, R_5=H, R_6=H
Fasamycin B R_1=H, R_2=H, R_3=Cl, R_4=Cl, R_5=H, R_6=H
1.Fasamycin C R_1=H, R_2=H, R_3=H, R_4=H, R_5=H, R_6=CH_3
2.Fasamycin D R_1=H, R_2=Cl, R_3=H, R_4=H, R_5=H, R_6=CH_3
3.Fasamycin E R_1=Cl, R_2=Cl, R_3=H, R_4=H, R_5=H, R_6=CH_3

4.Formicamycin A R_1=H, R_2=Cl, R_3=H, R_4=H, R_5=H,
5.Formicamycin B R_1=Cl, R_2=Cl, R_3=H, R_4=H, R_5=H,
6.Formicamycin C R_1=H, R_2=Cl, R_3=Cl, R_4=H, R_5=CH_3,
7.Formicamycin D R_1=Cl, R_2=Cl, R_3=Cl, R_4=H, R_5=H
8.Formicamycin E R_1=Cl, R_2=Cl, R_3=Cl, R_4=H, R_5=CH_3
9.Formicamycin F R_1=Cl, R_2=Cl, R_3=H, R_4=Cl, R_5=CH_3
10.Formicamycin G R_1=H, R_2=Cl, R_3=Cl, R_4=Cl, R_5=CH_3
11.Formicamycin H R_1=Cl, R_2=H, R_3=Cl, R_4=Cl, R_5=CH_3
12.Formicamycin I R_1=Cl, R_2=Cl, R_3=Cl, R_4=Cl, R_5=H
13.Formicamycin J R_1=Cl, R_2=Cl, R_3=Cl, R_4=Cl, R_5=CH_3
14.Formicamycin K R_1=H, R_2=Cl, R_3=Br, R_4=Cl, R_5=CH_3
15.Formicamycin L R_1=Cl, R_2=Cl, R_3=Br, R_4=Cl, R_5=CH_3
16.Formicamycin M R_1=H, R_2=Br, R_3=H, R_4=H, R_5=CH_3

Figure 5.8: Structures of the fasamycin (**1-3**) and formicamycin (**4-16**) congeners from *S. formicae*. All congeners from the formicamycin BGC in *S. formicae* are di-methylated on ring D and O-methylated at the base of ring A. All formicamycins are O-methylated on ring B but fasamycins are not. Some formicamycins can also be O-methylated at R5.

Most methyltransferase enzymes transfer a methyl group from the cofactor SAM to a carbon, oxygen or nitrogen group on the relevant substrate(s) (Hertweck *et al.*, 2007). BlastP analysis of both ForW and ForT shows that they contain conserved domains of the O-methyltransferase family of enzymes and ForM contains a known methyltransferase domain. To confirm the roles of each individual methyltransferase enzyme within the formicamycin BGC, their encoding genes were deleted using CRISPR/Cas9. Three biological replicates of each mutant were made and cultured under formicamycin producing conditions, before analysis by LCMS was conducted by Dr Zhiwei Qin at the JIC. Deletion of *forM* resulted in the production of no fasamycin or formicamycin congeners, with no evidence of any new intermediates or shunt metabolites present in the culture extract. Complementation *in trans* under the native promoter restored production of the fasamycin and formicamycin congeners, although production levels were considerably lower than

the wild-type, perhaps due to changes in transcription levels caused by changing the location of the gene within the genome (**Figure 5.9**). The fact that the *ΔforM* mutant produces no fasamycin or formicamycin congeners or any biosynthetic intermediates suggests that the reaction catalysed by ForM may be an essential step in this biosynthetic pathway.

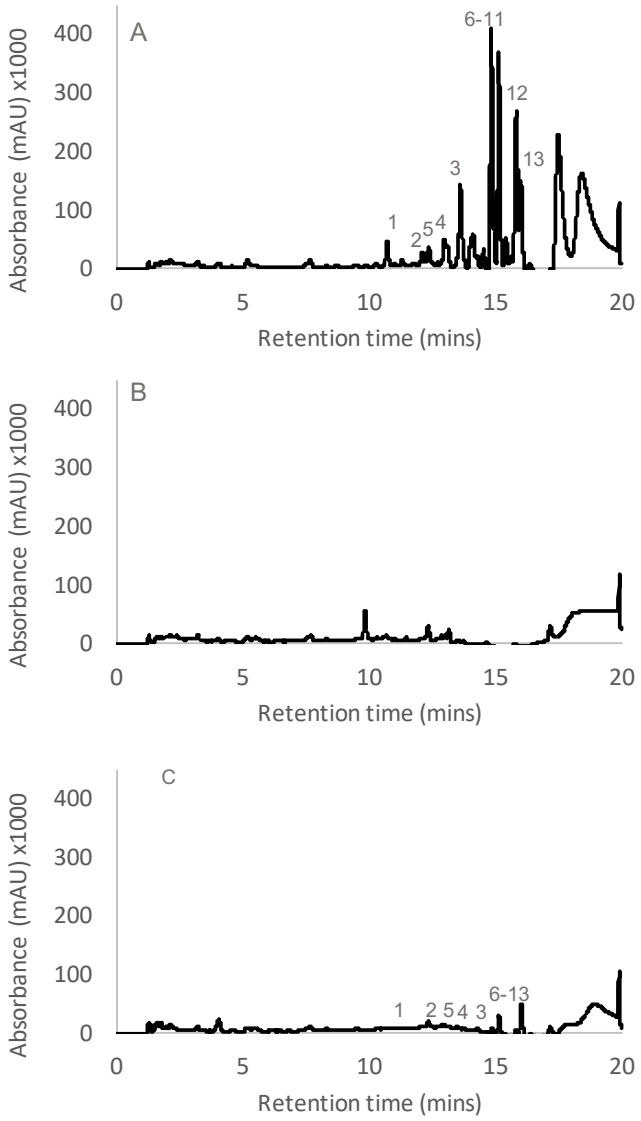


Figure 5.9: HPLC traces (UV 285 nm) of extracts from: (A) *S. formicae* wild-type; (B) *S. formicae* Δ*forM*; (C) *S. formicae* Δ*forM*: φBT1 *forM pforM*. Deletion of *forM* abolishes the production of fasamycins and formicamycins but production returns on complementation under the native promoter. HPLC(UV) LCMS analysis conducted by Dr Zhiwei Qin at the John Innes Centre.

Deletion of *forT* also abolished the production of fasamycins, formicamycins or any biosynthetic intermediates. Genetic complementation *in trans* under the native

promoter was unsuccessful, however complementation *in trans* under the constitutive promoter *permE** resulted in restored formicamycin biosynthesis but at lower levels than the wild-type (**Figure 5.10**). This is likely due to differences in transcription levels by placing the gene at a different location in the genome. These results suggest that methylation by ForT is also an essential step during the biosynthesis of the fasamycins and formicamycins.

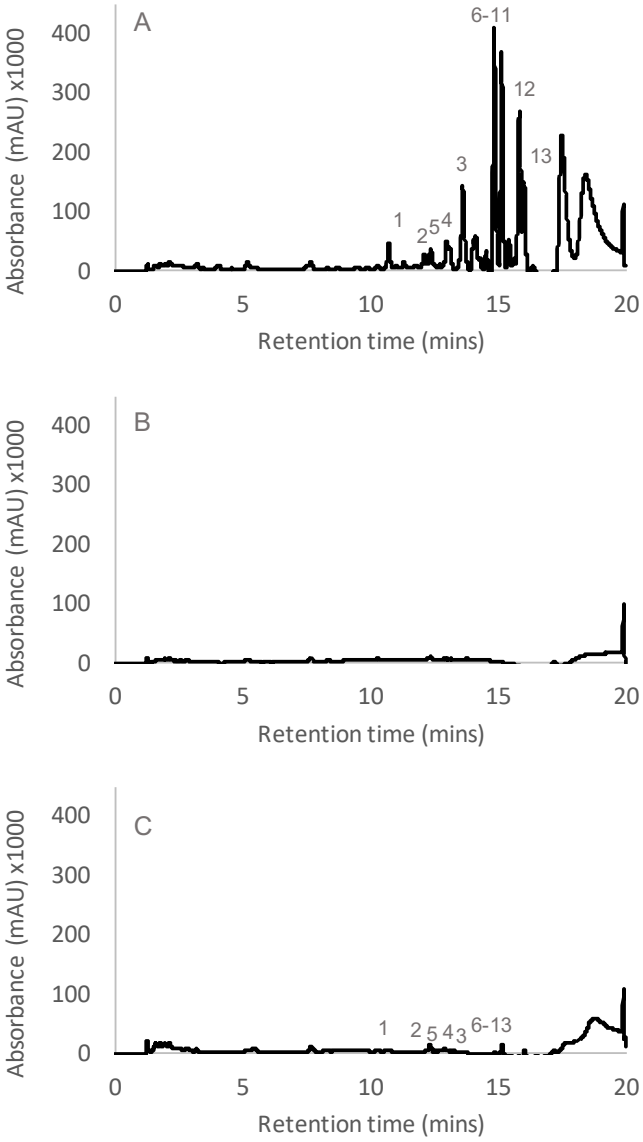


Figure 5.10: HPLC traces (UV 285 nm) of extracts from: (A) *S. formicae* wild-type; (B) *S. formicae* Δ*forT*; (C) *S. formicae* Δ*forT*: φBT1 *forT permE**. Deletion of *forT* abolishes the production of fasamycins and formicamycins but production returns on complementation under the native promoter. HPLC(UV) LCMS analysis conducted by Dr Zhiwei Qin at the John Innes Centre.

Deletion of *forW* abolished the production of the formicamycin congeners and resulted in just the fasamycins C-E (**1-3**) being produced, with fasamycin E (**3**) accumulating in the culture extracts (**Figure 5.11**). Complementation *in trans* under the native promoter restored production of all the formicamycin congeners. This suggests that ForW is essential for the conversion of fasamycins to formicamycins and indicates that ForX and ForY do not function in the absence of ForW.

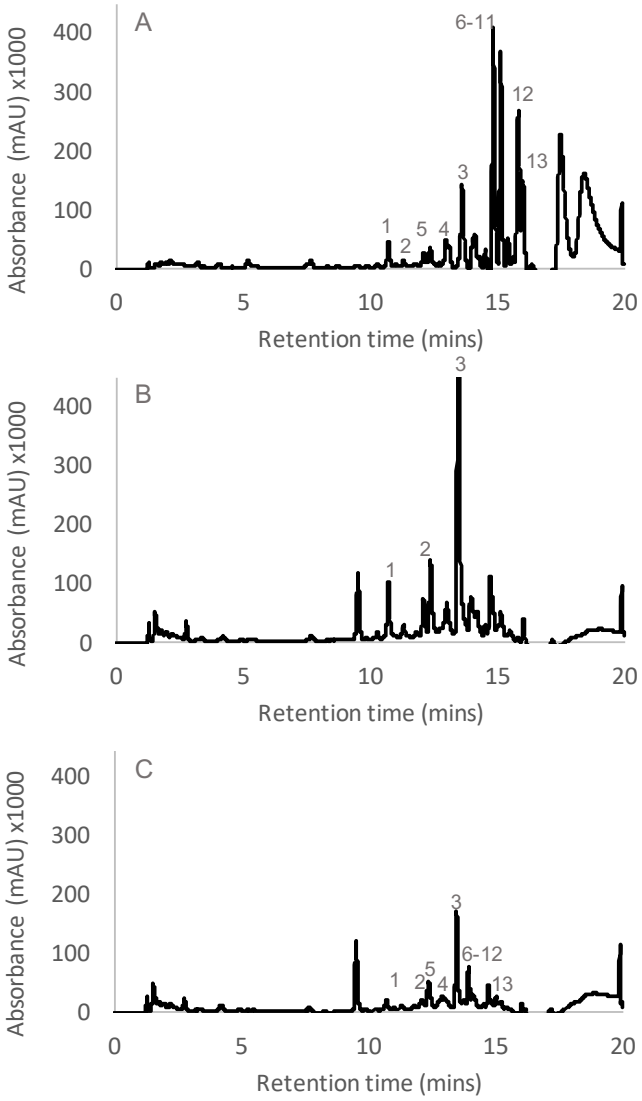


Figure 5.11: HPLC traces (UV 285 nm) of extracts from: (A) *S. formicae* wild-type; (B) *S. formicae* Δ*forW*; (C) *S. formicae* Δ*forW*: φBT1 *forW pforU*. Deletion of *forW* abolishes the production of formicamycins and results in an accumulation of fasamycins in the culture extracts. This implies that *forW* is essential for the conversion of a fasamycin to a formicamycin. Production of the formicamycins returns on complementation under the native promoter. HPLC(UV) LCMS analysis conducted by Dr Zhiwei Qin at the John Innes Centre.

Using these data, we can begin to make predictions about the methylation of the fasamycins and how this affects their conversion to a formicamycin. We hypothesise that the first methylation reaction to occur during formicamycin biosynthesis is di-methylation at C18 by the actions of either ForM or ForT, as all congeners from the formicamycin BGC contain this moiety and deletion of these enzymes results in a complete block on the pathway progression, implying it is an essential part of the early biosynthesis. A similar reaction is seen during the biosynthesis of benastatin, where the methyltransferase BenF catalyses the rare geminal *bis*-methylation to add a dimethyl group onto the polyketide chain (Xu, Schenk and Hertweck, 2007). BlastP analysis reveals that ForM shows low sequence identity to BenF, whereas ForT has 95% coverage and 49% similarity to BenF. Instead, ForM is more similar to RoqN (18.07% identity and 94% coverage) which has been shown to catalyse an O-methylation reaction during the biosynthesis of the alkaloid antimicrobial, meleagrin (Newmister *et al.*, 2018). Based on this information, we predict that ForT might be responsible for the addition of the di-methyl group to ring D of the formicamycin backbone. Further supporting this hypothesis, the di-methyl group is present on fasamycins A and B from cosAZ154; this BGC contains a homolog to ForT but is missing a significant match to ForM, implying ForM is not the enzyme responsible for this reaction. Instead we predict that ForM catalyses the addition of the O-methyl group at C3 at the base of ring A. This reaction is also predicted to be essential for progression of the formicamycin biosynthetic pathway in *S. formicae* as deletion of the gene encoding ForM resulted in no congeners being produced and all products from the pathway contain this moiety. However, the fasamycins isolated from *S. albus* carrying cosAZ154 do not have this methyl group and a homolog of ForM is missing from this BGC. It is possible that over the course of evolution, addition of the *forM* gene contributed to some sort of survival advantage and therefore became an essential component of the formicamycin BGC. During these experiments, we were not able to identify or isolate any congeners or biosynthetic intermediates that lacked either the dimethyl group of the O-methyl group at C3, implying that these molecules must not exist as enzyme-free intermediates (Qin *et al.*, 2017). We therefore predict that ForM and ForT accept substrates that are still bound to the

ACP and act before the intermediate is released, decarboxylated and cyclised, forming fasamycin C.

Following cyclisation and removal of the ACP, the next stage of formicamycin biosynthesis is predicted to be halogenation by ForV. The action of the halogenase generates the activated oxygen in the flavin cofactor for use by the Baeyer-Villiger monooxygenase, ForX, to expand ring C and form the seven membered lactone structure. These lactone intermediates contain the di-methyl group on ring D and the O-methylation at the base of ring A, however they do not contain the O-methylation on ring B that all of the formicamycin congeners do (**Figure 5.12**). This is therefore predicted to be an important stage of the conversion from a fasamycin to a formicamycin in addition to the two-step ring-expansion, ring-contraction mechanism described previously. It appears that the O-methylation of ring B is catalysed by ForW, as deletion of this enzyme blocked the pathway at fasamycin production and prevented the conversion of fasamycins to formicamycins. ForW is most similar to the O-methyltransferase LaPhzM which methylates the aromatic iodinin to form myxin, a phenazine antibiotic produced by *Lysobacter antibioticus* OH13. Interestingly, the methylated myxin has been shown to have a higher bioactivity than the iodinin precursor (Jiang *et al.*, 2018). Based on the data we have available, we predict that ForW acts after ForX, generating a highly reactive intermediate that is the substrate for reduction by ForY that enables the fasamycin precursor to be converted into the more bioactive formicamycin. Currently, we have been unable to isolate this intermediate for analysis, however, if it is highly reactive it is unlikely to be apparent in culture extracts. The formicamycin scaffold generated by the combined actions of ForX, ForW and ForY can then be halogenated at additional points on the carbon chain to produce congeners with further increased bioactivity. Some formicamycins can also be additionally O-methylated at R5. The data here do not reveal which methyltransferase is responsible for catalysing this reaction, however, some O-methyltransferases, including LaPhzM, have been shown to O-methylate at multiple positions on their aromatic substrates (Jiang *et al.*, 2018). Therefore, we predict that ForW also carries out this role, as it is the only

methyltransferase encoded in the formicamycin BGC that is thought to accept free, non-ACP bound substrates.

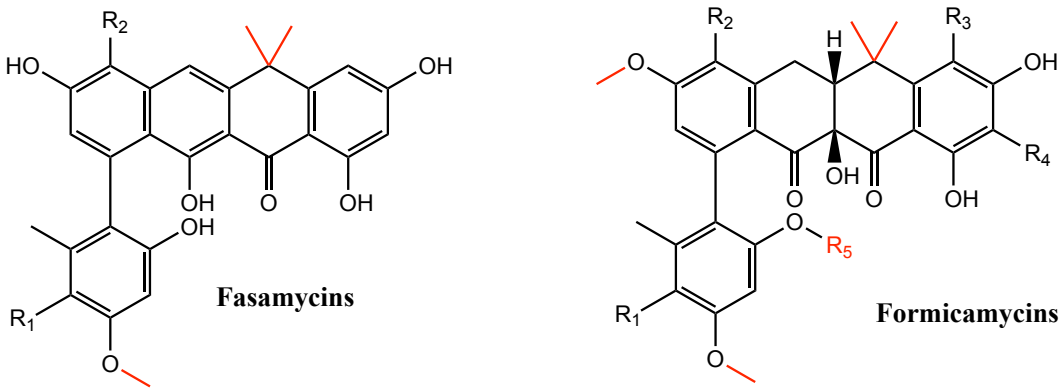


Figure 5.12: Compared to the fasamycins, the formicamycins display two additional points for methylation, highlighted here in red. Based on the data available, we predict that ForM and ForT are responsible for methylation at the sites common between fasamycins and formicamycins, with ForT catalysing the formation of the di-methyl group, and that these enzymes act early in the biosynthesis of the pentacyclic fasamycin scaffold. In contrast, ForW seems to be involved in the methylation of those sites specific to the formicamycins as deletion of *forW* has no effect on fasamycin biosynthesis.

5.4 Discussion

Previous work in this thesis indicated that fasamycins are the biosynthetic precursors of the formicamycins and that the formicamycins are more potent antibiotics. The aim of this chapter was to identify the biosynthetic and evolutionary link between these two distinct chemical scaffolds. We identified two other BGCs in the database that encode for fasamycins; cosAZ154 which has been shown to encode biosynthesis of fasamycin molecules similar to those produced by the *for* BGC and a putative BGC in *S. kanamyceticus* that remains to be experimentally characterised. No other BGCs in the database appear to encode for formicamycins, making the BGC in *S. formicae* unique. *Streptomyces* bacteria encounter many stresses in their natural environments and being non-motile, they have to deal with these stresses, often by the production of secondary metabolites. Most *Streptomyces* genomes contain a core region that encodes all the essential proteins such as those involved in DNA replication, transcription and translation, while the 'arms' of the chromosome often contain the BGCs encoding for non-essential secondary metabolites. At these

chromosome arms, significant exchange of genetic material between strains can occur via horizontal gene transfer, making the acquisition of advantageous genetic material common (Challis and Hopwood, 2003). We predict that the formicamycin BGC in *S. formicae* evolved from a fasamycin BGC via multiple insertion events as a result of the high rates of horizontal gene transfer and gene rearrangement that occur at the chromosome arms in these organisms. These genes were likely acquired from other soil-dwelling actinomycetes over the course of evolution and became permanent as they provided a significant competitive advantage by increasing the potency of the natural products encoded by the BGC.

The novel ring-expansion, ring-contraction mechanism of formicamycin biosynthesis described here appears to involve the activities of four enzymes, ForV, ForX, ForW and ForY, as all are essential for the conversion of fasamycins to formicamycins. Flavin dependent oxygenase enzymes like ForX and ForY are common in biosynthetic pathways and Baeyer-Villiger reactions are particularly prevalent in the biosynthesis of type 2 polyketides (Hertweck *et al.*, 2007). One of the most well studied Baeyer-Villiger oxygenase reactions is the conversion of the inactive precursor premithramycin B to the bioactive metabolite mithramycin by the atypical BVMO, MtmOIV in *S. argillaceus*. MtmOIV has been shown by crystallisation to noncovalently bind the FAD cofactor near the opening of the large substrate binding cavity designed for accepting large polyketide backbones (Beam *et al.*, 2009; Tolmie, Smit and Opperman, 2019). Although we have so far been unable to obtain pure ForX, it is likely that the structure and mechanism would be very similar to MtmOIV as it also falls into the same group of atypical BVMOs. Interestingly, cappable RNA sequencing results from **chapter 4** show that ForV, ForX, ForW and ForY are all expressed on the same transcript in *S. formicae* under the control of the promoter, *pforU*. Transcriptional coupling is often indicative of cooperative functionality in the encoded proteins and we predict that these enzymes all function together in the conversion of fasamycins to formicamycins, perhaps acting almost instantaneously, one after the other. Although many natural product biosynthesis pathways are generally viewed as linear routes from precursor to product, cooperative interactions between biosynthetic enzymes are actually quite common in the assembly of

macromolecules due to the dynamic nature of the cellular metabolism. Even outside of interactions between domains of multimeric proteins such as the ketosynthase, monomeric proteins can cooperate using ligands to aid these interactions (Porter and Miller, 2012). In the case of formicamycin biosynthesis, three of these four enzymes, ForV, ForX and ForY likely bind to flavin, so it is possible that this may be responsible for the functional coupling between these enzymes. ForV, ForX, ForW and ForY are predicted to function on cyclised substrates that have been released from the ACP. The enzymes involved in the cyclisation of the polyketide backbone during formicamycin biosynthesis are explored in the following chapter.

This work also led to the isolation of five new bioactive secondary metabolites from the formicamycin BGC and demonstrated that by genetically engineering a T2PKS gene cluster we can obtain novel, bioactive natural products. This method of genetically manipulating antibiotic producing organisms to produce 'unnatural natural products' has been around for some time and has been applied to multiple PKSs, such as the erythromycin biosynthesis pathway, to generate novel analogues of known compounds (Bérdy, 2005). The insight gained into the biosynthesis of these polyketide antibiotics will inform future drug discovery efforts, both in the natural product field and in synthetic chemistry. Furthermore, as a result of the experiments conducted in this chapter, we now have a dedicated fasamycin producing strain, *S. formicae ΔforX,* which will be useful for further biosynthetic studies and the isolation of high titres of fasamycin congeners for *in vitro* experiments, including bioassays and feeding studies. The nature of the biosynthetic pathway means that it is not possible to create a dedicated formicamycin producer that does not first make fasamycins, as they are the biosynthetic precursors, however, it may be possible to create enhanced production strains by combining the biosynthetic knowledge gained in this chapter with our knowledge of formicamycin regulation. This idea will also be explored further in the following chapter.

6 Genetic engineering of the formicamycin BGC to obtain novel, pre-defined products, by-products and biosynthetic intermediates

Traditionally, biosynthetic pathways are viewed as linear routes through pre-defined, enzymatically controlled biosynthetic steps that reach completion on production of a single defined end product. In reality, the biosynthesis of secondary metabolites in microorganisms is inherently more complicated. Variations in culture conditions of the microorganism, as well as the levels of starter units and cofactors available in the environment, will affect the activities of the biosynthetic enzymes and therefore the rates at which these reactions occur, resulting in different levels of pathway intermediates and biproducts accumulating at any one time (Sanchez and Demain, 2008). In addition, some of the proteins encoded by secondary metabolite BGCs are reliant on the presence or activity of other proteins from the biosynthetic pathway, like the well characterised KS_{α} and KS_{β} subunits, so if one is inhibited in some way, the other will also not function (Chen, Re and Burkart, 2018). Because of this complexity, spontaneous side reactions can occur, generating additional bi-products and shunt metabolites that differ from the main product of the pathway. Like many biosynthetic pathways, formicamycin biosynthesis in *Streptomyces formicae* is further complicated by the fact that it does not have a defined 'end product' and instead produces a complex mixture of both fasamycins and formicamycins. We have so far identified 16 fasamycin and formicamycin congeners that all originate from the formicamycin biosynthetic pathway, in addition to the six bioactive Baeyer-Villiger intermediates that accumulate on deletion of the oxidoreductase, ForY (**Chapter 5**). The aim of this chapter is to look for further congeners from the formicamycin biosynthetic pathway that may have been missed during the initial analysis of culture extracts due to the fact that production levels of biproducts and shunt metabolites are generally lower than the main pathway products. By identifying additional biosynthetic intermediates and shunt metabolites we aim to further characterize the biosynthetic pathway and define the exact steps needed for production of both fasamycins and formicamycins. This analysis may also enable us to identify further bioactive compounds from this diverse secondary metabolite

pathway. Until this point, the functions of the cyclase enzymes encoded by the formicamycin BGC have been predicted, but they will be fully investigated in this chapter. Furthermore, by combining our knowledge of formicamycin biosynthesis gained throughout this thesis with our knowledge of the gene cluster regulation, we aimed to produce multiple strains of *S. formicae* with more efficient and directed biosynthesis of specific formicamycin pathway products, intermediates and shunt metabolites. If successful, these strains could be used to overproduce specific bioactive compounds from the formicamycin BGC at a level that might be appropriate for industrial production of these clinically relevant metabolites.

6.1 Identifying further congeners from the formicamycin biosynthetic pathway

Analysis of ethyl acetate extracts of wild-type *S. formicae* cultures by LCMS results in an extremely complex dataset, with approximately 3000 unique ions present, including those representing the fasamycins and formicamycins. This, in combination with the aromatic, polycyclic nature of the formicamycins limits the effectiveness of tools such as the Global Natural Product Social Molecular Networking (GNPS) web-platform for the identification of novel congeners from the formicamycin biosynthetic pathway (Wang *et al.*, 2016). Dr Zhiwei Qin (JIC) therefore developed a de-replication method using mutants generated previously in this thesis to identify new halogenated ions in mass spectrometry (MS) data that originate from the formicamycin BGC. Profiling Solutions software (Shimadzu Corporation) was used to filter down the dataset by starting with the 3000 ions from the wild-type extracts, then removing mass spec data from blank media that was not inoculated with *S. formicae,* in order to remove ions present in the growth medium. Next, as most fasamycin and formicamycin congeners are chlorinated and only two other BGCs in the *S. formicae* genome contain putative halogenase encoding genes, it was hypothesised that any chlorine containing ions present in the dataset would likely derive from the formicamycin biosynthetic pathway. As such, the ions present in extracts of *S. formicae ⊿forV* (the halogenase deletion mutant) were then removed from the dataset to leave only ions of molecules that are halogenated by ForV. To further filter the results, the ions from extracts of *S. formicae ⊿for* (with the entire

formicamycin BGC deleted) were removed, leaving only halogenated compounds that originate from the formicamycin BGC. This filtered dataset was significantly less complex than the original unfiltered results, with around 200 unique ions present (Qin *et al.*, 2019).

6.1.1 6-chlorogenistein

Manual curation of the filtered dataset by Dr Zhiwei Qin showed that most of the remaining ions corresponded to the previously identified fasamycin and formicamycin congeners. Analysis of the ions in the dataset also revealed the presence a compound with the molecular mass of 304.0139 and a predicted chemical formula of $C_{15}H_9ClO_5$. Purification and subsequent NMR analysis of this molecule by Dr Zhiwei Qin showed it was a chlorinated derivative of the isoflavone, Genistein, specifically 6-chlorogenistein (**Figure 6.1**).

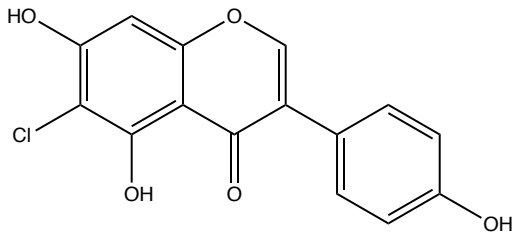


Figure 6.1: 6-chlorogenistein was isolated from culture extracts of wild-type *S. formicae* and the structure solved by NMR.

As discussed previously, halogenated natural products often display potent bioactivity. Previous studies have focussed on identifying chlorinated natural products by PCR screening microbes for conserved sequences present in genes encoding for FAD-dependent halogenases. This approach was used to identify the chlorinated Genistein molecules 3',8-dichlorogenistein and 8-chlorogenistein from *Actinoplanes* sp. HBDN08. These molecules were shown to display antioxidant and antitumor activities, however there was no evidence of antibacterial activity (Xiang *et al.*, 2010). The bioactivities of these other chlorinated genistein metabolites were reported at 25 μg/ml, therefore bioassays of the isolated 6-chlorogenistein from *S. formicae* were set up against *B. subtilis* at equivalent concentrations. As with other

studies, no antibacterial activity was seen from the 6-chlorogenistein isolated from the *S. formicae* extracts, even when the concentration was increased by 10 and 100 fold (data not shown). We therefore conclude that, like other chlorinated genistein metabolites, 6-chlorogenistein does not have any significant bioactivity against Gram-positive bacteria.

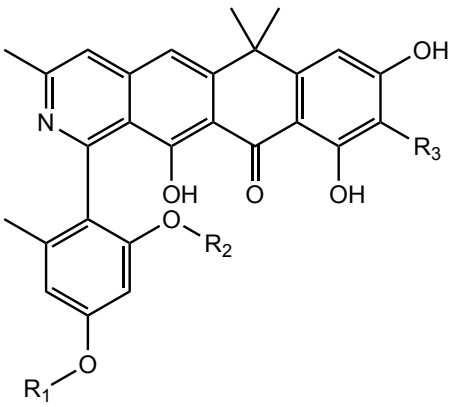
Based on structural analysis, we also assume that 6-chlorogenistein is not related to the formicamycins and does not directly originate from the formicamycin BGC. Genistein has been found in significant levels in soy beans and the above experiment conducted by Dr Zhiwei Qin was conducted on SFM agar (soy flour, mannitol agar) (**Table 2.1**) (Fukutake *et al.*, 1996). Therefore, it is assumed that ForV may have some promiscuous activity on this metabolite from the growth media, allowing chlorination to produce 6-chlorogenistein during fermentation. This further supports the hypothesis put forward in **Chapter 3.3** that the halogenase ForV can non-specifically halogenate multiple aromatic substrates. This broad substrate recognition is likely facilitated by the actions of a carrier protein that may also be encoded in the formicamycin BGC that brings a broad range of substrates to the halogenase. In this way, ForV recognises the carrier rather than the substrate, allowing it to act promiscuously on multiple substrates such as the isoflavone genistein and multiple positions on the fasamycin/formicamycin backbone. This hypothesis is explored further later in this chapter.

6.1.2 Formicapyridines; novel bi-products from the formicamycin biosynthesis pathway

Further analysis of the filtered mass spectra data by Dr Zhiwei Qin also resulted in the identification of three chlorinated metabolites (**21-23**) with a close structural relationship to the fasamycins and formicamycins but varying in the number of methyl groups present and surprisingly, the presence of a single nitrogen atom. By searching for equivalent ions lacking chlorination in the un-filtered dataset, three additional congeners similar to these new molecules were identified (**18-20**), bringing the total number of new compounds up to six. In addition, by repeating the fermentation of wild-type *S. formicae* cultures supplemented with 2mM sodium

bromide, which we have previously shown to induce biosynthesis of brominated formicamycin congeners, three additional bromine containing metabolites (**24-26**) were isolated with MS characteristics similar to those of the six new metabolites. The predicted molecular formulae from the mass spectrometry analysis suggested that these compounds represented novel structures. We therefore hypothesised that these nine new nitrogen-containing compounds represent a family of novel biosynthetic congeners from the formicamycin biosynthetic pathway that were likely missed during our initial analysis due to low production levels (Qin *et al.*, 2019).

As such, large scale cultures of wild-type *S. formicae* (14L made into approximately 450 agar plates) were grown, extracted with ethyl acetate and small amounts of the compounds were isolated for structural elucidation by Dr Zhewei Qin. Full analysis by NMR revealed the nine compounds were polyketide alkaloids with high structural similarity to the formicamycins, with the exception of the presence of a pyridine at ring B (**Figure 6.2**). We therefore named these new compounds the formicapyridines.



18 Formicapyridine A R_1=H, R_2=H, R_3=H
19 Formicapyridine B R_1=CH_3, R_2=H, R_3=H,
20 Formicapyridine C R_1=CH_3, R_2=CH_3, R_3=H,
21 Formicapyridine D R_1=H, R_2=H, R_3=Cl,
22 Formicapyridine E R_1=CH_3, R_2=H, R_3=Cl,
23 Formicapyridine F R_1=CH_3, R_2= CH_3, R_3=Cl,
24 Formicapyridine G R_1=H, R_2= H, R_3=Br,
25 Formicapyridine H R_1=CH_3, R_2= H, R_3=Br,
26 Formicapyridine I R_1=CH_3, R_2= CH_3, R_3=Br

Figure 6.2: De-replication of the mass spectra resulted in the identification of the formicapyridines; pyridine containing compounds with high structural similarity to the formicamycins with various levels of methylation that can be chlorinated or brominated on supplementation of the growth media.

Quantitative analysis of LCMS data from extracts of wild-type *S. formicae* using absorbances at 390 nm show that the six formicapyridines are present (no brominated congeners as the media was not supplemented with sodium bromide), but at approximately 100-fold lower levels than the fasamycins and formicamycins, which is why they were missed during the initial investigations. Further interrogation of LCMS data shows that the *S. formicae* Δ*for* cluster deletion mutant does not produce the formicapyridines, whereas production of all six metabolites returns in the Δ*for* mutant *in trans* complemented with the PAC (215-G). Similarly, *S. formicae* Δ*forV* accumulates the non-chlorinated formicapyridines A-C (**18-20**) and does not produce formicapyridines D-F (**21-23**), the chlorinated formicapyridines. Production of all formicapyridine congeners is restored by *in trans* complementation with *forV* under the control of *pforU* (see **Chapter 3.3**) (**Figure 6.3**). This confirms that the formicapyridines are products of the formicamycin biosynthetic pathway (Qin *et al.*, 2019).

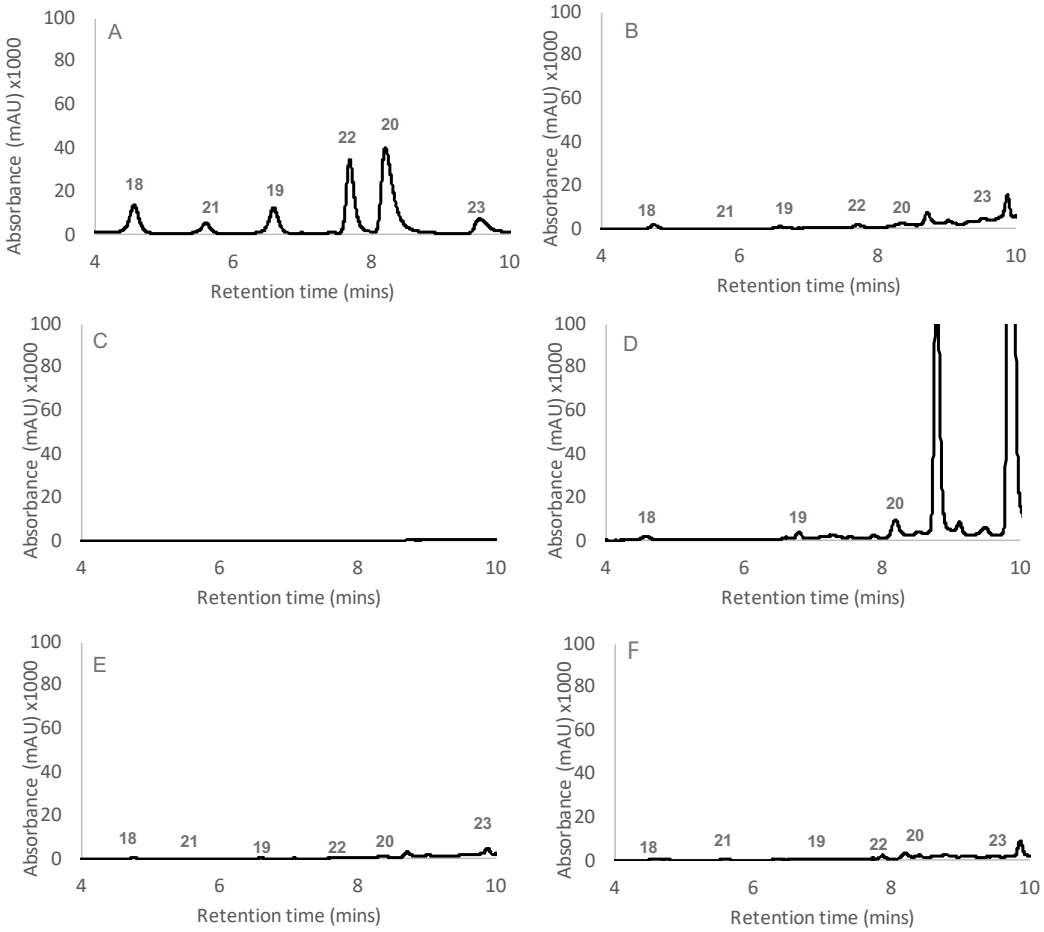


Figure 6.3: Reconstituted HPLC (UV 390 nm) showing (A) formicapyridine standards alongside culture extracts of (B) wild-type *S. formicae*; (C) *S. formicae* Δfor; (D) *S. formicae* Δ*forV;* (E) *S. formicae* Δfor: φC31 for 215-G; (F) *S. formicae* Δ*forV:* φBT1 *pforU forV*. Deletion of the formicamycin BGC abolished formicapyridine production and complementation with the PAC 215-G restores production at approximately wild-type levels. Deletion of the halogenase *forV* results in production and accumulation of only the non-chlorinated formicapyridines, with all six congeners returning on *in trans* complementation. Retention times are displayed between 4 and 10 minutes even though the method was run for 20 minutes as per previous experiments. This is because the antifungal compounds produced by *S. formicae* also absorb at UV 390nm and at much higher intensity than the formicapyridines, making visual identification challenging when viewing the whole dataset. Masses of all peaks have been confirmed by mass spectrometry to identify the compound. HPLC(UV) LCMS conducted by Dr Zhiwei Qin and the John Innes Centre.

To identify whether the formicapyridines display bioactivity against *B. subtilis,* disc diffusion assays were set up using 5 μg/ml of each compound, as this is the average concentration used for illustrating fasamycin and formicamycin bioactivity. No bioactivity was seen at this concentration, or at ten-fold higher concentration (50 μg/ml). Small zones of inhibition were seen at 100-fold (500 μg/ml) concentration, implying the MIC for these compounds is approximately 100-fold higher than that of

the fasamycin and formicamycins. To confirm this was not as a result of reduced diffusion from the paper disc due to altered polarity of the formicapyridines, assays were set up to test the growth of *B. subtilis* in liquid culture in the presence of compounds **18-23**. Small 1 ml cultures containing 50 μg/ml formicapyridine (or a hygromycin positive control) were inoculated in 12-well plates with *B. subtilis* in exponential phase and grown for 7 hours at 37°C, shaking at 220 rpm. After this incubation, the cultures were harvested and plated in serial dilutions in triplicate for colony count using the Miles and Misra protocol (Miles, Misra and Irwin, 1938). This showed that there was no significant difference in colony forming units (CFU) per ml (CFU/ml) between the cultures containing formicapyridines and the methanol solvent control, implying that the formicapyridines do not have significant bioactivity against Gram-positive pathogens (**Figure 6.4**).

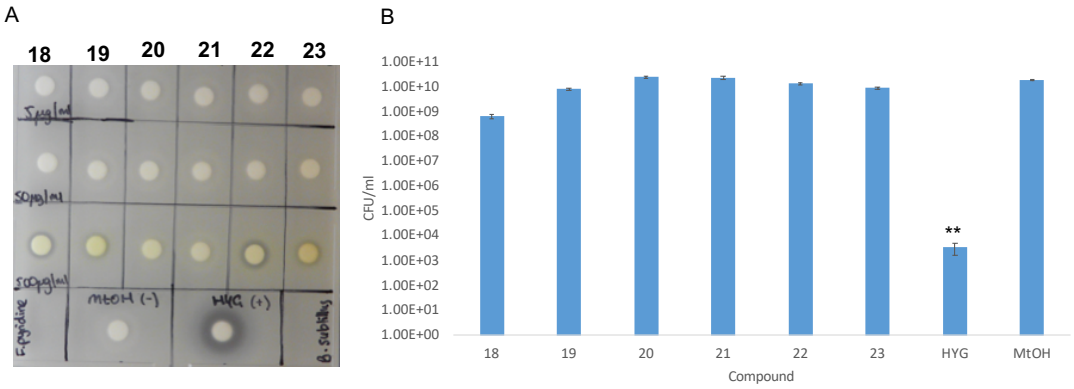


Figure 6.4: (A) The formicapyridines (**18-23**) were tested at 5, 50 and 500 μg/ml using disc diffusion assays against *Bacillus subtilis.* 20 μl of each solution is applied to each disc. No significant zone of inhibition was seen at 5 or 50 μg/ml, and only small zones were seen at 500 μg/ml. Hygromycin positive control is at 50 μg/ml. (B) In liquid culture, 50 μg/ml of each compound was added to 1 ml liquid culture of exponential phase *B. subtilis* in LB broth and grown at 37°C for 7 hours. Cells were harvested, serially diluted and plated for colony count to calculate the number of colony forming units present in the sample following Miles and Misra protocol (Miles, Misra and Irwin, 1938). There is no significant difference between the formicapyridine-containing cultures and the methanol (MtOH) solvent control, showing these compounds do not have significant antibacterial activity against Gram-positive pathogens. The only culture to have significantly lower growth than the negative control is the positive control with 50 μg/ml Hygromycin present (two tailed t-test, p = 0.005).

Furthermore, samples of formicapyridine A were sent to Inspiralis to test for both inhibition of topoisomerase IV relaxation and inhibition of gyrase supercoiling alongside the fasamycins and formicamycins. Assays were run using enzymes from both Gram-positive and Gram-negative bacteria as described previously (**Chapter 3.5**). In all assays, the IC50s values for the formicapyridine were higher than those

for fasamycin and formicamycin, confirming that these compounds have less potent bioactivity against the enzymes tested (**Table 6.1**). While it is possible that the pyridine ring changes the mechanism of action of these compounds, with the above bioassay data, we predict that alteration of ring B in this manner is enough to reduce the potency of these compounds, perhaps by changing the binding potential of these compounds to their target.

Table 6.1: Samples of formicapyridine A were tested by Isomerase therapeutics (NRP) for inhibitory activity against bacterial topoisomerase IV and DNA gyrase. In all assays conducted, the IC_{50} values for the formicapyridines were significantly higher than the tested fasamycin and formicamycin, implying that the potency of these compounds is reduced.

	Average IC_{50} (μM)			
	E. coli gyrase supercoiling	*S. aureus* gyrase supercoiling	*E. coli* topoisomerase IV relaxation	*S. aureus* topoisomerase IV relaxation
Ciprofloxacin	0.27	52.85	1.21	2.87
Fasamycin E	16.86	1.93	2.41	6.23
Formicamycin J	29.21	2.36	6.64	6.28
Formicapyridine A	>100	36.98	10.78	25-50

At this point, little is known about how the *for* BGC synthesises the formicapyridines. Aromatic nitrogen containing metabolites originating from T2PKS biosynthesis pathways are rare, with only a few examples having been reported to date (**Figure 6.5**). Jadomycins are polyketide natural products originating from *Streptomyces venezuelae* with nitrogen and oxygen containing heterocycles. These secondary metabolites display potent activity against both Gram-positive and Gram-negative bacteria as well as cytotoxicity against cancer cell lines. The mechanism of nitrogen incorporation into the jadomycin backbone remains unclear, but it is predicted to occur via the spontaneous, non-enzymatic incorporation of amino acids such as isoleucine during biosynthesis (Sharif and O'Doherty, 2012). The pyxidicyclines are anthracycline-related quinone like molecules with a nitrogen-containing tetracene structure originating from a T2PKS pathway from the myxobacterium *Pyxidicoccus*

fallax An d48. The pyxidicyclines have been shown to be moderate inhibitors of bacterial topoisomerase IV, with no significant activity against bacterial DNA gyrase, however, they were also shown to be potent inhibitors of human topoisomerase I and therefore have potential as novel anticancer agents. In the future, it would be interesting to screen the formicapyridines against mammalian topoisomerase enzymes and cancer cell lines to determine if they have any inhibitory activity. In contrast to the biosynthesis of the jadomycins, the mechanism of nitrogen incorporation in pyxidicyclines is thought to be via the attachment of a serine to the polyketide chain by a dedicated AMP-dependent synthetase/ligase and aromatase/cyclase di-domain protein, PcyJ, which then closes this ring with the help of two other cyclase enzymes, PcyK and PcyL (Panter *et al.*, 2018). At this point, it is unclear whether incorporation of the nitrogen and subsequent formation of the pyridine ring of the formicapyridines is spontaneous or enzymatically controlled, however no equivalent enzyme to PycJ is found within the formicamycin BGC, suggesting that biosynthesis of the formicapyridines in *S. formicae* may occur via a novel mechanism (Qin *et al.*, 2019).

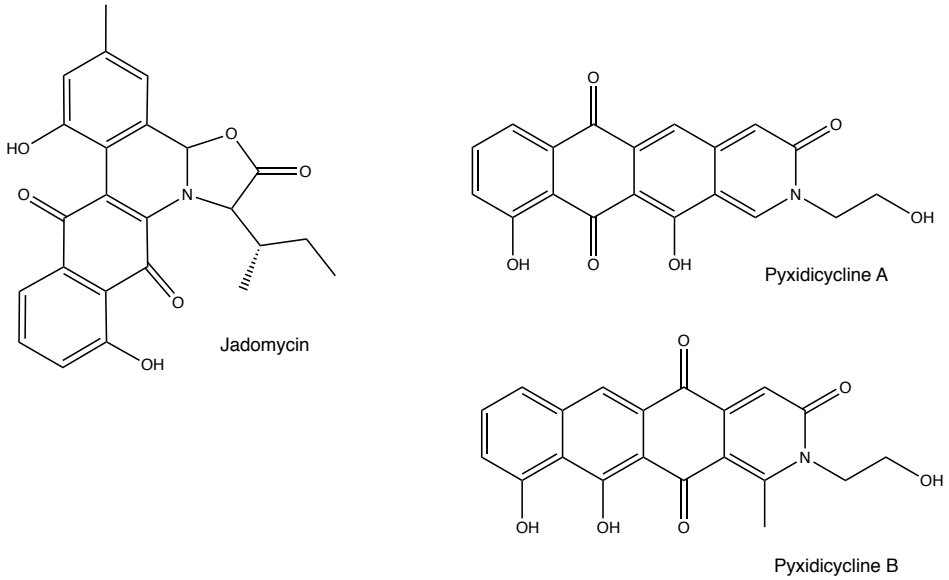


Figure 6.5: Structures of other natural products from T2PKS pathways that contain aromatic nitrogens. During the biosynthesis of both the jadomycins and the pyxidicyclines, amino acids are attached to the polyketide chain before cyclisation to allow nitrogen incorporation into the ring structures, either enzymatically or spontaneously. It is currently unclear how nitrogen is incorporated into the formicapyridine scaffold, although no enzyme candidates can be identified either within the BGC or near the cluster boundaries implying it may occur spontaneously during cyclisation.

6.2 The functions of the five putative cyclases in the biosynthesis of fasamycins, formicamycins and formicapyridines

The low titres of formicapyridines present in the wild-type extracts led us to hypothesise that these compounds are shunt metabolites produced spontaneously during the cyclisation of the poly-β-keto intermediate, specifically when cyclisation of ring B becomes derailed. To further investigate the biosynthesis of the formicapyridines, the roles of the cyclase enzymes encoded by the formicamycin BGC were considered. Sometimes, cyclisation of poly-β-keto intermediates in T2PKS biosynthesis is spontaneous, however, more often the process is controlled by dedicated cyclase enzymes or chaperone proteins that help direct the reaction to generate the desired final product. We hypothesised that generation of the pentacyclic formicamycin backbone from the poly-β-ketone tridecaketide (**17**) described previously requires the combined activities of five putative cyclases and antibiotic biosynthesis monooxygenases encoded in the formicamycin BGC; ForD, ForL, ForR, ForS and ForU (**Table 6.2**). These gene products are some of the few that remain uncharacterised in the formicamycin BGC at this point and therefore their function is currently unclear.

During fasamycin and formicamycin biosynthesis, we assume that ring A is closed first, followed by the sequential cyclisation and aromatisation of rings E, D, C and B. However, if this series of cyclisation steps is derailed before completion, specifically before complete cyclisation of ring B, a highly reactive β-ketoacid species would be generated, that might be prone to spontaneous decarboxylation and transamination. The *for* BGC does not encode for a transaminase/aminotransferase and a search of the *S. formicae* genome does not reveal any transaminase or aminotransferase gene candidates that might be involved in this reaction therefore we predict that the incorporation of nitrogen into the formicapyridine backbone may be spontaneous and/or as a result of the actions of endogenous enzymes from the cellular milieu. Once transaminated, the species could then theoretically undergo cyclisation, dehydration and oxidation to generate a formicapyridine (**Figure 6.6**).

Table 6.2: Bioinformatic analysis of each of the putative cyclases and antibiotic biosynthesis monooxygenases was conducted using BlastP. This analysis shows that these five genes show similarity to other polyketide cyclases and antibiotic biosynthesis monooxygenases in the database and therefore they are predicted to be involved in aromatisation of the poly-β-ketone intermediate during formicamycin biosynthesis. This table also appears in (Qin *et al.*, 2019).

Name	**Protein ID**	**Top BlastP hit**	**Annotation**	**% coverage**	**% amino acid identity**
ForD	WP_098245758.1	WP_003962252.1	Polyketide cyclase (*TcmN*), *Streptomyces clavuligerus* ATCC 27064	95	73
ForL	WP_055544278.1	WP_076971949.1	*TcmI* family type 2 polyketide cyclase, *Streptomyces sparsogenes*	97	52
ForR	WP_098245762.1	AUI41024.1	Polyketide WhiE II *Streptomyces* sp. (*abxR*), cupin domain	96	75
ForS	WP_098245763	SDF47214.1	Antibiotic biosynthesis monooxygenase, *Lechevalieria fradiae*	96	48
ForU	WP_098245765.1	PZS28546.1	Antibiotic biosynthesis monooxygenase, *Pseudonocardiales* bacteria	84	46

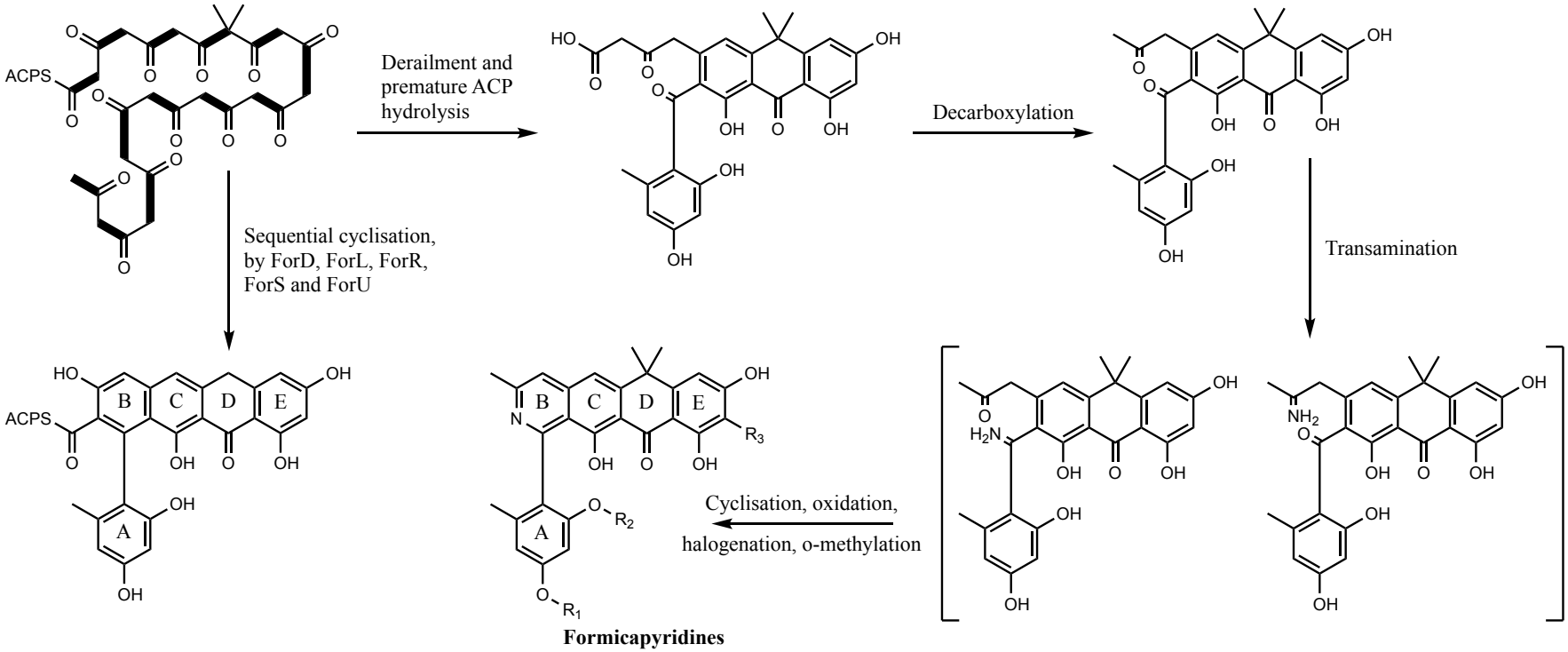


Figure 6.6: We hypothesise that cyclisation of the poly-β-ketone tridecaketide to form a fasamycin backbone requires the activity of some/all of the five putative cyclases and antibiotic biosynthesis monooxygenase enzymes encoded in the formicamycin BGC; ForD, ForL, ForR, ForS and ForU. If this sequential cyclisation is disrupted, generating a highly reactive β-ketoacid species, spontaneous decarboxylation and transamination may form the formicapyridine backbone as shown. Adapted from (Qin *et al.*, 2019).

We aimed to characterise the roles of the putative cyclase and antibiotic biosynthesis monooxygenase enzymes encoded in the formicamycin BGC and specifically to identify which of the five cyclase enzymes were responsible for closing ring B in order to experimentally characterise formicapyridine biosynthesis. In addition, by identifying a single gene product that controls production of the formicapyridines, specifically cyclisation of ring B, it may be possible to engineer a mutant strain of *S. formicae* with enhanced levels of formicapyridine production and possibly even abolish fasamycin and formicamycin biosynthesis to generate a specific formicapyridine producer. To determine the roles of these enzymes in fasamycin, formicamycin and formicapyridine biosynthesis, CRISPR-Cas9 was used to generate in frame, unmarked deletion mutations in each gene and examine the effect of these mutations on fasamycin, formicamycin and formicapyridine biosynthesis.

Bioinformatic analysis of ForD shows homology to multiple polyketide cyclase/dehydratase enzymes including the N-terminal aromatase/cyclase domain of the multifunctional tetracenomycin protein TcmN. The crystal structure of TcmN reveals an interior binding pocket for binding nascent polyketide intermediates and shows it plays a central role in the cyclisation and aromatisation of tetracenomycin, particularly catalysing the sequential cyclisation and aromatisation of rings 1 and 2 via multiple dehydration reactions (Ames *et al.*, 2008). We therefore hypothesise that ForD might catalyse the cyclisation of rings E and D on the formicamycin backbone. To experimentally characterise the role of ForD in the biosynthesis of fasamycins, formicamycins and formicapyridines, three biological replicates of *S. formicae ΔforD* were cultivated under formicamycin producing conditions. Culture extracts were analysed by HPLC(UV) and LCMS by Dr Zhiwei Qin at the JIC. Deletion of *forD* completely abolished biosynthesis of fasamycins, formicamycins and formicapyridines with no new shunt metabolites or biosynthetic intermediates evident in the culture extract. Genetic complementation of *forD in trans* under the control of the native promoter (*pforT*) reinstated production of all congeners from the pathway (**Figure 6.7**). This suggests that ForD plays a central role in the biosynthesis of all metabolites originating from the formicamycin biosynthetic pathway.

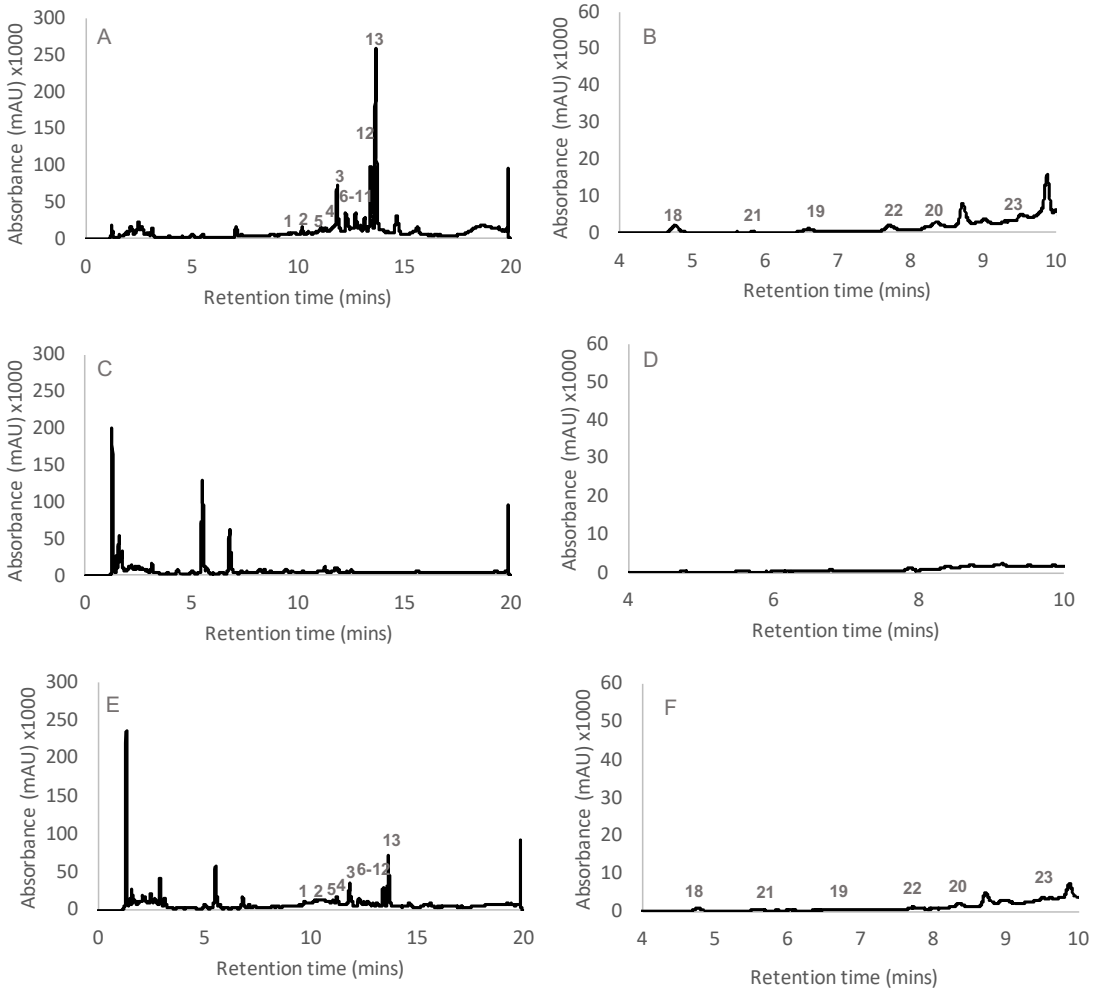


Figure 6.7: Reconstituted HPLC-UV at 280 nm to capture fasamycins and formicamycins (**1-13**) and 390 nm to capture formicapyridines (**18-23**) in culture extracts of (A and B) wild-type *S. formicae* at 280nm and 390 nm respectively; (C and D) *S. formicae* Δ*forD* at 280nm and 390 nm respectively; (E and F) *S. formicae* Δ*forD*: φBT1 *forD pforT* at 280nm and 390 nm respectively. n=3 replicates of each mutant and complementation strain. Deletion of the putative cyclase *forD* abolished fasamycin, formicamycin and formicapyridine biosynthesis. Production is reinstated on *in trans* complementation under the native promoter. Figures are scaled comparatively within the same wavelengths. Peak intensity cannot be compared between different wavelengths. HPLC (UV) LCMS conducted by Dr Zhiwei Qin, JIC.

ForL belongs to the polyketide synthase cyclase family of enzymes and shows closest homology to the polyketide cyclase TcmI from the tetracenomycin biosynthetic pathway. TcmI has been shown to catalyse the closure of the final ring of tetracenomycin F2 to produce tetracenomycin F1 (Thompson *et al.*, 2004). This reaction is remarkably similar to the predicted reaction needed to cyclise ring B, therefore we predict that ForL may be the enzyme responsible for catalysing cyclisation of ring B. If this were true, we might expect that a deletion mutant of *forL*

may only produce formicapyridines, due to derailed formicamycin biosynthesis as a result of no cyclisation of ring B. However, HPLC(UV) and LCMS analysis of culture extracts of *S. formicae ΔforL* by Dr Zhiwei Qin showed no fasamycins, formicamycins or formicapyridines are produced and no other biosynthetic intermediates or shunt metabolites are evident. Genetic complementation of *forL in trans* under the control of the native promoter (*pforM*) reinstated production of all congeners from the pathway, implying that ForL also plays an essential role in the cyclisation of the poly-β-ketone tridecaketide (**Figure 6.8**).

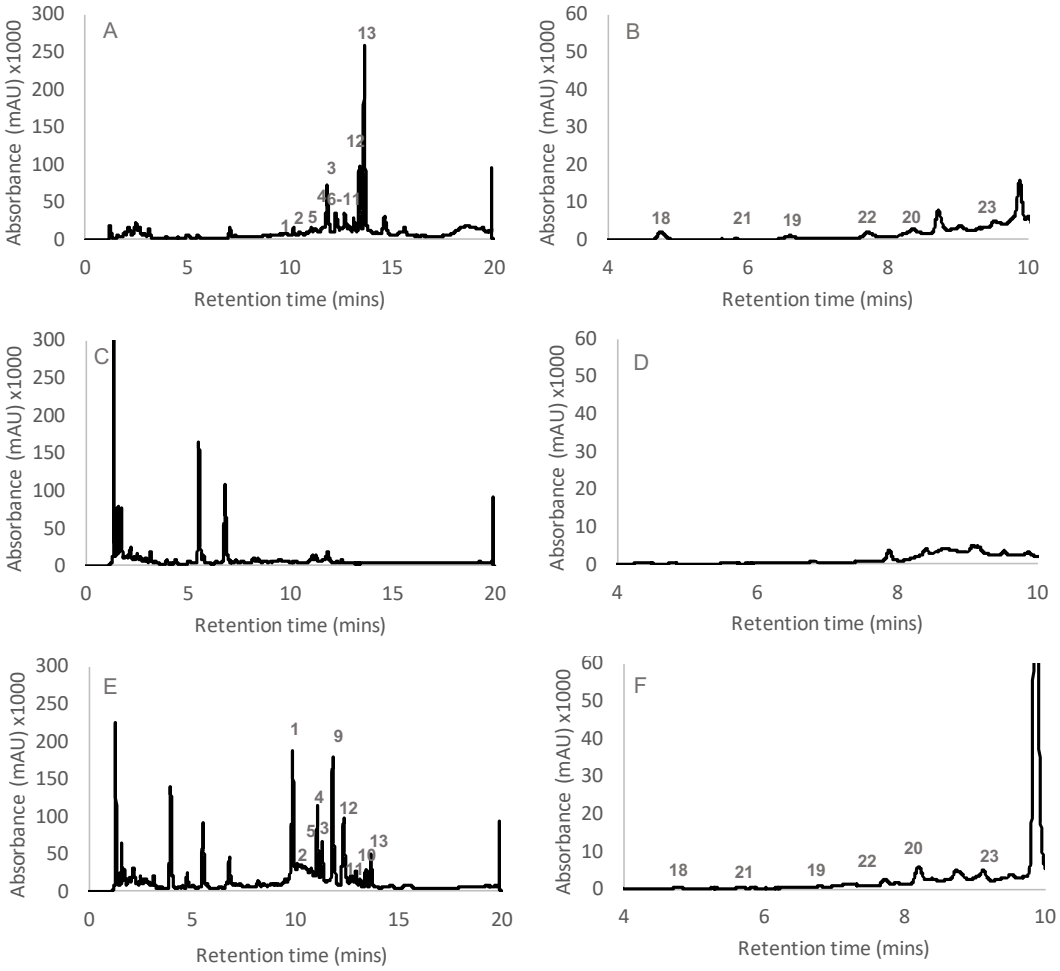


Figure 6.8: Reconstituted HPLC-UV at 280 nm to capture fasamycins and formicamycins (**1-13**) and 390 nm to capture formicapyridines (**18-23**) in culture extracts of (A and B) wild-type *S. formicae* at 280nm and 390 nm respectively; (C and D) *S. formicae* Δ*forL* at 280nm and 390 nm respectively; (E and F) *S. formicae* Δ*forL*: φBT1 *forL pforM* at 280nm and 390 nm respectively. n=3 replicates of each mutant and complementation strain. *S. formicae* Δ*forL* does not produce any fasamycin, formicamycin or formicapyridine congeners. Production is reinstated on *in trans* complementation under the native promoter. HPLC (UV) LCMS conducted by Dr Zhiwei Qin.

ForR is a member of the cupin-like superfamily with close homology to the polyketide cyclase RemF involved in the cyclisation of the pentacyclic polyketide core of resistomycin in *Streptomyces resistomycificus* (Silvennoinen, Sandalova and Schneider, 2009). When grown under formicamycin producing conditions, HPLC(UV) and LCMS analysis of culture extracts of *S. formicae ΔforR* show no fasamycins, formicamycins, formicapyridines or related ions showing it also likely plays an important role in the cyclisation of the formicamycin core. Genetic complementation of *forR in trans* under the control of the native promoter (*pforT*) was unsuccessful, resulting in HPLC (UV) profiles that looked the same as traces for *S. formicae ΔforR.* The complementation was therefore repeated using a construct with *forR* placed under the control of the constitutive and high-level ermE* promoter. When the new complementation mutant was grown under formicamycin producing conditions and analysed by HPLC (UV) LCMS, all formicamycin congeners returned, but at lower levels than is seen in the wild type culture extracts (**Figure 6.9**). It is unclear why complementation of *forR* under the control of *pforT* was unsuccessful in this case, as the *pforT* promoter has been used to complement other deletion mutations in genes under its control during this work. When complementing genes *in trans* on integrative vectors such as pMS82, the new location of the gene can be important in determining the levels of transcription, so it may be that in this case, transcription levels of *forR* were too low under *pforT* but the stronger promoter *permE** generated enough transcript for complementation. Nevertheless, this shows that all three putative cyclases encoded within the formicamycin biosynthesis pathway (ForD, ForL and ForR) are essential for the biosynthesis of any metabolites from the formicamycin pathway. It may be that these enzymes all function as a multi enzyme complex and that removal of one cyclase de-stabilises the complex, resulting in no activity by the other two enzymes. This is explored further later in this chapter.

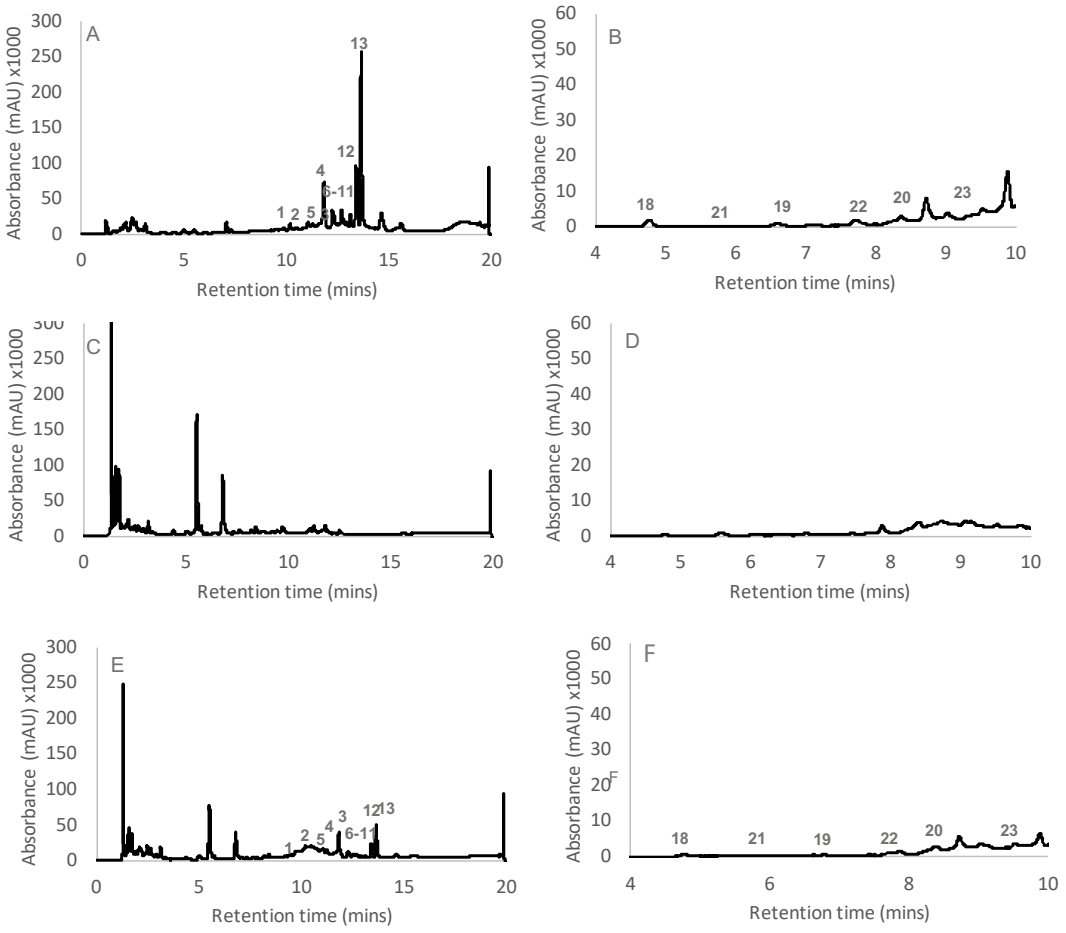


Figure 6.9: Reconstituted HPLC-UV at 280 nm to capture fasamycins and formicamycins (**1-13**) and 390 nm to capture formicapyridines (**18-23**) in culture extracts of (A and B) wild-type *S. formicae* at 280nm and 390 nm respectively; (C and D) *S. formicae* Δ*forR* at 280nm and 390 nm respectively; (E and F) *S. formicae* Δ*forR*: φBT1 *forRA permE** at 280nm and 390 nm respectively. n=3 replicates of each mutant and complementation strain. Deletion of *forR* abolishes formicamycin, fasamycin and formicapyridine production. Complementation under the constitutive promoter permE* reinstated production of all congeners, but at lower levels than the wild-type strain. HPLC (UV) LCMS conducted by Dr Zhiwei Qin.

ForS is an antibiotic biosynthesis monooxygenase (ABM) protein. Members of this protein superfamily are involved in the biosynthesis of several antibiotics by *Streptomyces* species and often function as multimers, either with each other or with proteins from other superfamilies (Sciara *et al.*, 2003). Replicates of *S. formicae ΔforS* were grown under formicamycin producing conditions and culture extracts were analysed by HPLC(UV) and LCMS by Dr Zhiwei Qin. Interestingly, the levels of formicamycin production in *S. formicae ΔforS* were reduced to about a third of the levels produced by the wild-type strain. In contrast, levels of formicapyridines in the culture extracts increased by approximately 25 times in the mutant.

Complementation of *forS in trans* under the control of *pforT* was unsuccessful, however, *in trans* complementation of *forS* under the control of *permE** resulted in partial complementation, with production levels of formicamycin congeners still being less than in the wild-type strain, but formicapyridine production being reduced to wild-type levels once again (**Figure 6.10**).

The increase in formicapyridine titres in *S. formicae ΔforS* suggests that ForS plays an important role during the closure of ring B. However, formicamycins are still present in the extracts, albeit at lower levels than in the wild-type extracts, demonstrating that cyclisation of ring B can still occur without ForS and that the enzyme is not solely responsible for catalysing the reaction. Instead, it may be that ForS functions as a chaperone that supports the other cyclase enzymes during the aromatisation of the polyketide backbone and that without it, the cyclisation of the polyketide is less efficient and more shunt metabolites are produced. Interestingly, the levels of fasamycins present is approximately equal in both the wild-type and the Δ*forS* extract samples. This could suggest that ForS is somehow involved in the conversion of a fasamycin to a formicamycin, as fasamycin biosynthesis is unaffected but formicamycin biosynthesis is significantly reduced (**Figure 6.11**).

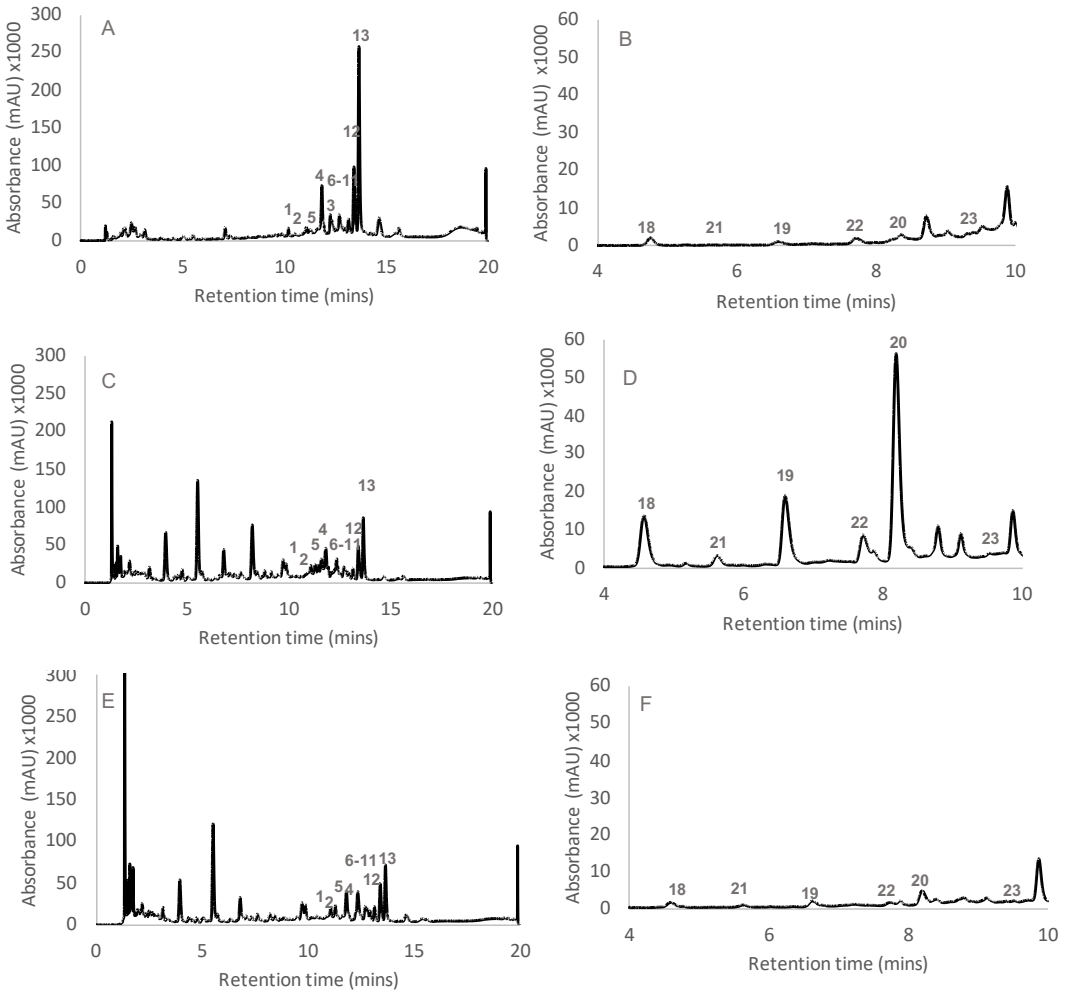


Figure 6.10: Reconstituted HPLC-UV at 280 nm to capture fasamycins and formicamycins (**1-13**) and 390 nm to capture formicapyridines (**18-23**) in culture extracts of (A and B) wild-type *S. formicae* at 280nm and 390 nm respectively; (C and D) *S. formicae* Δ*forS* at 280nm and 390 nm respectively; (E and F) *S. formicae* Δ*forS*: φBT1 *forS permE** at 280nm and 390 nm respectively. n=3 replicates of each mutant and complementation strain. Deletion of *forS* reduces fasamycin and formicamycin titres to approximately one-third of the wild-type levels. In contrast, formicapyridines are produced at 25-times higher titres in *S. formicae* Δ*forS.* Complementation under the constitutive promoter permE* returned production of all congeners to approximately wild-type levels. HPLC (UV) LCMS conducted by Dr Zhiwei Qin.

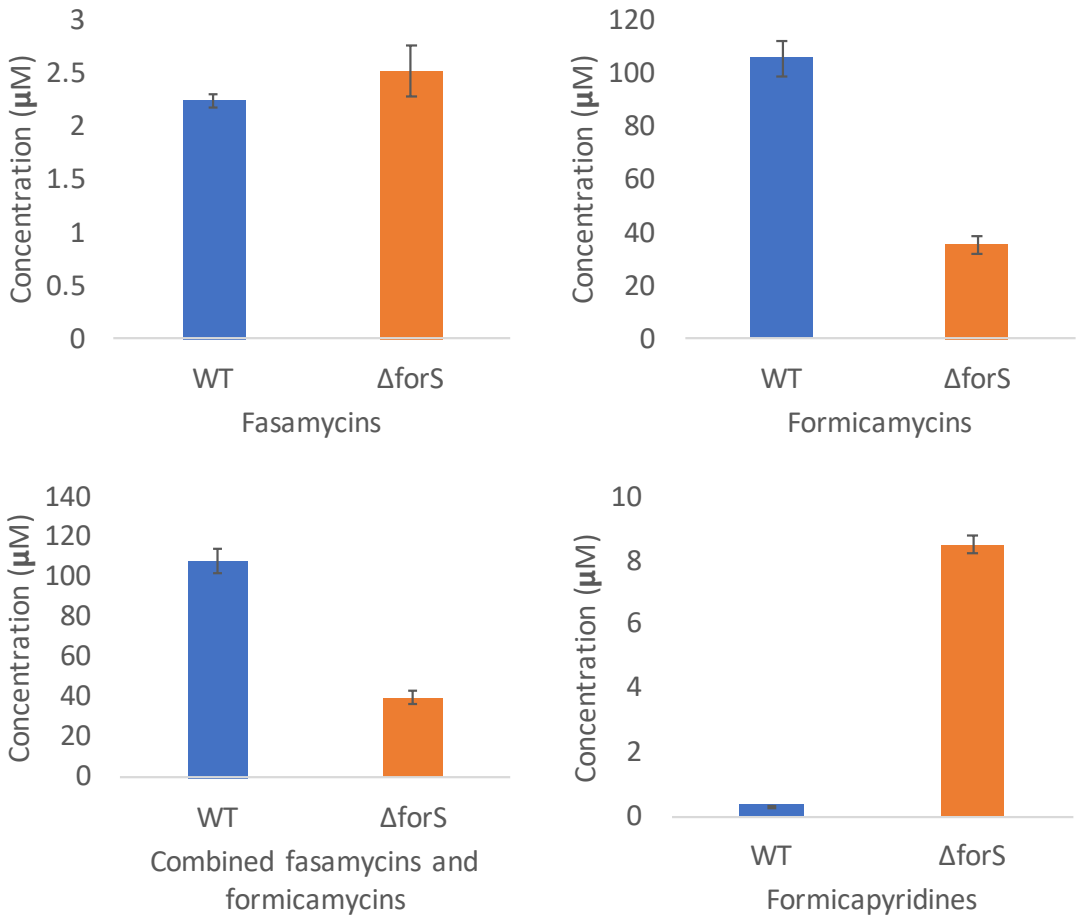


Figure 6.11: Titres of compounds present in the fermentation extracts were measured by quantitative integration of HPLC(UV) and mass ion data using calibrations based on standards to calculate the relative amounts of each compound in the extracts of the wild-type (WT) and the *S. formicae* Δ*forS* (Δ*forS*) mutant. The *forS* deletion mutant produces less formicamycins than the wild-type but more formicapyridines. Error bars represent standard error of the mean. Data generated by Dr Zhiwei Qin for (Qin *et al.*, 2019). Figure adapted and reproduced with permission.

During further analysis of *S. formicae* Δ*forS*, a new minor congener was identified that is not present in any wild-type fermentations. Growth of *S. formicae* Δ*forS* was therefore scaled up and 3.4 mg of the new compound was isolated for structural elucidation by NMR by Dr Zhiwei Qin. This new metabolite is a new fasamycin congener that is not halogenated but contains a carboxyl group on ring B (**Figure 6.12**). The new fasamycin congener, named fasamycin F (**27**), appears to be a cyclised polyketide intermediate that has been hydrolysed from the ACP but has not undergone subsequent decarboxylation to form the fasamycin scaffold. This further suggests that ForS is involved in the final step of cyclisation of the fasamycin

backbone and although not essential, supports the role of other enzymes involved in modifying ring B (Qin *et al.*, 2019).

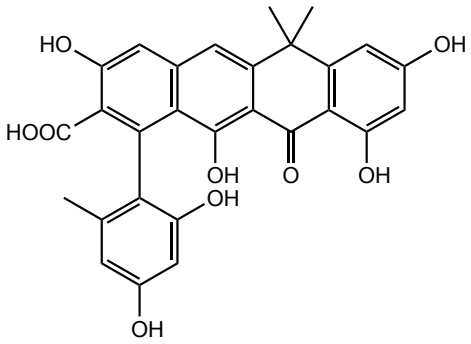


Figure 6.12: Structure of the new fasamycin congener, **fasamycin F**, isolated from *S. formicae ΔforS.* Fasamycin F is the only compound isolated from the *for* BGC that contains the carboxylic acid group on ring B.

ForU is also a member of the ABM protein superfamily and shows the greatest homology to the monooxygenase ActVA-orf6 from the actinorhodin biosynthetic pathway in *S. coelicolor*, a protein that itself has homology to TcmI, the closest relative of ForL (Sciara *et al.*, 2003; Thompson *et al.*, 2004). Extracts of *S. formicae ΔforU* were cultivated under formicamycin producing conditions and analysed by HPLC(UV) and LCMS by Dr Zhiwei Qin, JIC. Once again, no fasamycin, formicamycin, formicapyridine or related ions were present in the resulting dataset. This implies that ForU also plays an important role in the cyclisation of the core carbon backbone during formicamycin biosynthesis. Genetic complementation of *forU in trans* under the control of the native promoter resulted in only the production of low levels of the non-halogenated fasamycin, formicamycin and formicapyridine congeners (data not shown). When designing pCRISPomyces-2 constructs, care is taken not to delete regions of neighbouring genes where the coding regions overlap. For example, there is a nine base pair overlap between *forU* and *forV*, therefore this region was not included in the deletion of *forU*. Interestingly, *in trans* complementation of *S. formicae ΔforU* with a construct containing a fusion of the coding regions for both *forU* and *forV* to *permE** does restore production of all fasamycins, formicamycins and formicapyridines, including the chlorinated congeners (**Figure 6.13**). This could imply that deletion of *forU* results in a polar effect of the transcription of downstream

forV, perhaps due to the presence of a ribosome binding site (RBS) for *forV* within the coding region of *forU.* Alternatively, it could be that as well as performing a role in the cyclisation of the poly-β-ketone tridecaketide, ForU has a role in allowing activity of ForV, perhaps by acting as the carrier for substrates to the active site. In this case, the activity of ForV would be sensitive to the levels of ForU present in the cell. When *S. formicae ΔforV* was complemented *in trans* with *forV* under the control of *pforU*, expression levels of *forU* were unaffected, as it remained in its native locus, therefore any transcribed *forV* from the complementation construct would be functional due to the presence of ForU. However, when *S. formicae ΔforU* was complemented with *forU* under the native promoter, some ForU is produced, as a fully cyclised fasamycin 1 is seen in the culture extract, but perhaps expression levels were lower, as seen in the complementation of other genetic mutants in this project. In this case, transcription of *forV* would be unaffected, but the resulting protein may not be functional, or may only function at low levels.

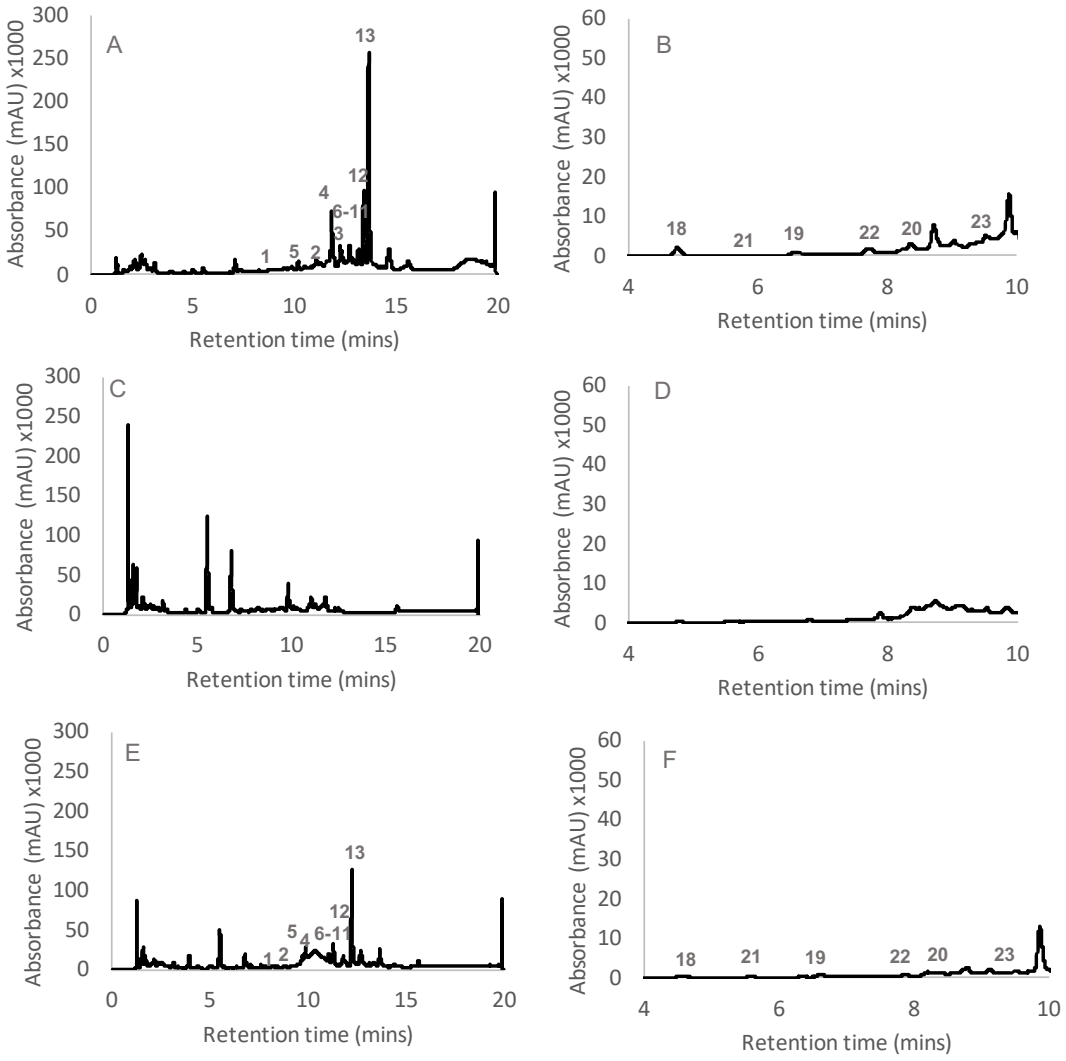


Figure 6.13: Reconstituted HPLC-UV at 280 nm to capture fasamycins and formicamycins (**1-13**) and 390 nm to capture formicapyridines (**18-23**) in culture extracts of (A and B) wild-type *S. formicae* at 280nm and 390 nm respectively; (C and D) *S. formicae* Δ*forU* at 280nm and 390 nm respectively; (E and F) *S. formicae* Δ*forU*: φBT1 *forU-V permE** at 280nm and 390 nm respectively. n=3 replicates of each mutant and complementation strain. Deletion of *forU* abolishes fasamycin, formicamycin and formicapyridine production. Complementation under the native promoter was unsuccessful, but complementation with a construct containing *forU* and downstream *forV* under the control of *permE** restored fasamycin formicamycin and formicapyridine production. HPLC (UV) LCMS conducted by Dr Zhiwei Qin.

These results show that ForD, L, R and U are required for the production of all the pathway derived metabolites and therefore likely play major roles in the cyclisation of the pentacyclic formicamycin backbone. In addition, removal of any one gene individually abolishes fasamycin, formicamycin and formicapyridine production. This implies that these cyclase enzymes work together in a large, multi-enzyme complex to cyclise the poly-β-keto intermediate, rather than as independent proteins. In

contrast, ForS is not essential for the biosynthesis of any congeners from the formicamycin biosynthetic pathway, therefore, ForS does not have an essential role in cyclisation of the carbon backbone. However, ForS does appear to play an important role in the overall productivity of the formicamycin biosynthesis pathway. Deletion of *forS* resulted in less formicamycin biosynthesis and increased levels of shunt metabolites like the formicapyridines. It also affected the decarboxylation of ring B after cleavage of the ACP, as shown by the isolation of the new fasamycin congener, fasamycin F. Therefore, it may be that ForS stabilises the multi-enzyme cyclase complex for optimal formicamycin production and/or ForS may act as a chaperone to modulate the biosynthetic pathway and direct nascent intermediates for complete cyclisation. In this way, ForS ensures that most of the poly-β-keto intermediate is fully cyclised and becomes a fasamycin and, ultimately, a formicamycin, with only low levels of formicapyridines and other shunt metabolites being produced as side products. Furthermore, ForS may be involved in chaperoning the cyclised intermediate to other enzymes further downstream in the pathway for conversion into formicamycins. This may explain why *S. formicae* Δ*forS* makes wild-type levels of fasamycins but significantly less formicamycins.

Similarly, it is possible that ForU has a role in the activity of the halogenase ForV, perhaps by acting as a carrier or by tethering ForV to the cyclase complex in order to guide the cyclised intermediate to the halogenase for chlorination. Evolutionary analysis of fasamycin and formicamycin BGCs suggests that ForU and ForV are always encoded on the same transcript, even on cosAZ154 which shows significant rearrangement of other genes compared to the formicamycin BGC in *S. formicae,* indicating they may be functionally linked (**Chapter 5.1**). If ForU does stabilise the interaction between the cyclase enzymes and the halogenase, removal of ForU would not only affect halogenation but would prevent the activity of the cyclases ForD, ForL and ForR, explaining the loss of any fasamycins and formicamycins in the extract of the deletion mutant. In future, it would be interesting to co-express and purify ForU and ForV together to determine whether *in vitro* halogenation activity is dependent on the presence of ForU.

Overall, this experiment highlights the importance of ABM domain containing proteins, especially ForS, for the productivity of the formicamycin biosynthesis pathway. Genes encoding for ABM domains are encoded in many PKS pathways and they have a variety of biological functions. Based on these data, we suggest one of their functions may be to fine-tune PKS pathways by stabilising multi-enzyme complexes or guiding intermediates through various stages of biosynthesis to generate the desired pathway products with high efficiency. Similar pathways to the formicamycin BGC that also encode for ABM family proteins include the xantholipin biosynthesis pathway in *Streptomyces flavogriseus* and the pradimicin biosynthetic pathway in *Actinomadura hibisca* P157-2 (**Figure 6.14**). Xantholipin is a hexacyclic aromatic polyketide with antibacterial activity against Gram-positive pathogens in addition to significant anticancer activity. The BGC encodes for three putative cyclases and four ABM family proteins. The predicted biosynthesis pathway only requires a single monooxygenase although one of the ABM family monooxygenases is predicted to assist the Baeyer-Villiger monooxygenase during the oxidation to produce the xanthone ring (Zhang *et al.*, 2012). Currently there is no predicted role for the other ABM family proteins encoded in this pathway and it is possible that they are chaperones that support the activities of the encoded cyclases. Similarly, the pradimicins are pentacyclic aromatic polyketides with some structural similarities to the formicamycins. The BGC encoding for pradimicins contains three putative cyclases and two putative ABM domain proteins, PdmH and PdmI. Heterologous expression of these genes has shown that in addition to all three cyclases, the ABM domain monooxygenase PdmH must also be expressed for complete pradimicin biosynthesis, as it is required for the formation of the quinone moiety. Experimental characterisation has shown that the three cyclases function as a multi-enzyme complex and the authors of the study propose that the complex is able to span the entire length of the uncyclized polyketide to 'lock the backbone in place' and prevent undesired spontaneous cyclisation reactions. Furthermore, co-expression of PdmH with the core biosynthetic machinery changes the predominant product of the pradimicin biosynthetic pathway and decreases the levels of production of the major congeners to less than 10% of the normal levels (Zhan, Watanabe and Tang, 2008). This suggests that the cyclase enzymes from the pradimicin biosynthetic pathway

also require the activity of the ABM family protein PdmH to chaperone or fine-tune the biosynthesis in the same way as ForS in the biosynthesis of the formicamycins.

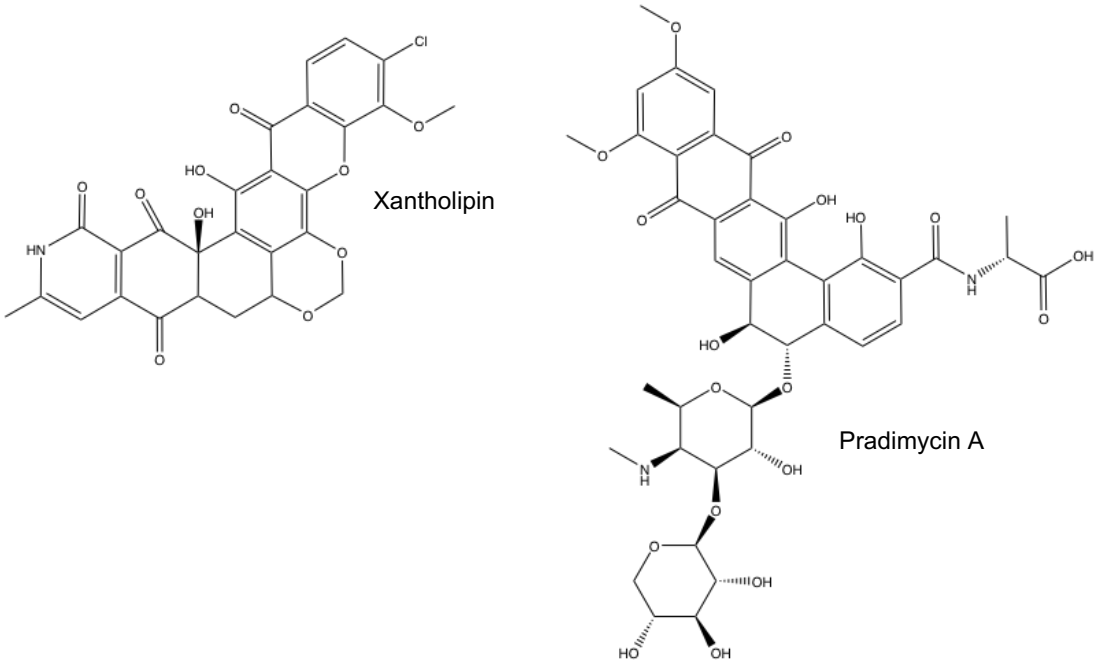


Figure 6.14: Structures of the polyketides xantholipin and pradimycin which also have ABM domain containing proteins encoded within their BGCs. Is it possible that some of the ABM domain proteins involved in the biosynthesis of these molecules act as chaperones to guide the cyclisation of the multiple rings as ForS does during the biosynthesis of the formicamycins. Structures downloaded from PubChem and adapted in ChemDraw for figure.

Furthermore, many PKS pathways encode for PKS cyclases with an ABM domain fused to the end of the protein, such as the polyketide synthase CurD from *Streptomyces ambofaciens* (WP_053126740.1) which contains an ABM domain at its N-terminus. If ABM domains do indeed act as chaperones that support cyclisation reactions during polyketide biosynthesis, it is possible that these genes may represent examples of highly evolved systems where these chaperones have become an essential part of the core essential biosynthetic machinery and are now transcribed as a single gene product with the cyclase, further highlighting their important functions in these complex secondary metabolite biosynthesis pathways. Phylogenetic analysis of the top 100 BlastP hits against ForS shows that the ABM domain containing proteins cluster broadly into two main groups; those associated with cyclases and those with predicted monooxygenase activity. Within the monooxygenase cluster, there are two further sub-groups, within which ForS and its

closest homologs cluster (**Chapter 8.4**). It is possible that these two groups of monooxygenases represent one group of chaperones and another group of those are monooxygenases with enzymatic activity. However, too few ABM proteins have so far been characterised to make any firm conclusions regarding this hypothesis at this time.

Using the information gained in this chapter, we predict that the cyclases ForD, ForL and ForR are essential for forming the pentacyclic fasamycin scaffold, with ForL being the most likely candidate for closure of ring B. The ABM proteins ForS and ForU function as chaperones that support the cyclisation process. ForD, ForL, ForR and ForU likely form a multi-enzyme complex where the presence of all enzymes is necessary for any one of them to function. This complex may engulf the entire formicamycin polyketide during cyclisation to form the pentacyclic scaffold. ForS is important for the final step in the cyclisation of the fasamycin scaffold and reduces the formation of the less bioactive formicapyridine shunt metabolites. ForU also appears to play a role as a chaperone protein and its function may be linked to the functionality of the halogenase, ForV. More information is needed to further confirm the roles of these ABM family proteins, such as coexpression of the proteins for *in vitro* biochemistry studies, however, this work provides much information about the biosynthesis of the fasamycins, formicamycins and formicapyridines in *S. formicae* as well as indicating potential roles of ABM family proteins in other T2PKS pathways (Qin *et al.*, 2019).

6.3 Engineering highly efficient strains of *S. formicae* that overproduce specific pathway products

Using the knowledge of fasamycin, formicamycin and formicapyridine biosynthesis gained above in combination with previous biosynthetic and regulatory data presented earlier in this thesis, we hypothesised that it would be possible to generate mutant strains of *S. formicae* with enhanced biosynthetic efficiency that could specifically produce pre-defined pathway products. For example, we know from previous work that removing the major repressor of the pathway, ForJ, results in increased expression of the biosynthetic machinery and therefore increased titres of

formicamycins in the culture extracts. Furthermore, experiments conducted by Dr Zhiwei Qin at the JIC show that *S. formicae ⊿forJ* also produces two new formicamycin congeners, represented by the additional peaks on the HPLC trace (**28** and **29**) (**Figure 6.15**).

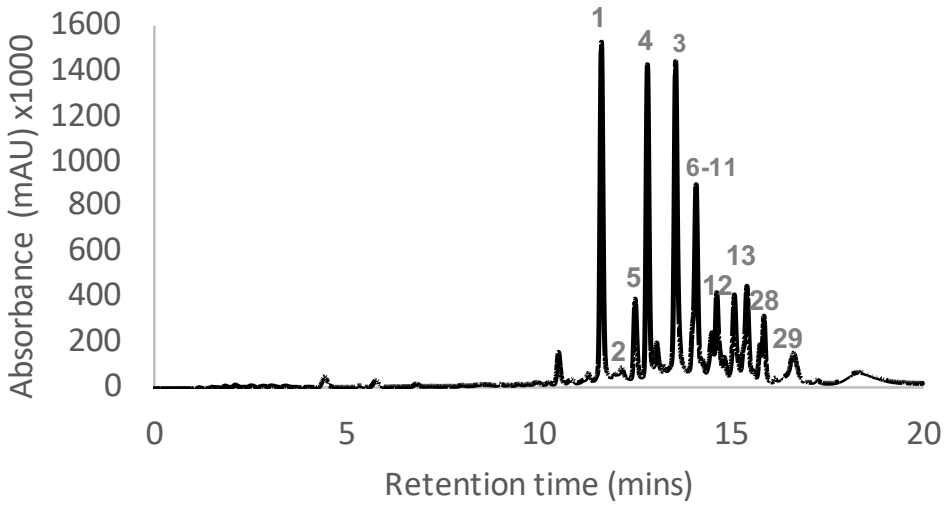


Figure 6.15: Reconstituted HPLC-UV at 280 nm in culture extracts of *S. formicae ⊿forJ*. Previously in this thesis, deletion of the pathway repressor, *forJ*, has have been shown to cause accumulation of high levels of formicamycin congeners in the culture extract of the mutant compared to the wild-type strain. However, in addition to the known formicamycin congeners (**1-12**) two new peaks were visible of the HPLC trace (**28,29**). HPLC (UV) LCMS conducted by Dr Zhiwei Qin.

When cultures of *S. formicae ⊿forJ* were scaled up, small amounts (approximately 2 and 7 mg) of the two compounds were isolated and their structure solved by NMR by Dr Zhiwei Qin. This showed that the peaks represent new formicamycin congeners with additional chlorination sites on the carbon backbone (**Figure 6.16**).

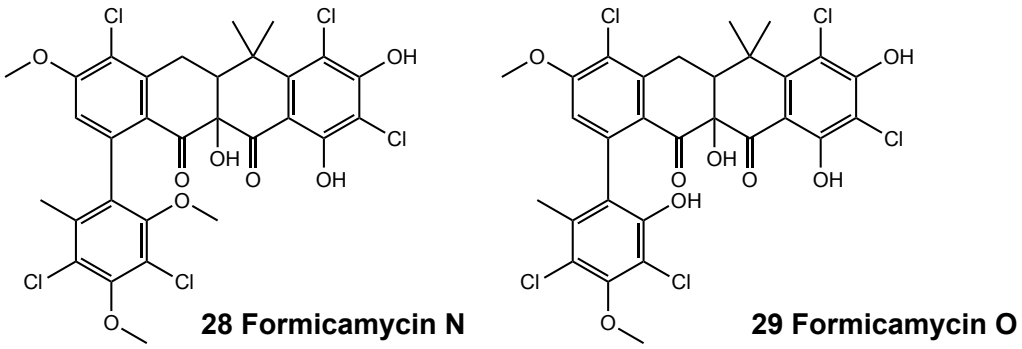


Figure 6.16: *S. formicae ⊿forJ* was cultured at large scale (4L MYM agar) and compounds extracted for structural elucidation by NMR by Dr Zhiwei Qin. This showed that the new peaks represented two new formicamycin congeners, formicamycin N (**28**) and formicamycin O (**29**) that have five chlorine atoms on the carbon scaffold.

These new formicamycin congeners are chlorinated at five positions on the carbon backbone, where previously isolated congeners have a maximum of four chlorine moieties present on the molecule. We have shown that ForV can halogenate multiple positions on the formicamycin backbone as well as several other aromatic substrates. The overexpression of the formicamycin pathway caused by deletion of *forJ* likely results in the accumulation of formicamycins to high titres as well as increased levels of biosynthetic enzymes, including the halogenase ForV, in the cellular environment. As such, the halogenase is likely to chlorinate at these additional positions that are perhaps less energetically favourable and therefore do not occur when substrate and/or enzyme concentrations are lower in the wild-type strain. Currently we do not know whether these additional chlorine moieties increase the bioactivity of formicamycins N and O compared to the previously isolated molecules, however this will be investigated in the future.

Based on the above observation, we hypothesised that by combining the deletion of the pathway repressor with multiple biosynthetic mutations presented earlier in this thesis, we could generate strains of *S. formicae* that would over-produce either fasamycins, formicapyridines or the lactone intermediates, depending on which biosynthetic genes were deleted. In addition, the overexpression of the pathway in combination with these biosynthetic mutations may also encourage the production of additional novel congeners by increasing the occurrence of energetically less favourable reactions. As such, *S. formicae* Δ*forJ* was subjected to further genetic manipulation using the pCRISPomyces-2 system to generate *S. formicae* Δ*forJ* Δ*forV*, a strain lacking the major pathway repressor and the halogenase. Extracts of three biological replicates of this mutant were analysed by HPLC(UV)/LCMS by Dr Zhiwei Qin and Hannah McDonald. As expected, *S. formicae* Δ*forJ* Δ*forV*, overproduced specifically the non-halogenated fasamycin C (**1**) to approximately 3 times the level of the *S. formicae* Δ*forV* single mutant (**Figure 6.17**)**.** No new congeners were isolated from this experiment, further confirming our hypothesis that ForV acts on the fasamycin backbone early in the biosynthesis and performs a gatekeeper function to allow progression of the pathway.

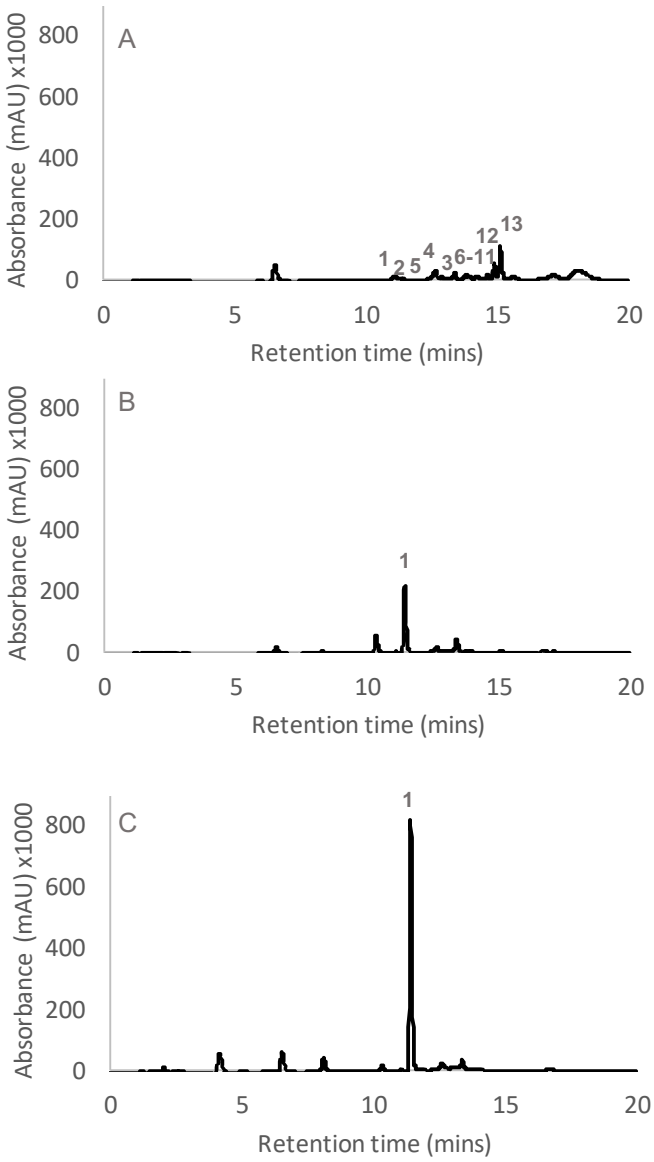


Figure 6.17: Reconstituted HPLC-UV at 280 nm in culture extracts of (A) wild-type *S. formicae*; (B) *S. formicae* Δ*forV*; (C) *S. formicae* Δ*forJ* Δ*forV*. As shown previously, deletion of *forV* results in only the production of the non-halogenated fasamycin C (compound 1). When combined with the deletion of the pathway repressor, *forJ*, deletion of *forV* results in accumulation of compound 1 to approximately 286% of the wild-type strain. HPLC (UV) LCMS conducted by Dr Zhiwei Qin and quantification/integration conducted by Hannah McDonald.

Next, the *S. formicae* Δ*forJ* strain was used to generate *S. formicae* Δ*forJ* Δ*forX*, a strain lacking the major repressor and the Baeyer-Villiger monooxygenase. The single deletion mutant, *S. formicae* Δ*forX* generated only fasamycins, therefore we expected this strain to be a specific fasamycin over-producer. The extracts of three biological replicates of this mutant were quantitatively analysed by HPLC(UV)/LCMS by Dr Zhiwei Qin and Hannah McDonald. The resulting strain produced titres of

fasamycins that were approximately 6 times higher than that of the wild-type. This staggering increase in fasamycin accumulation also highlighted the presence of other, new peaks which had not been seen in previous experiments (**Figure 6.18**).

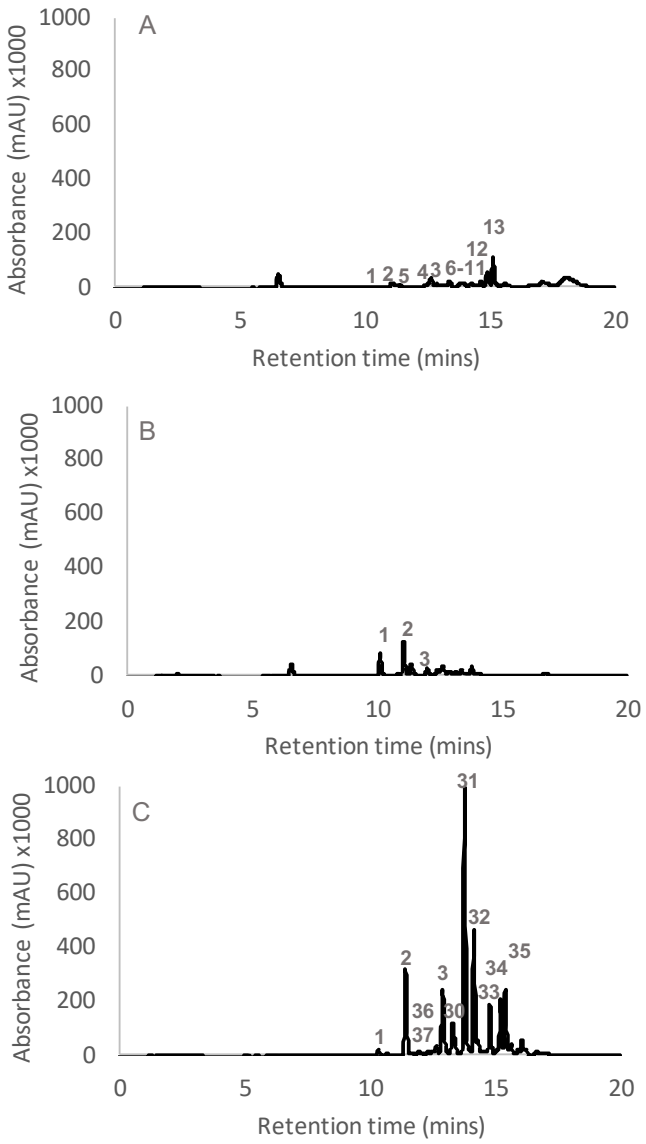


Figure 6.18: Reconstituted HPLC-UV at 280 nm in culture extracts of (A) wild-type *S. formicae*; (B) *S. formicae* Δ*forX*; (C) *S. formicae* Δ*forJ* Δ*forX*. As shown previously, deletion of *forX* results in only the production of fasamycins C-E (compounds 1-3). When combined with the deletion of the pathway repressor, *forJ*, deletion of *forX* results in accumulation of fasamycins to approximately 596% of the wild-type strain. There also appears to be other peaks present in the extracts of *S. formicae*Δ*forJ* Δ*forX* that do not appear in the single mutant. HPLC (UV) LCMS conducted by Dr Zhiwei Qin and quantification/integration conducted by Hannah McDonald.

As a result, *S. formicae* Δ*forJ* Δ*forX* fermentations were scaled up and the extracts were further analysed by HPLC and LCMS to identity the masses and predicted molecular formulae for each compound represented by these new peaks (**Table 6.3**). The masses imply that these new peaks represent a range of novel fasamycin congeners in addition to fasamycins C-E (**1-3**) that have been described previously in this thesis. Six of the eight new fasamycin congeners (**30-35**) were produced in high enough levels that they could be isolated from these large-scale cultures and their structures elucidated by NMR by Dr Zhiwei Qin (**Figure 6.19**). The remaining two compounds (**36** and **37**) are also predicted to be fasamycins, however, this is unconfirmed at this point as the yield of these compounds was too low to enable structural elucidation.

Table 6.3: Cultures of *S. formicae* Δ*forJ* Δ*forX* were scaled up and the mass of each new peak was calculated using mass spectrometry. The predicted molecular formulae suggest that *S. formicae* Δ*forJ* Δ*forX* produces the known fasamycins C-E (**1-3**) along with 8 novel fasamycin congeners. LCMS conducted by Dr Zhiwei Qin.

Compound number	Compound name	Molecular mass	Molecular formula
1	Fasamycin C	473	$C_{28}H_{24}O_7$
2	Fasamycin D	507	$C_{28}H_{23}O_7Cl$
3	Fasamycin E	540	$C_{28}H_{22}O_7Cl_2$
30	Fasamycin G	574	$C_{28}H_{21}O_7Cl_3$
31	Fasamycin H	574	$C_{28}H_{21}O_7Cl_3$
32	Fasamycin I	574	$C_{28}H_{21}O_7Cl_3$
33	Fasamycin J	588	$C_{29}H_{23}O_7Cl_2$
34	Fasamycin K	608	$C_{28}H_{20}O_7Cl_4$
35	Fasamycin L	608	$C_{28}H_{20}O_7Cl_4$
36	Unknown	538	$C_{28}H_{20}O_7Cl_2$
37	Unknown	572	$C_{28}H_{19}O_7Cl_3$

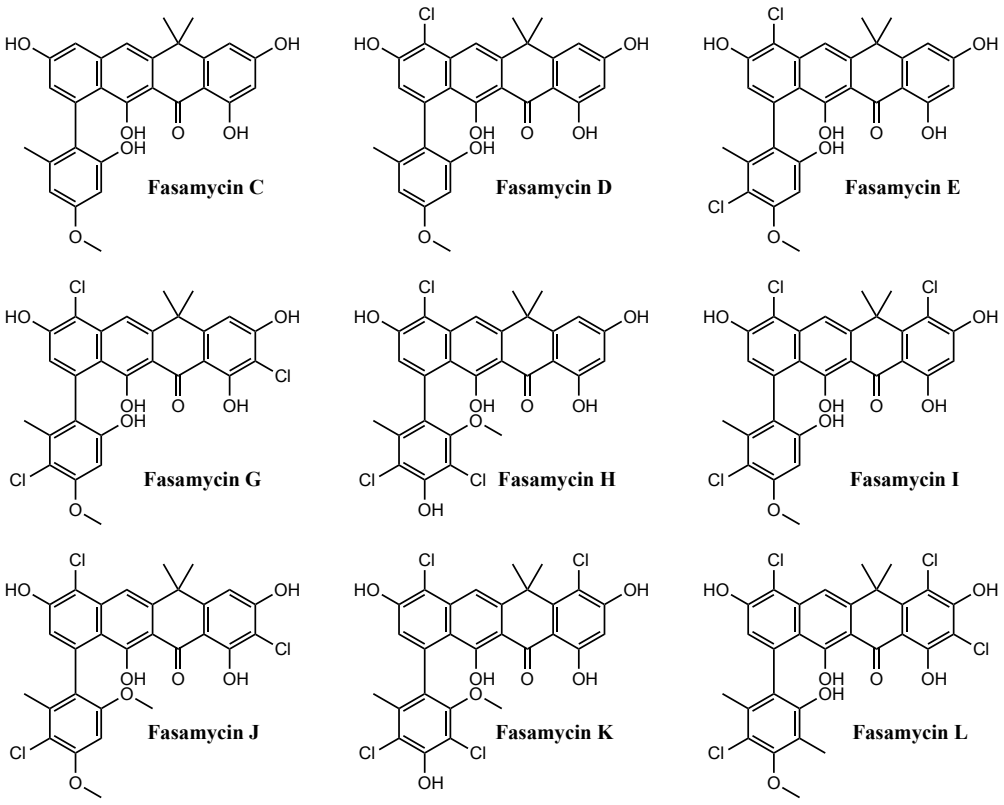


Figure 6.19: Nine fasamycin congeners were produced in high enough titres from *S. formicae* Δ*forJ* Δ*forX,* that they could be isolated for structural elucidation by NMR by Dr Zhiwei Qin. Three of the congeners were the previously known fasamycins C-E. The other 6 compounds produced by *S. formicae* Δ*forJ* Δ*forX* were novel fasamycin congeners G-L which have not been described previously.

The new fasamycin congeners display halogenation at additional positions along the fasamycin backbone that has not been seen in previously isolated fasamycins. As discussed previously, ForV can halogenate the fasamycin and formicamycin backbone at multiple positions on the carbon chain. The overexpression of the formicamycin BGC caused by deletion of *forJ* will lead to both accumulation of fasamycins and increased levels of ForV. As ForX is also removed in this strain, these extra fasamycin precursors cannot progress through the pathway for conversion into a formicamycin. As such, the fasamycin precursors that accumulate are more likely to encounter high levels of the halogenase which will start to chlorinate at additional positions on the carbon backbone. In addition, fasamycins H and K represent the only compounds isolated from the formicamycin BGC that do not contain a methyl group at the base of ring A**.** Work in the previous chapter predicted that ForM was the enzyme responsible for this reaction and concluded that methylation at this

position occurred early in the biosynthesis as no congeners had been isolated without this moiety and deletion of *forM* abolished the production of all previously known fasamycin and formicamycin congeners. The discovery of these new compounds that lack this moiety is not necessarily converse to this theory; we propose that whilst clearly not essential, methylation by ForM occurs early in the biosynthesis of the formicamycins. The overexpression of the formicamycin BGC in *S. formicae* Δ*forJ* Δ*forX* will mean that fasamycin precursors accumulate much more rapidly in this strain than in the wild-type and it is possible that ForM, even though likely overexpressed itself, is overwhelmed by all the additional substrates, meaning that some intermediates are shunted further down the pathway, missing this step of the biosynthesis. These two congeners also display chlorination on ring A, an addition that has not been seen on any previously described fasamycins or formicamycins. It may be that normally the methyl group would prevent chlorination at this position and in its absence, ForV is able to add a chlorine to this position. All of the new fasamycin congeners contain the di-methylation on ring D like all other metabolites isolated from this pathway, confirming our previous hypothesis that this is essential for fasamycin and formicamycin biosynthesis. Furthermore, none of these new fasamycins are methylated on ring B, which agrees with our previous hypothesis that methylation here is part of the conversion of a fasamycin to a formicamycin. Due to the deletion of *forX*, this cannot occur in this double mutant as the pathway is stopped at fasamycin biosynthesis.

Using the same method of combining mutations, we also demonstrate that a mutant of *S. formicae* that overproduces the lactone intermediates can be generated by deleting the oxidoreductase gene in combination with Δ*forJ* to form the strain *S. formicae* Δ*forJ* Δ*forY.* HPLC(UV)/LCMS analysis of this mutant by Dr Zhiwei Qin shows that no fasamycins, formicamycins or formicapyridines are produced, only the lactone intermediates, which accumulated to significantly higher titres than in the *S. formicae* Δ*forY* strain (**Figure 6.20**).

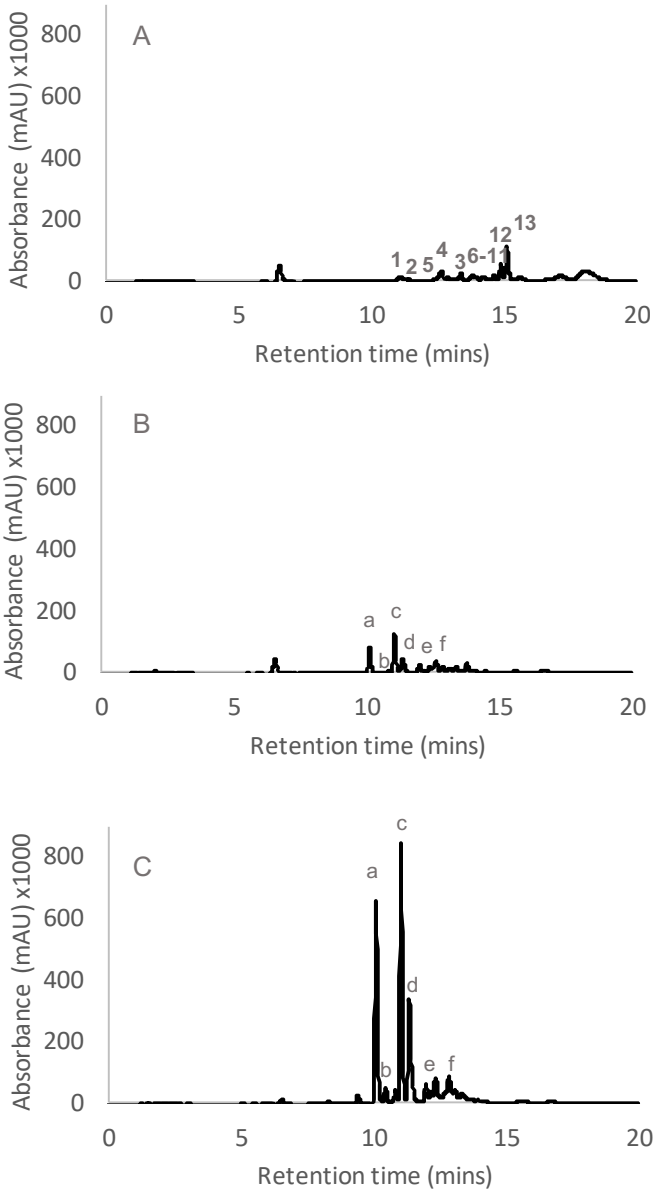


Figure 6.20: Reconstituted HPLC-UV at 280 nm in culture extracts of (A) wild-type *S. formicae*; (B) *S. formicae* Δ*forY*; (C) *S. formicae* Δ*forJ* Δ*forY*. As shown previously, deletion of *forY* results in only the production of the lactone intermediates (**a-f**). When combined with the deletion of the pathway repressor, *forJ*, deletion of *forY* results in accumulation of the lactone intermediates. HPLC (UV) LCMS conducted by Dr Zhiwei Qin.

Finally, we aimed to generate a strain of *S. formicae* that could overproduce the formicapyridines. When studying the roles of the cyclase enzymes above, deletion of *forS* resulted in up-regulation of the formicapyridines and a significant reduction in the production of the fasamycin and formicamycin congeners (**Chapter 6.2**). The double mutant *S. formicae* Δ*forJ* Δ*forS* was generated and extracts analysed by HPLC(UV)/LCMS by Dr Zhiwei Qin (**Figure 6.21**). The results show that the strain *S.*

formicae Δ*forJ* Δ*forS* overproduces formicapyridines at much higher levels than the single Δ*forS* mutant.

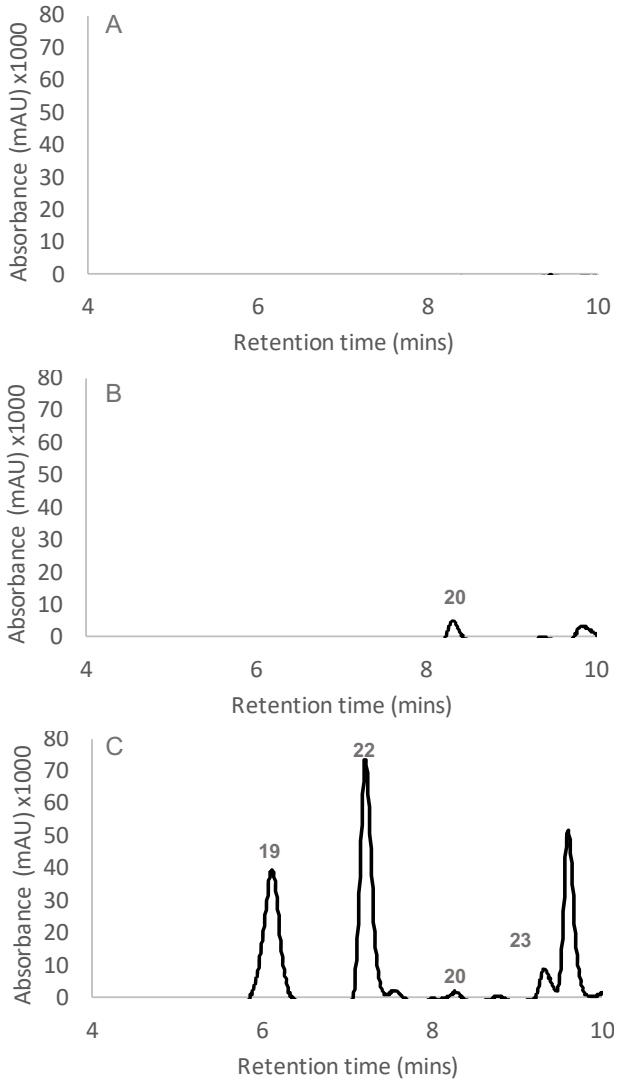


Figure 6.21: Reconstituted HPLC-UV at 390 nm in culture extracts of (A) wild-type *S. formicae*; (B) *S. formicae* Δ*forS*; (C) *S. formicae* Δ*forJ* Δ*forS*. As shown previously, deletion of *forS* results in increased production of formicapyridines **18-23**. When combined with the deletion of the pathway repressor gene, *forJ*, deletion of *forY* results in further accumulation of the formicapyridines. HPLC (UV) LCMS conducted by Dr Zhiwei Qin.

This library of strains that can specifically over-produce various biosynthetic congeners, intermediates and shunt metabolites from the formicamycin pathway can be used in the future to isolate high titres of individual compounds for further *in vitro* work. This work has also led to further novel fasamycin and formicamycin congeners being isolated from *S. formicae* and the *for* BGC. Furthermore, preliminary analysis suggests that *S. formicae* Δ*forJ* Δ*forX* produces the new fasamycin congeners

during liquid culture in TSB medium (**Table 2.1**) (personal communication with Dr Zhiwei Qin, March 2019). It is currently unknown whether all the double mutants generated here can produce compounds during liquid culture but this will be the subject of further investigation in the Wilkinson laboratory. If these strains do all produce during liquid fermentation, they could be very attractive for the industrial production of these clinically relevant natural products.

6.4 The proposed fasamycin, formicamycin and formicapyridine biosynthesis pathway in *S. formicae*

Using the data described above in combination with the biosynthetic data collected throughout this thesis, we are now able to propose a complete and detailed biosynthetic pathway for the fasamycins, formicamycins and formicapyridines. Condensation of acetyl-CoA by the ACC carboxylase enzymes, ForH and ForI, results in the formation of the malonyl-CoA extender units required for polyketide chain formation. Feeding data described previously (**chapter 3.4**) shows that these multiple malonyl-CoA starter units are condensed, presumably by the combined actions of the minimal PKS, ForA, ForB and ForC as well as ForE, to form a tridecaketide intermediate (**17**). We predict that this intermediate then enters the multi-enzyme complex formed by ForD, ForL, ForR and ForU for sequential cyclisation to form a pentacyclic structure, with the aid of the chaperone ForS to increase pathway efficiency. Based on bioinformatic analysis above we predict that ring A is closed first, followed by rings E and D, which are cyclised by ForD. Cyclisation of ring C likely occurs next, followed by formation of ring B, a reaction which is probably catalysed by ForL. It is thought that the methyltransferase ForT acts next, possibly as part of the multi-enzyme complex or closely linked, to catalyse the addition of the di-methyl group on ring C, as all isolated metabolites from the formicamycin BGC contain this moiety. The cyclised, di-methylated pentacyclic intermediate produced by the cyclase/methyltransferase complex is then cleaved from the ACP by the actions of ForN followed by decarboxylation (either spontaneous or catalysed by ForQ) with the aid of ForS, to form fasamycin C. During the process of cyclisation and decarboxylation, some intermediates can undergo spontaneous side-reactions to

generate shunt metabolites like the formicapyridines. The formicapyridines can be halogenated by ForV and O-methylated at various positions on the carbon chain. The pathway products are next likely O-methylated by ForM after cyclisation and liberation from the ACP, as both fasamycins and formicamycins are usually O-methylated at the base of ring A and deletion of this enzyme abolishes the production of any congeners, implying it works closely with the core multi-enzyme complex to produce the non-halogenated fasamycin C.

The fasamycin C formed by the above reactions can then be brought to ForV for halogenation, perhaps by the chaperone ForU. The activity of ForV activates molecular oxygen to generate hydroperoxy flavin that is used by the Baeyer-Villiger monooxygenase, ForX, to convert the fasamycin precursors into the lactone intermediates (**Chapter 5.2**). We predict that the lactone intermediates are then O-methylated on ring B by ForW, generating the highly reactive intermediate for reduction by ForY and conversion of the lactone intermediates into the formicamycin backbone. Formicamycins A and B can then be further halogenated by ForV to form the full family of formicamycin congeners seen in extracts of wild-type *S. formicae* (**Figure 6.22**).

Figure 6.22 (see following page for figure legend)

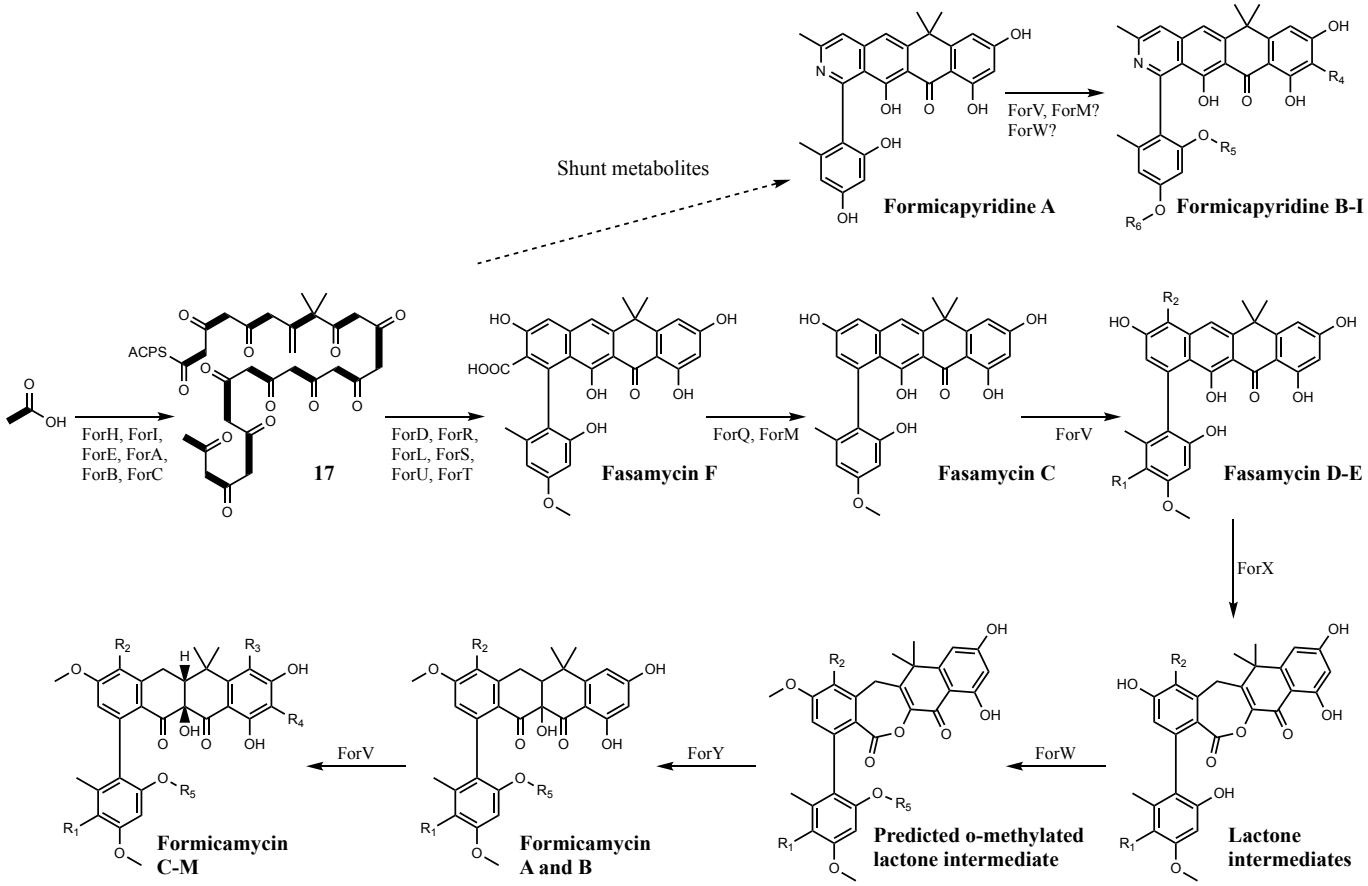


Figure 6.22 (previous): Schematic of the full fasamycin, formicamycin and formicapyridine biosynthetic pathway in *S. formicae*. The malonyl-CoA starter units synthesised by ForH and ForI are used to by ForA, ForB, ForC and ForE to assemble the polyketide chain. This is cyclised by ForD, ForR, ForL, ForS and ForU and di-methylated by ForT to form the core carbon scaffold. This is then released from the ACP and decarboxylated either spontaneously or by ForQ to form fasamycin F. During this process, ForM can methylate ring A. Shunt metabolites like the formicapyridines can also be formed during this stage of the biosynthesis due to a leaky pathway. Both the fasamycins and the formicapyridines generated at this stage of the biosynthesis can be halogenated by ForV at multiple positions on the carbon chain (R_1-R_4). The formicapyridines can also be O-methylated by one of the methyltransferase enzymes encoded in the pathway at R_5, either ForM or ForW. During the halogenation of the fasamycins and formicapyridines, activated molecular oxygen is generated in the form of the flavin hydroperoxide that is used by the Baeyer-Villiger ForX to form the lactone intermediates. These are then predicted to be O-methylated by ForW to generate a reactive intermediate for reduction by ForY, forming the formicamycin backbone. This formicamycin backbone can then be further halogenated by ForV to generate the full family of congeners.

Further work is currently being conducted in the Wilkinson laboratory to build the formicamycin pathway in a heterologous host a single enzyme unit at a time. As well as providing a platform to reconstruct the formicamycin pathway to generate further novel analogues, this work will provide more information about how these enzymes function together in the native pathway and provide more evidence to support the proposed biosynthetic pathway outlined above.

6.5 Discussion

This chapter aimed to further elucidate the formicamycin biosynthetic pathway and has resulted in a much greater understanding of the functions of many of the enzymes encoded by the formicamycin BGC. Furthermore, using targeted metabolomics, we have identified a new family of pyridine containing polyketide natural products from the formicamycin BGC, showing that this single T2PKS pathway can synthesise three distinct but related families of compounds; the fasamycins, the formicamycins and the formicapyridines. The results presented here show that the production of this complex mixture of secondary metabolites occurs via a multi-step pathway that consists of several routes for the production of bi-products and shunt metabolites like the formicapyridines. Many of the enzymes involved in formicamycin biosynthesis appear to interact, with the activity of one affecting the function of another. For example, the cyclase enzymes, ForD, ForL, ForR and ForU appear to function as a multi-enzyme complex, as deletion of one abolishes the

biosynthesis of any products or intermediates. Accessory enzymes like the halogenase, ForV and the methyltransferase, ForT also appear to interact with enzymes from this complex or other chaperones that function closely with this multi-enzyme structure. The importance of multi-enzyme interactions is well recognised in modular systems such as T1PKS pathways and is generally accepted as being fundamental for the proper progression of these biosynthetic pathways (Tran *et al.*, 2010; Robbins *et al.*, 2016; Dodge, Maloney and Smith, 2018; Kosol *et al.*, 2018). In iterative systems like the T2PKSs, interactions between domains of the ketosynthase enzymes are well characterised, but protein-protein interactions between the tailoring enzymes are less well recognised. Type 2 polyketide biosynthesis is related to Type 2 fatty acid biosynthesis, where the ACP performs the central role for recognition by the multiple biosynthetic enzymes (Chen, Re and Burkart, 2018). While the ACP has been shown to play an important role in substrate recognition in other T2PKS pathways during the early stages of biosynthesis, mainly by the KS, we believe that some of the enzymes involved in methylation and cyclisation of the formicamycin backbone recognise substrates that have already been liberated from the ACP. This suggests that other proteins encoded within the formicamycin biosynthetic pathway are likely responsible for the coordination of the actions of these multiple enzyme units instead.

Many complex pathways like the formicamycin biosynthesis pathway encode for dedicated chaperone proteins that help direct nascent intermediates into defined reaction channels in order to synthesise more of the desired final product and less of the shunt metabolites (Hertweck *et al.*, 2007). Through our investigations into the functions of the putative cyclase enzymes encoded by the formicamycin pathway, we have highlighted the importance of chaperones in the formicamycin biosynthetic pathway, particularly ForS, the putative ABM, that appears to act as a chaperone during the cyclisation of the polyketide backbone and the formation of the pentacyclic fasamycin scaffold. Mutational analysis showed that ForS is not essential for the biosynthesis of the fasamycins and formicamycins, but removal of this enzyme reduces the efficiency of the pathway, resulting in less of the desired final products and the generation of higher levels of shunt metabolites. There is also

evidence in this thesis that the function of the halogenase ForV, may be dependent on an additional chaperone or carrier that so far remains unidentified.

Whilst the formicapyridines do not display significant antibacterial activity, their discovery provides significant knowledge about the formicamycin biosynthetic pathway. Furthermore, they represent novel structures and nitrogen containing compounds from T2PKS pathways are relatively rare. Those examples that do exist in the literature have been shown to have cytotoxic activity and therefore the potential anticancer activity of the formicapyridines will be investigated in the future. The biosynthetic studies presented in this chapter also led to the discovery of 10 new fasamycin congeners and 2 new formicamycin congeners that have not previously been isolated from either the formicamycin BGC or any other fasamycin pathways in the database. Due to time limitations, we are yet to identify whether these new congeners display any bioactivity, however, it is assumed that these metabolites will display similar bioactivities to the previously isolated fasamycin and formicamycin congeners due to their significant structural similarity. It will be interesting to determine whether the additional halogenation and methylation on these new congeners increases the potency of the compounds. These future bioactivity studies will be made easier due to the development of a library of *S. formicae* mutants that produce high titres of specific pathway products and intermediates. These strains can be used to isolate large quantities of compound for use in *in vitro* studies. Furthermore, there is potential for these strains to be used in industry for the production of these clinically relevant compounds, particularly if we are able to show that they consistently produce during liquid culture.

This work highlights the importance of multi-enzyme interactions in the biosynthesis of complex, polycyclic natural products. Over the course of evolution, intricate systems have developed within the formicamycin biosynthetic pathway that tightly control the cyclisation of the extended polyketide backbone and the subsequent tailoring steps that occur. By suppressing undesirable, spontaneous cyclisation reactions that might generate shunt metabolites and other minor congeners, nature can ensure that the desired secondary metabolites are generated with the highest efficiency by the host microorganism. This work also shows that by genetically

manipulating biosynthetic pathways, not only can we learn about how these secondary metabolites are synthesised *in vitro*, but we can generate increased chemical diversity from these pathways and encourage the biosynthesis of 'unnatural' natural products that are either not made by the wild-type strain, or are made at levels that are too low for isolation and structural elucidation.

7 Conclusions and further work

7.1 Characterising the biosynthesis of the novel antibiotics, the formicamycins, in *Streptomyces formicae*

The aim of this work was to begin characterising antibiotic production in *Streptomyces formicae*, a novel strain isolated from a fungus-farming plant ant nest. The genome sequence of *S. formicae* revealed that it is a talented strain, with at least 45 secondary metabolite BGCs, many of which appear novel and display low homology to other BGCs in the database. Prior to the start of this project, previous work from the Hutchings and Wilkinson laboratories had shown that *S. formicae* produces metabolites with antibacterial and antifungal activity under normal laboratory conditions, including the novel pentacyclic polyketides, the formicamycins. However, little was known about the biosynthesis of these molecules and no attempts to genetically modify the organism had been made. During the course of this project, a CRISPR-Cas9 platform for the efficient genetic engineering of *S. formicae* was developed and optimized, allowing numerous mutants to be generated for biochemical studies. Using these protocols, we demonstrate that a single T2PKS BGC in *S. formicae* is responsible for the biosynthesis of three groups of natural products; the fasamycins, the formicamycins and the formicapyridines. We have characterised the functions of most of the biosynthetic enzymes encoded by this BGC, a process which has revealed additional novel intermediates and shunt metabolites, as well as shedding light on the biosynthesis of both the formicamycins and other related aromatic type 2 polyketides.

Biosynthetic pathways are often perceived as direct and linear routes that result in a single defined end-product, but most secondary metabolite pathways are actually significantly more complex. Like the formicamycin pathway, some BGCs encode for the production of multiple related structures, for example the T1PKS BGC in *S. venezuelae* that is responsible for the biosynthesis of the 12-membered ring macrolides methymycin and neomethymicins as well as the 14-membered ring macrolides narbomycin and pikromycin. In this case, the ability to produce multiple

structures is due to the ability of the pathway to terminate the polyketide chain at two different points of assembly, resulting in two macrolactones of different sizes (Xue *et al.*, 1998). In the formicamycin BGC, the production of a mixture of these three distinct compounds seems to be the result of a complex multi-step biosynthetic pathway that is prone to the build-up of precursors, intermediates and spontaneously generated side-products. Fasamycins A and B were previously isolated from environmental DNA, however fasamycins C-E, as well as the formicamycins and the formicapyridines from *S. formicae*, represent novel structures (Feng, Kallifidas and Brady, 2011; Qin *et al.*, 2017). Our evidence suggests that the fasamycins are the biosynthetic precursors of the more bioactive formicamycins. Through a collaborative project with Dr Zhiwei Qin in Prof Barrie Wilkinson's group at the JIC, we identified the biosynthetic link between these distinct but related compound families. Our results describe a novel ring-expansion, ring-contraction mechanism via a Baeyer-Villiger monooxygenase and an oxidoreductase that converts the fasamycin precursor to the formicamycin. This work also led to the isolation and characterisation of five bioactive lactone intermediates that also originate from the formicamycin BGC. As part of the same project, we were able to propose a mechanism for the evolution of the formicamycin BGC from a fasamycin BGC, by the horizontal gene transfer of four genes encoding the monooxygenase, the oxidoreductase, a MarR regulator and an MFS family transporter. In addition to the eDNA isolated from the metagenomics study in Arizona that has been shown to produce fasamycins A and B, we propose that another actinomycete, *S. kanamyceticus*, also contains a fasamycin BGC, but this is yet to be experimentally confirmed. A search of the database implies that *S. formicae* is the only genome sequenced strain that has the potential to make both fasamycins and formicamycins.

We have also characterised the role of the putative cyclase enzymes encoded within the BGC, as well as the methyltransferases involved in formicamycin biosynthesis. We have demonstrated that the majority of the core biosynthetic machinery functions as a multi-enzyme complex, where the presence of all the enzyme components is necessary to generate any products or intermediates. In addition, we have highlighted the importance of chaperone proteins, particularly ABMs that direct

the biosynthetic pathway and guide nascent intermediates for the most efficient production of the main pathway products. Antibiotic production by microorganisms comes at an expense and therefore secondary metabolite pathways evolve to become highly efficient in order to produce only those metabolites with the highest bioactivities that will provide the greatest competitive advantage. In the case of the formicamycin BGC, this is predicted to be through the development of dedicated chaperone proteins that guide nascent intermediates from one enzyme to the next, reducing the spontaneous generation of less bioactive side-products like the formicapyridines. Multi-enzyme interactions are well established in linear systems like the T1PKSs, however, understanding of more complex iterative pathways like the formicamycin BGC often fall behind. Our work shows that the communication between these multiple enzymes goes beyond the recognition of the ACP and often involves dedicated proteins or protein subunits that coordinate these protein-protein interactions. Furthermore, we have shown that the single halogenase encoded within the BGC is able to chlorinate and brominate at multiple positions on the formicamycin backbone, as well as other aromatic substrates. The halogenase also plays a gatekeeper role in controlling the progression of the biosynthetic pathway from fasamycin biosynthesis to formicamycin production. This is also predicted to be depended on interactions with other biosynthetic enzymes encoded in the cluster, providing more evidence that progression of formicamycin biosynthesis pathway relies on protein-protein interactions, this time coordinated by a common flavin ligand.

The fasamycins and formicamycins display potent antibacterial activity against Gram-positive pathogens including drug-resistant clinical isolates as well as membrane permeabilised Gram-negative strains. The work in this thesis shows that the fasamycins and formicamycins are able to inhibit the activity of bacterial topoisomerase enzymes *in vitro*. In contrast, the formicapyridines do not display significant bioactivity, showing that ring B is an important part of the core pharmacophore involved in inhibition of topoisomerases. Previous work suggests that the tricyclic chloro-*gem*-dimethyl-anthracenone sub-structure is responsible for the inhibition of fatty acid biosynthesis by the antibacterial agents, fasamycins A/B

and BABX (Feng *et al.*, 2012). This evidence, in addition to the lack of resistance developed in MRSA during previous work, led us to hypothesise that the fasamycins and formicamycins from *S. formicae* may be able to inhibit two distinct intracellular targets; DNA topoisomerases and fatty acid synthases. For this reason, we propose that these compounds represent a promising new group of antimicrobials for clinical development. By inhibiting two distinct molecular targets, the development of resistance in the clinic would be reduced and the compounds could be used to treat infections that are already resistant to currently available antibiotics.

The genetic tools developed during this thesis were also applied to the cluster situated regulators within the formicamycin BGC. Thus, we have begun to understand the complex regulatory cascade responsible for controlling both the biosynthesis of formicamycins and the host resistance mechanisms in *S. formicae*. A single MarR regulator, ForJ, is predicted to repress the majority of the biosynthetic genes in the formicamycin BGC. Activation of formicamycin biosynthesis is controlled by the TCS ForGF, which is predicted to induce production of the malonyl-CoA starter units required for polyketide biosynthesis, however, the activation signal for the ForG sensor kinase is currently unknown. Our evidence also suggests that the other MarR regulator encoded within the BGC, ForZ, controls the expression of the host-resistance mechanism in response to accumulating levels of products from the pathway. By applying next generation sequencing technologies, we have also characterised the transcriptional organisation of the *for* BGC as well as mapping the transcriptional start sites across the genome for those genes that are expressed under formicamycin producing conditions. As well as the *for* BGC, the genes encoding the production of the antifungal compounds that are made by *S. formicae* were also likely expressed during this experiment, as these are usually also visible when culture extracts are analysed by HPLC(UV), therefore we may be able to use this data for further characterisation of another BGC in the future.

Overall, the work in this thesis has begun to provide a detailed understanding of the biosynthesis of novel antibiotics from *S. formicae*, with a focus on the biosynthesis of metabolites encoded by the *for* BGC. Through the extensive genetic manipulation of the pathway, we have been able to put forward a detailed proposed biosynthetic

pathway with the functions of most of the gene products described. This comprehensive knowledge of formicamycin biosynthesis can inform future drug discovery efforts by both natural products chemists and synthetic chemists. Using the combined knowledge of formicamycin biosynthesis and regulation, we generated a library of *S. formicae* strains that produce high titres of specific pathway products that can be used in the future for multiple applications. Furthermore, the genetic manipulation of the *for* BGC enabled the discovery of further novel natural products that were not present in the extracts of the wild-type strain, such as biosynthetic intermediates and the novel fasamycin and formicamycin congeners generated when high levels of precursors accumulate and get shunted through alternative biosynthetic routes. This project highlights the huge potential of genetic engineering of novel environmental isolates in the discovery of novel, clinically relevant natural products.

Whilst this work represents significant progress towards characterising antibiotic production in *S. formicae,* a great deal of novel biochemistry still remains to be discovered from this strain. The compounds studied during this work originate from just one of the 45 BGCs within the genome, many of which look novel. Analysis of culture extracts of *S. formicae* grown under standard laboratory conditions shows the presence of multiple peaks, some of which correspond to compounds from the *for* BGC and some of which have been shown by bioassay guided fractionation to represent the antifungal compound(s) (**Figure 1.2**). These antifungal compounds have potent activity against the multi-drug resistant *Lomentospora prolificans*. *L. prolificans* is an emerging human pathogen and there are currently no clinically approved treatments for systemic infections with this agent. Therefore, we predict that any compounds with activity against this fungus are likely to have novel structures and novel mechanisms of action. Mass spectrometry analysis indicates that the antifungal activity may be due to a mixture of polyenes, however, attempts to isolate and purify the antifungal compounds to solve their three-dimensional structure have so far been unsuccessful. The compounds co-elute and even after fractionation, the properties of the compounds change after relatively short periods of time. This could be due to degradation or it could be conformational if the

compounds are able to switch between different states. The priority now is to identify the BGC encoding for these compounds. Polyene antifungal compounds like nystatin and candicidin often originate from T1PKS BGCs (Seipke *et al.*, 2012). Due to the modular architecture of T1PKS BGCs, once the cluster is identified, it may be possible to predict the structure from the sequence and modules present. Furthermore, we could then use the protocols developed during this thesis to genetically modify the cluster to subtly change the structure of the compound(s) produced and therefore make them easier to isolate and purify.

7.2 The potential for the discovery of more novel natural products from *Streptomyces formicae*

In addition to the antifungal compounds, there is significant potential for the discovery of many more natural products from *S. formicae*. Many of the other BGCs in the *S. formicae* genome look promising as a potential source of novel antimicrobial compounds, as several have less than 10% homology to known BGCs in the database, according to AntiSMASH. The hope is that because most of the BGCs encoded within the genome look novel, any new compounds isolated from the strain will have novel structures and may therefore be valuable in the treatment of drug-resistant infections. The low homology to other BGCs in the database means it can be challenging to predict the structures of the products synthesised by the proteins encoded, however, there are some BGCs that look like interesting candidates for further investigation. For example, two putative peptide synthetase modules within cluster 5 show homology to the daptomycin peptide synthetases. Daptomycin is a potent antimicrobial that acts via the physical perforation of the cell membrane. As such, it is more challenging for pathogens to develop resistance to daptomycin, making it a valuable treatment against drug-resistant infections (Hur, Vickery and Burkart, 2012). According to AntiSMASH, the putative proteins encoded by *S. formicae* show approximately 50% coverage and identity to the synthetases in daptomycin biosynthesis, implying the core peptide products may be similar. No structural predictions about the product can be made at this time, however, the accessory enzymes encoded in the *S. formicae* BGC are different to those in the

daptomycin BGC, implying that the overall structure of the compounds encoded could be significantly different, potentially representing a novel agent of this class (**Figure 7.1**). In addition, NRPS genes within cluster 31 show high homology to the core NRPS genes from the BGCs encoding for feglymycin, teicoplanin and balhimycin biosynthesis, therefore we predict that this BGC encodes a glycopeptide antibiotic (**Figure 7.2**). Glycopeptides are rare and are also valuable in the fight against drug-resistant infections because as well as often being active against Gram-negative pathogens, resistance often develops slowly to these compounds due to their ability to inhibit cell wall biosynthesis by binding to cell wall precursors rather than transpeptidase enzymes (Yim *et al.*, 2014). Characterising the molecules produced by these biosynthetic pathways is important in order to discover more novel antimicrobials from this talented strain.

Figure 7.1: AntiSMASH analysis shows that 15% of the genes in BGC 5 in *S. formicae* show similarity to genes from within the daptomycin BGC. Much of the similarity lies within the core NRPS genes, shown in purple, implying that the core peptide structures may be similar.

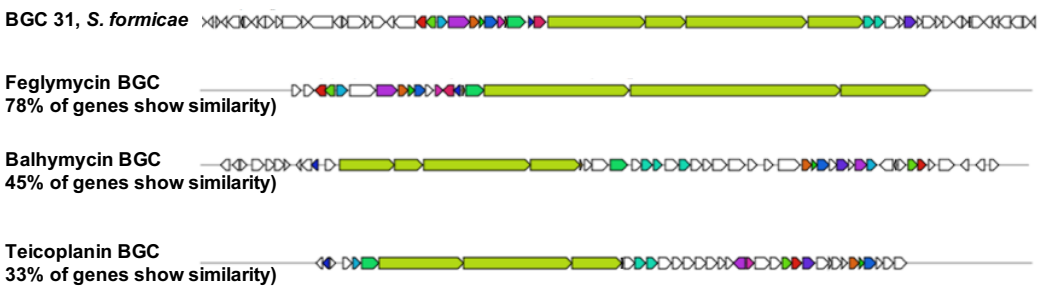


Figure 7.2: AntiSMASH analysis shows that core NRPS genes encoded in BGC 31 in *S. formicae* show significant similarity to the NRPS genes in other glycopeptide BGCs. Much of the similarity lies within the core NRPS genes, shown in pale green, implying that products may be structurally similar.

During the course of this PhD project, some initial work has been conducted to begin to identify other novel compounds encoded in the putative BGCs in the *S. formicae* genome. It was assumed that under standard laboratory conditions, all the BGCs except the formicamycin BGC and the antifungal BGC are silent or not expressed. Therefore, several pleiotropic and genotypic methods of switching on BGCs were trialled. Media screens in both solid and liquid culture did not yield any significant

changes to the secondary metabolite profile. Including glass beads in the liquid medium before inoculation of *S. formicae* sporadically induces formicamycin production, however, this effect is variable and unreliable, so for the isolation of large amounts of formicamycin congeners, the use of mutants generated in this thesis remains the more attractive method. RNA polymerase and ribosomal engineering using the antibiotics rifampicin and streptomycin have been shown in other actinomycetes to cause significant upregulation of expression of secondary metabolite genes, therefore we applied the same approach to *S. formicae* (Wang, Hosaka and Ochi, 2008; Hosaka *et al.*, 2009). For *S. formicae,* much higher concentrations of antibiotics than those used in the original studies had to be used because the strain appeared to possess some natural resistance to both antibiotics. Furthermore, the majority of the resistant isolates that were obtained did not contain mutations in the desired genes or DNA regions and showed no differences in secondary metabolite profiles. We predict that due to the prolific biosynthetic potential of *S. formicae,* it will encode within its genome many host-resistance mechanisms and may therefore be able to evade toxicity to many antibiotic compounds, particularly those like rifampicin and streptomycin that also originate from *Streptomyces* species. We therefore predict that these methods are unlikely to be a good way of encouraging antibiotic production in *S. formicae.* Similarly, efforts were made to pleiotropically alter the expression of secondary metabolite BGCs by mutating the highly conserved TCS MtrAB, which has been shown previously by Nicolle Som (Hutchings laboratory) to increase antibiotic production in *S. venezuelae* (Som *et al.*, 2017). Whilst this method did lead to new peaks appearing in the culture extracts of *S. formicae* during analysis by HPLC/LCMS, different traces were evident on every LCMS run and with every biological replicate of each mutant strain (data not shown). This implies that the disruption global regulators in talented strains like *S. formicae* has a dynamic effect on the secondary metabolism. Due to the fact that the wild-type *S. formicae* already produces such a complex mixture of secondary metabolites under normal laboratory conditions, it can be difficult to identify new compounds in the extracts generated during pleiotropic studies like the ones described here. These methods often result in changes to the expression of multiple BGCs across the genome, resulting in even more complex mixtures of products in the

culture extracts. Therefore, we predict that targeted genetic approaches will be a more effective method for identifying novel secondary metabolites from this talented strain.

Interestingly, deletion of the *for* BGC conducted in **chapter 3.1**, while abolishing the biosynthesis of all formicamycin congeners, did not abolish the antibacterial activity. HPLC(UV) results presented in this thesis clearly show that no fasamycin or formicamycin congeners are present in the extracts of *S. formicae Δfor*, however, when tested against Gram-positive strains like *B. subtilis* and MRSA, the zone of inhibition is actually consistently larger than the wild-type strain (**Figure 7.3**).

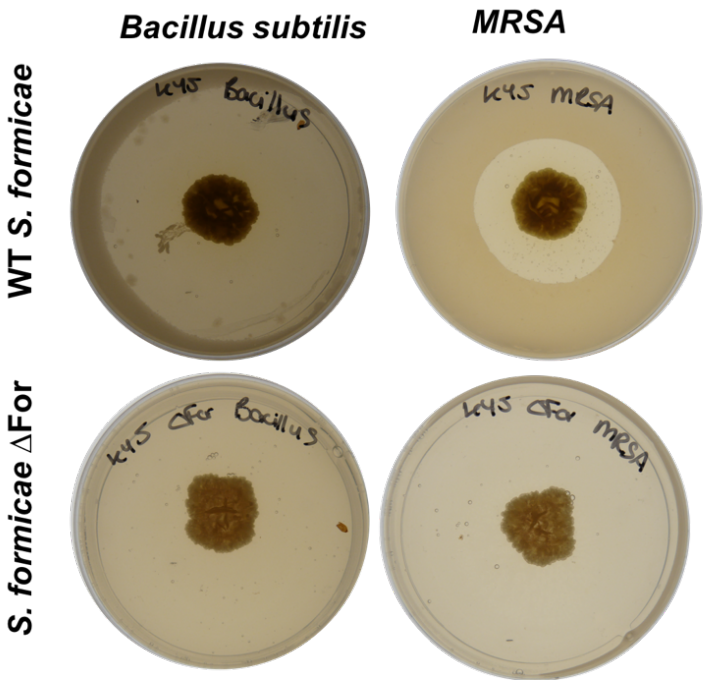


Figure 7.3: Bioassays against Gram-positive strains *B. subtilis* and MRSA show that *S. formicae* Δ*for* generates much larger zones of inhibition than the wild-type strain, almost completely clearing a small 9mm plate. *S. formicae* Δ*for* does not produce any formicamycin congeners, therefore the metabolites must originate from another BGC within the genome.

This implies that deletion of the formicamycin BGC induces the expression of another BGC elsewhere in the genome. We predicted that this could be due to loss of cross-cluster regulation through the deletion of the regulatory genes in the formicamycin BGC. During the ChIP-sequencing experiment conducted in chapter 4, the MarR regulator ForJ and the response regulator ForF appeared to bind to the promoters of genes outside of the formicamycin BGC (**chapter 4.3.1**). There are multiple examples

where cluster-situated regulators from one BGC are involved in the regulation of another cluster elsewhere in the genome. For example, JadR1 in *Streptomyces venezuelae*, the main activator of the jadomycin BGC, has also been shown to repress chloramphenicol biosynthesis (Xu *et al.*, 2010; Chater, 2016). We therefore hypothesised that either ForJ or ForF were responsible for the regulation of another BGC elsewhere in the genome that encodes for the production of an antimicrobial. If correct, removal of the formicamycin BGC in *S. formicae ⊿for* would therefore remove this regulatory control and possibly induce expression of the alternative antimicrobial. To test this theory, constructs containing each of the regulators, either ForJ or the TCS ForGF, were introduced into *S. formicae ⊿for* in order to complement the loss of the regulatory elements without re-introducing formicamycin biosynthesis in the strain. However, none of the complemented strains lost their bioactivity, with the resulting zones of inhibition against *B. subtilis* remaining the same as the *S. formicae ⊿for* control (**Figure 7.4**).

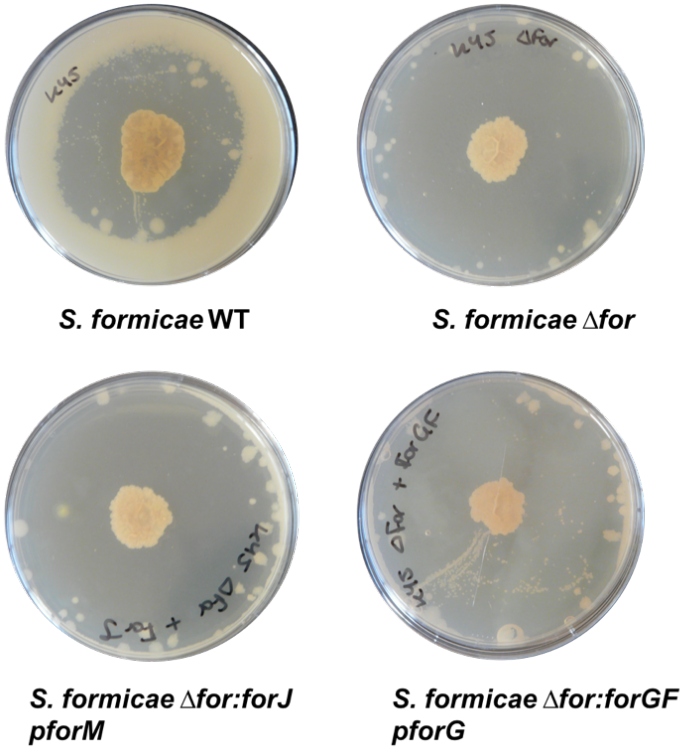


Figure 7.4: Bioassays against Gram-positive strain *B. subtilis* show that the increased bioactivity of *S. formicae* Δ*for* compared to the wild-type strain is not rescued on complementation with either *forJ* or forGF. This implies that expression of the new BGC is not controlled by cross regulation from either of these cluster situated regulators from within the formicamycin BGC.

This suggests that neither ForJ or ForF are responsible for the cross-regulation of the BGC that is switched on in *S. formicae ⊿for.* It is possible that this increase in bioactivity is to do with the other TCS, ForBBCC, that is encoded at the edge of the *for* BGC, as this would have been included in the original whole-cluster deletion, before the cluster edges had been characterised. It could also be that the loss of the formicamycin BGC induces expression of alternative BGC(s) elsewhere in the genome due to the increased availability of precursors and biosynthetic resources. The biosynthesis of the formicamycins will be metabolically expensive for *S. formicae* and therefore the deletion of the BGC results in much less carbon and other resources being used by the cell. These resources are then able to flux through alternative pathways, giving rise to the production of different secondary metabolites (Gomez-Escribano and Bibb, 2011). This will be investigated further in the future. As yet, we have been unable to identify any peaks by either HPLC or LCMS that correspond to this antibacterial activity, however, the potent activity of this new antimicrobial against MRSA makes it a priority for further investigation.

In addition to searching for the new antibacterial compound, further work has been done to try and identify the BGC responsible for the production of the antifungal. Without a structure of the compound(s), we are unable to identify which of the 45 putative BGCs encoded within the genome are responsible for their production, however, we predict that the compounds originate from a T1PKS BGC. There are multiple clusters in the *S. formicae* genome that show some characteristics of T1PKS genes, however, only clusters 30 and 33 contain significant repeated T1PKS modular genes that look like they may be capable of producing a polyene antifungal. According to AntiSMASH analysis, cluster number 33 shows 54% similarity to the lasalocid BGC. Cluster number 30 shows 46% similarity to the filipin BGC, 27% similarity to nystatin and 28% similarity to candicidin, implying it might be more likely to synthesise a polyene antifungal (**Figure 7.5**). We know that the antifungals are produced under the same laboratory conditions as the formicamycins and therefore the BGC responsible should be expressed in the cappable RNA sequencing experiment conducted in chapter 4. Unfortunately, TSSs were mapped in both clusters 30 and 33, implying that some genes from both of these clusters were being

transcribed at the time of sampling, therefore we remain unsure which cluster is responsible for the biosynthesis of the antifungal compound(s). Furthermore, structural predictions of the products from each of the T1PKS encoded by clusters 30 and 33 can be made due to the modularity of the genes and the high percentage conservature to other sequenced PKSs in the database. This analysis shows that of the two products, cluster 30 is more likely to encode the production of a polyene, as the poly-β-keto product contains long chains of carbon-carbon double bonds that are characteristic of this class of compounds (**Figure 7.6**). In the future we intend to carry out full transcriptomic analysis of both the wild-type strain and *S. formicae Δfor* using RNA sequencing to identify which clusters are expressed in which strain to narrow down the producing BGC for both the antifungals and the new antibacterial compounds.

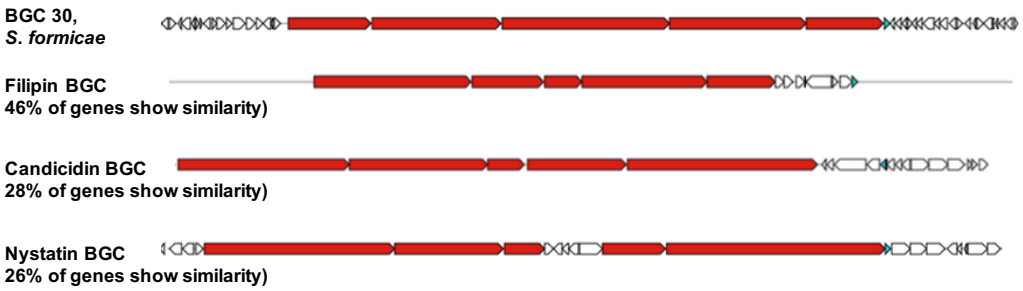


Figure 7.5: AntiSMASH analysis shows that cluster 30, the T1PKS gene cluster encoded within the *S. formicae* genome, shows high percentage homology to other clusters that are known to produce polyene antifungals, particularly in the core PKS genes shown in red. This is therefore thought to be the gene cluster responsible for the production of the antifungal compounds purified from extracts of *S. formicae* during earlier work (**Chapter 1.8**).

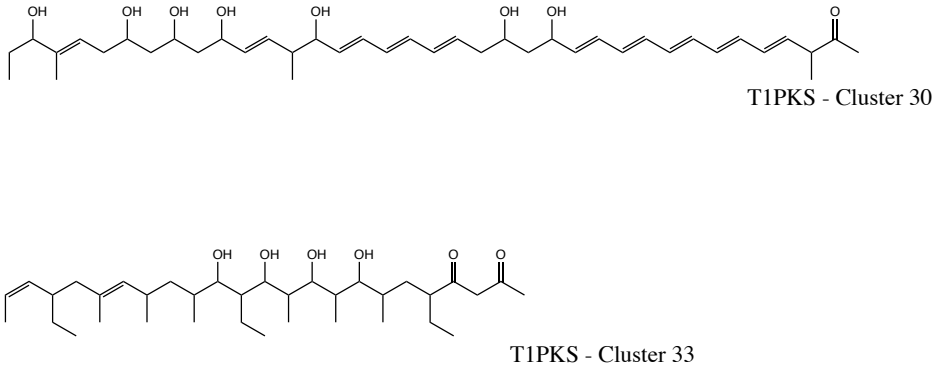


Figure 7.6: Structural predictions of the products encoded by the T1PKSs in gene clusters 30 and 33 in *S. formicae*. From this analysis, it appears that cluster 30 is more likely to synthesise a polyene antifungal, as the poly-β-keto intermediate contains multiple carbon-carbon double bonds which are a common feature of compounds of this class.

7.3 Preliminary results of a genome mining project using CRISPR/Cas9 in *S. formicae*

An alternative method to identify which genes are responsible for the biosynthesis of a secondary metabolite of interest is to genetically disrupt BGCs and examine the effect on compound production in the mutant strain. The linear genomes of *Streptomyces* have been shown to contain the essential genetic machinery in the centre, with BGCs encoding for secondary metabolites usually found on the chromosome arms, therefore deletion of large regions of these non-essential genes does not usually result in developmental defects. For example, large deletions of PKS and NRPS clusters across the genome have been conducted in *Streptomyces coelicolor* using PCR targeting with no detrimental effects on sporulation of growth reported (Zhou *et al.*, 2012). Until the development of CRISPR, this approach would not have been possible in non-model actinomycetes like *S. formicae* as cosmid libraries were not available. In recent years however, CRISPR has been used in several non-model *Streptomyces* species to identify the BGCs responsible for the production of specific metabolites. *Streptomyces* sp. SD85 was isolated from marine sediment in Singapore and shown to produce the polyene macrolactam, sceliphrolactam. The BGC responsible for the biosynthesis of sceliphrolactam was identified using the pCRISPR-Cas9 system to make a small 883 bp deletion within the predicted cluster that inactivated one of the core PKS modules and therefore

abolished sceliphrolactam production (Low *et al.*, 2018). The deletion of the formicamycin BGC in *S. formicae* was the first examples of CRISPR being used to delete a whole BGC of 46 Kb in a non-model actinomycete and represents the largest published CRISPR deletion (46kbp) made in a *Streptomyces* species to date. Following on from this work, we hypothesised that we could use the CRISPR protocols optimised for use in *S. formicae* during this thesis to identify which BGCs are responsible for the production of the new antibacterial and antifungal compounds produced by *S. formicae* and the *S. formicae* Δ*for* mutant by deleting whole BGCs across the genome and examining the effects on compound production in the resulting mutants. To this end, the antiSMASH output from the *S. formicae* genome was manually examined and 8 candidate BGCs were identified that were predicted to encode for either antifungal or antibacterial compounds (**Table 1.1**). These clusters, ranging from 59133 base pairs to 208356 base pairs, were deleted using the pCRISPomyces-2 system as described previously in this thesis. Once the mutants were generated and confirmed, bioassays were run to compare the activity of the mutant strain to the controls. During the generation of these mutants, it was noted that the deletion of large areas of DNA occurs with lower efficiency than smaller deletions like the targeting of individual genes presented earlier in this thesis. This is predicted to be due to the homologous repair being more challenging and therefore less efficient over such large regions of DNA. As a result of this lower efficiency, some of the experiments only resulted in a single, confirmed whole cluster deletion mutant. For this work to be completed, a full repertoire of at least three biological replicates for each strain will need to be generated. Nevertheless, the results of the initial bioactivity assays are shown below with the number of confirmed mutants generated for each cluster indicated (**Figure 7.7**). Representative images of all of the bioassays conducted during this experiment are included in the appendix of this thesis (**Chapter 8.4**).

Colour	Level of inhibition
	Strong (i.e. cleared a small 9mm plate)
	Good
	Average
	Small zone
	No inhibition

Knockout		Bioactivity				
Strain	# of replicates	*B. subtilis*	MRSA	*E. coli*	*C. albicans*	*L. prolificans*
S. formicae	3					
S. formicae Δ5	2					
S. formicae Δ6 (a,b,c,d)	3					
S. formicae Δ8	3					
S. formicae Δ27	4					
S. formicae Δ29 (a,b,c)	2					
S. formicae Δ30	1					
S. formicae Δ31	2					
S. formicae Δ34 (a,b,c,)	3					

Knockout		Bioactivity				
Strain	# of replicates	*B. subtilis*	MRSA	*E. coli*	*C. albicans*	*L. prolificans*
S. formicae Δfor	3					
S. Formicae Δfor Δ5	3					
S. formicae Δfor Δ6 (a,b,c,d)	2					
S. formicae Δfor Δ8	3					
S. formicae Δfor Δ27	2					
S. formicae Δfor Δ29 (a,b,c)	2					
S. formicae Δfor Δ30	1					
S. formicae Δfor Δ31	1					

Figure 7.7: Summary of bioactivity results from whole cluster deletion mutants presented as a heat map of level of inhibition with the number of biological replicates of each strain indicated. Each individual replicate was bioassayed at least 3 times on separate agar plates. All plates were overlayed after 7 days of incubation at 30°C.

In general, these results show that *S. formicae* Δ*for* has greater antibacterial activity against Gram-positive pathogens and stronger antifungal activity than the wild-type strain. This is consistent with observations made during HPLC analysis of the formicamycin cluster deletion mutants conducted earlier in this thesis. None of the strains generated showed any antibacterial activity against *E. coli* suggesting the antibacterials produced will not be effective against Gram-negative pathogens. Interestingly, deleting some of the clusters in the wild-type background seemed to increase antibacterial activity against Gram-positive pathogens. This may be due to

increased carbon flux through the formicamycin BGC when other secondary metabolites are removed. More significantly, deletion of clusters 6 and 31 reduces the bioactivity of *S. formicae ⊿for* against Gram-positive pathogens, implying that both of these clusters produce antimicrobial compounds and the potent bioactivity of *S. formicae ⊿for* noted above may be due to a mixture of molecules from both BGCs (**Figure 7.8**). From these bioassays, it also appears that cluster number 6 is expressed in the wild-type strain under normal laboratory conditions, with the bioactivity of the wild-type strain consisting of a mixture of compounds from cluster 6 and the formicamycin BGC.

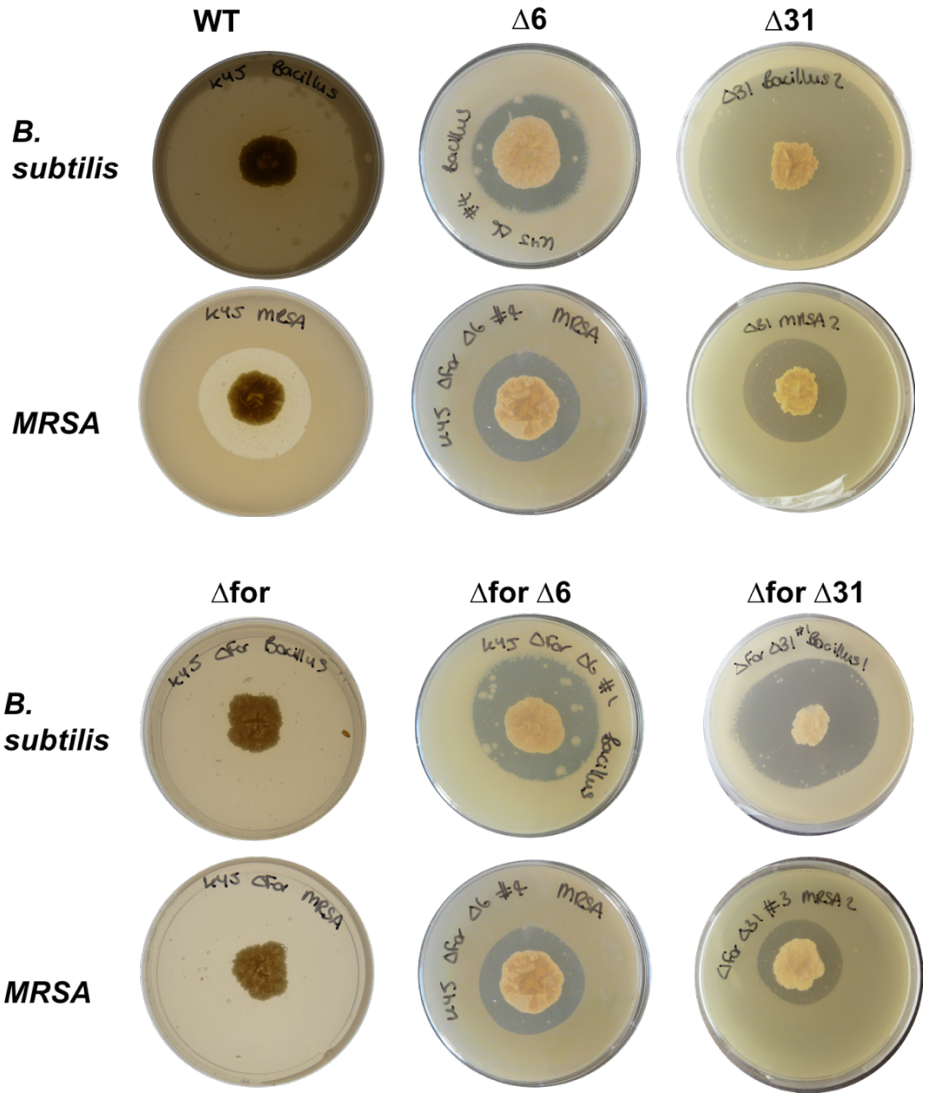


Figure 7.8: Deletion of cluster number 6 and cluster number 31 significantly reduced the antibacterial activity of *S. formicae* and *S. formicae ⊿for* against both *B. subtilis* and MRSA. This suggests that both of these clusters are responsible for producing compounds with antimicrobial activity and the potent bioactivity of *S. formicae ⊿for* is due to a mixture of these secondary metabolites.

Although annotated as one large cluster by AntiSMASH, cluster number 6 is actually thought to consist of 4 smaller clusters in close proximity; two NRPS clusters, a terpene and an aminoglycoside. The terpene synthase is predicted to make 2-methylisoborneol, an odorous secondary metabolite like geosmin with no reported bioactivity (Komatsu *et al.*, 2008). Both non-ribosomal peptides and aminoglycosides have been reported to have antibacterial activity, therefore there is no way to determine which cluster(s) is/are responsible for the production of the antibacterial activity seen here. Early results from Mellisa Davie (undergraduate student in the Wilkinson Laboratory) suggest that there may be low levels of a telomycin-like compound in extracts of *S. formicae Δfor.* It is possible that cluster 6a is responsible for the synthesis of this molecule, however, further investigation is needed (personal communication with Professor Barrie Wilkinson, April 2019). Deletion of cluster 31 in the wild-type strain has little to no effect on the anti- Gram-positive activity, suggesting that this BGC is not expressed in the wild-type under formicamycin producing conditions. However, deletion of cluster 31 in *S. formicae Δfor* significantly reduces the bioactivity, especially against MRSA, implying that the production of this molecule is induced by deletion of the formicamycin BGC. Cluster 31 also encodes for a predicted NRPS and is thought to be responsible for the production of a novel glycopeptide antibiotic as described above (**Chapter 7.2**). If a new glycopeptide antibiotic could be identified from *S. formicae* this could be extremely valuable in the fight against drug-resistant infections. Indeed, analysis of both clusters 6 and 31 shows they have low homology to other BGCs in the database, therefore these compounds are likely to have novel structures. Having identified two more BGCs within *S. formicae* that produce antibacterial compounds and having a strain where these BGCs are switched on, we can now conduct further work on the clusters to identify the compounds produced. We can then use the genetic techniques developed in this thesis to study these BGCs in similar detail to the formicamycin cluster as presented in this thesis. It would also be interesting to generate a *S. formicae Δfor Δ6 Δ31* triple cluster deletion mutant as this strain may either have no antibacterial activity, or more cryptic clusters may be induced in the absence of all three antibiotics.

Deletion of clusters 30 and 31 also appears to reduce the antifungal activity of *S. formicae* although only moderately. As we know what the antifungal peaks look like on an HPLC trace, the whole cluster deletion mutants generated from this work have been analysed by HPLC for comparative metabolomics by Johannes Rassbach (former undergraduate student, JIC) to identify which strains can and cannot produce the antifungal compounds, however, the results were inconclusive. Production of the antifungals is not evident in *S. formicae ⊿30,* however, *S. formicae ⊿29* and *S. formicae ⊿31* also had altered levels of production (personal communication with Johannes Rassbach and Prof Barrie Wilkinson, March 2019). This could imply that cluster 30 is the BGC responsible for production of the antifungal, as predicted previously, and genetic manipulation of neighbouring clusters might affect production levels. It is possible that a regulator encoded in one cluster affects the other neighbouring clusters or, if the borders of the clusters are not called accurately by AntiSMASH, deletion of these neighbouring regions may have removed one or more genes involved in the biosynthesis of the antifungal. Nonetheless, there is only a single replicate for the deletion of cluster 30, therefore further work is needed to confirm that this is the BGC responsible for the biosynthesis of the antifungal compounds.

Although these results are preliminary, this work shows that CRISPR can be used for much larger deletions than previously demonstrated. However, using CRISPR in *S. formicae* to delete such large regions of DNA was less efficient than the generation of smaller deletions presented earlier in this thesis. The Cas9 enzyme has been reported to have toxicity in some strains although this is predicted to be low in *S. formicae* as many CRISPR mutants have been generated in this thesis that have been genetically complemented with no detrimental effects evident in the resulting strains. Genetic engineering with CRISPR can result in off target effects due to the ability of the protospacer to mis-match and allow double strand breaks elsewhere in the genome (Alberti and Corre, 2019). It is not known whether these off target effects become more common when larger regions of DNA are targeted, however, if the homologous repair is less efficient across larger regions of DNA this might allow for more off target events to occur during these experiments. With whole cluster

deletions of the extensive sizes described here it is difficult to genetically complement the deletion without using cosmid libraries or PAC clones which are expensive to generate and screen. Without genetic complementation, it is difficult to identify any off target effects that may have occurred elsewhere in the genome as a result of the CRISPR editing. Although the mutants generated do not display any obvious phenotypic differences to the wild-type controls, it will be necessary to sequence their genomes to confirm the desired edits have been made with no off target effects. However, it is worth noting that off target effects are possible regardless of the method used to introduce genetic mutation and are not just a problem that is specific to CRISPR. As eluded to above, deletion of one cluster may affect the production of a secondary metabolite encoded by another cluster due to the deletion of a cross-cluster regulator, causing complications when applying this method of genome mining. An alternative approach to identifying BGCs using a CRISPR-guided genome mining approach would be to target just the core PKS or NRPS genes or modules. Deleting these core, essential genes would also prevent compound production, without the need for such large genetic deletions and without removing CSRs that may affect other BGCs in the genome. However, the challenge with this approach is that these regions of DNA tend to be highly repetitive, therefore designing unique guide RNAs for this work would be difficult and PCR amplifying DNA regions for the homologous repair template in these regions would likely also be a challenge. A recent study showed that CRISPR systems that utilise the pSG5 replicon, like pCRISPomyces2, are more likely to result in undesired recombination events in repetitive regions such as modular PKS genes. To overcome this problem, the authors of the study developed a new CRISPR system, pMWCas9, using the pIJ101 replicon along with an inducible Cas9, that showed much higher efficiency in the genetic manipulation of challenging DNA sequences (Mo *et al.*, 2019). This system may therefore be more appropriate for this type of genome editing in the future.

Overall however, this preliminary work shows that CRISPR/Cas9 can be used to mine the genomes of talented secondary metabolite producers. Similar approaches could potentially be used in other novel actinomycetes to delete BGCs for known molecules and thus identify novel natural products with clinically relevant bioactivities. Using

these new whole-cluster deletion mutants, we will apply further bioactivity guided metabolomics approaches to identify further new molecules from *S. formicae* in the future. In addition to the antibacterial and antifungal gene clusters identified here, there remain other cryptic BGCs in the *S. formicae* genome. We plan to capture these BGCs and heterologously express them in engineered hosts with less complex secondary metabolite profiles. We also plan to de-repress or activate the BGCs using genetic manipulation of the putative regulatory elements as has been presented in this thesis and elsewhere (Alberti *et al.*, 2019). The majority of potential novel biochemistry encoded in *S. formicae* remains to be discovered. Continued genome mining efforts will be conducted to explore the full potential for secondary metabolite production in this strain. Once novel compounds of interest are identified, their BGCs can be investigated in detail as we have done for the formicamycin gene cluster.

7.4 Final conclusions

S. formicae is a talented strain that is capable of producing multiple novel antimicrobial compounds of clinical interest. The work presented in this thesis shows an in depth characterisation of a T2PKS BGC in *S. formicae* that is responsible for the production of a novel family of polyketides, the formicamycins. We have shown how the formicamycins are biosynthesised by the host, how the pathway is regulated and what the mechanisms of host-resistance are, as well as beginning to define the mechanism of action of these promising new antibiotics. The molecular genetic techniques developed over the course of this project will allow further work to be conducted in order to characterise the additional natural products encoded by this strain. In addition to characterising formicamycin biosynthesis, we have already begun to characterise the products of other BGCs encoded within the genome of *S. formicae*. Our results imply that both clusters 6 and clusters 31 are involved in the production of potent antimicrobials. The peptide produced by cluster 31 seems to have particularly potent inhibitory activity against MRSA and is therefore a priority for further investigation. We also know that *S. formicae* produces antifungal compounds with potent activity against the drug-resistant *L. prolificans*. Confirming

the producing BGC and identifying the chemical structure(s) of the antifungal compound(s) will be the priority for future work. Furthermore, there remains numerous other putative BGCs encoded within the genome of *S. formicae* that appear novel and could be a source of further interesting biochemistry. Now that we have tools available for the genetic manipulation of *S. formicae,* we can investigate these silent clusters either in the host or in a heterologous producer. We predict that *S. formicae* will continue to be a source of novel antibacterials and antifungals in the future. Importantly, this work shows that searching novel environments such as the *Tetraponera* ant nests is a valid method for finding new species with prolific biosynthetic potential. In addition, identifying novel strains increases the likelihood of identifying BGCs with low homology to other known clusters and therefore increases the chances of identifying new antimicrobials like the formicamycins that might eventually be used to treat drug-resistant infections.

8 Appendix

8.1 Strains used and generated during this thesis

Strain	Description	Plasmid	Resistance	Source/Reference
E. coli				
Top10	F– *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara leu*) 7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G			Invitrogen™
BW25113	λ⁻, Δ(*araD-araB*)*567,* Δ*lacZ4787*(*::rrnB-4*)*, lacIp-4000*(*lacIQ*), *rpoS369*(*Am*)*, rph-1,* Δ(*rhaD-rhaB*)*568, hsdR514*	pIJ790	Cml^R	(Datsenko and Wanner, 2000)
DH10β	F– mcrA Δ(mrr-hsdRMSmcrBC) Φ80lacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara leu) 7697 galU galK rpsL nupG λ–			Invitrogen™
ET12567	*dam⁻ dcm⁻ hsdS⁻*	pUZ8002	Cml^R/Tet^R	(MacNeil *et al.*, 1992)
BL21	*fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdSλ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5*			(Studier and Moffattf, 1986)
S. coelicolor				
M1146	Superhost strain: M145 Δact Δred Δcpk Δcda			(Gomez-Escribano and Bibb, 2011)
Bioassay strains				

Bacillus subtilis	Wild-type strain 168		Gift from Nicola Stanley Wall, University of Dundee
E. coli	Wild-type		Lab stock
E. coli NR698	Membrane permeabilised via *imp* deletion		(Ruiz, Kahne and Silhavy, 2006)
Candida albicans	Clinical isolate		From Barrie Wilkinson Gift from Prof Neil Gow, U. Exeter
MRSA	Clinical isolate		Norfolk and Norwich University Hospital (UK)
VRE	Clinical isolate		Norfolk and Norwich University Hospital (UK)
Lamentospora prolificans	Environmental isolate		ATCC Culture collection
Bacillus subtilis PL39 *gyrA*::pMUTIN4 ermC gryA'-lacZ pSPACgyrA+	DNA Gyrase inhibition reporter	Erm^R	Gift from Prof Jeff Errington, University of Newcastle
Bacillus subtilis ypuA::pMUTIN4 ermC ypuA''-lacZ	Cell wall damage reporter	Erm^R	Gift from Prof Jeff Errington, University of Newcastle

*Bacillus subtilis fabHA::*pMUTIN4 ermC fabHA'-lacZ pSPAC-fabHA+	Fatty acid synthesis inhibition reporter		Erm^R	Gift from Prof Jeff Errington, University of Newcastle
Bacillus subtilis ϕ105 (lacZ fusion to a late promoter in a ϕ105 prophage)	DNA damage reporter		Cml^R	Gift from Prof Jeff Errington, University of Newcastle
*Bacillus subtilis helD::*pMUTIN4 helD-lacZ ermC	RNA polymerase inhibition reporter		Erm^R	Gift from Prof Jeff Errington, University of Newcastle
*Bacillus subtilis lial::*pMUTIN4 lial-lacZ ermC	Cell envelope reporter		Erm^R	Gift from Prof Jeff Errington, University of Newcastle
S. Kanamyceticus				
NRRL B2535	ATTC wild-type			Professor Barrie Wilkinson
***S. formicae* KY5**				
Wild-type				Lab stock
KY5001-003	*S. formicae* Δ*for*			This work and Qin *et al.*, 2017
KY5004-006	*S. formicae* Δ*for*: *ΦC31 for* 215-G	*pESAC-13* 215-G	Kan^R/Tsr	This work and Qin *et al.*, 2017

KY5007-010	*S. formicae* Δ*for*V			This work and Qin *et al.*, 2017
KY5011-013	*S. formicae* Δ*for*X			This work
KY5014-016	*S. formicae* Δ*for*Y			This work
KY5017-019	*S. formicae* Δ*for*V: *ΦBT1 forV pforU*	pRD004	Hyg^R	This work and Qin *et al.*, 2017
KY5020-022	*S. formicae ΔforX: ΦBT1 forX pforU*	pRD005	Hyg^R	This work
KY5023-025	*S. formicae ΔforY: ΦBT1 forY pforU*	pRD006	Hyg^R	This work and GenScript
KY5026-028	*S. formicae ΔforD*			This work
KY5029-031	*S. formicae ΔforL*			This work
KY5032-034	*S. formicae ΔforR*			This work
KY5035-037	*S. formicae ΔforS*			This work
KY5038-040	*S. formicae ΔforU*			This work
KY5041-043	*S. formicae ΔforD: ΦBT1 forD pforT*	pRD012	Hyg^R	This work
KY5044-046	*S. formicae ΔforL: ΦBT1 forL pforM*	pRD013	Hyg^R	This work
KY5047-049	*S. formicae ΔforR: ΦBT1 forR*	pRD014	Hyg^R	This work
KY5050-052	*S. formicae ΔforR: ΦBT1 forRA permE**	pRD015	Hyg^R	This work
KY5053-055	*S. formicae* Δ*for*S: *ΦBT1 for*S	pRD016	Hyg^R	This work
KY5056-058	*S. formicae* Δ*for*S: *ΦBT1 for*S *permE**	pRD017	Hyg^R	This work

KY5059-061	*S. formicae* Δ*for*U: *ΦBT1 for*U	pRD018	Hyg^R	
KY5062-064	*S. formicae* Δ*for*U: *ΦBT1 for*UV *permE**	pRD019	Hyg^R	This work
KY5065-067	*S. formicae ΔforM*			This work
KY5068-070	*S. formicae ΔforT*			This work
KY5071-073	*S. formicae ΔforW*			This work
KY5074-076	*S. formicae ΔforM: ΦBT1 forM pforM*	pRD023	Hyg^R	This work
KY5077-079	*S. formicae ΔforT: ΦBT1 forT pforT*	pRD023	Hyg^R	This work
KY5080-082	*S. formicae ΔforT: ΦBT1 forT permE**			
KY5083-085	*S. formicae ΔforW: ΦBT1 forW pforU*	pRD025	Hyg^R	This work
KY5086-088	*S. formicae ΔforJ*			This work
KY5089-091	*S. formicae ΔforGF*			This work
KY5092-094	*S. formicae ΔforZ*			This work
KY595-097	*S. formicae ΔforBBCC*			This work
KY5098-100	*S. formicae ΔforJ: ΦBT1 forJ pforM*	pRD030	Hyg^R	This work
KY5101-103	*S. formicae ΔforGF: ΦBT1 forGF pforG*	pRD031	Hyg^R	This work
KY5104-106	*S. formicae ΔforZ: ΦBT1 forZ pforZ*	pRD032	Hyg^R	This work
KY5107-109	*S. formicae ΔforBBCC: ΦBT1 forBBCC pforBB*	pRD033	Hyg^R	This work

KY5110-112	*S. formicae ΔforJ: ΦBT1 forJ 3x Flag*	pRD034	Hyg^R	This work
KY5113-115	*S. formicae ΔforGF: ΦBT1 forGF 3x Flag*	pRD035	Hyg^R	This work
KY5116-118	*S. formicae ΔforZ: ΦBT1 forZ 3x Flag*	pRD036	Hyg^R	This work
KY5119-121	*S. formicae Δfor:* ΦC31 *for 1-4 aac(3)IV*	pRD037	Apr^R	This work
KY5122-124	*S. formicae Δfor:* ΦC31 *for 1-7 aac(3)IV*	pRD038	Apr^R	This work
KY5125-127	*S. formicae Δfor:* ΦC31 *for 32-43 aac(3)IV*	pRD039	Apr^R	This work
KY5128-130	*S. formicae Δ for:* ΦC31 *for 36-43 aac(3)IV*	pRD040	Apr^R	This work
KY5131-133	*S. formicae Δfor*AA			This work
KY5134-136	*S. formicae ΔforK*			This work
KY5137-139	*S. formicae Δfor*AA: ΦC31 f*orAA pforAA*	pRD043	Hyg^R	This work
KY5140-142	*S. formicae ΔforK*: ΦC31 f*orK pforM*	pRD044	Hyg^R	This work
KY5143-145	*S. formicae Δfor*AA, *ΔforK*			This work
KY5146-148	*S. formicae ΔforJ*: ΦC31 f*orK pforM*	pRD046	Hyg^R	This work
KY5149-151	*S. formicae ΔforJ*: ΦC31 f*orAA pforAA*	pRD047	Hyg^R	This work
KY5152-154	*S. formicae ΔforJ, ΔforX*			This work
KY5155-157	*S. formicae ΔforJ, ΔforV*			This work
KY5158-160	*S. formicae ΔforJ, ΔforY*			This work

KY5161-163	*S. formicae ΔforJ, ΔforS*	This work
KY5164-165	*S. formicae Δ5*	This work
KY5166-168	*S. formicae Δfor Δ5*	This work
KY5169-171	*S. formicae Δ6*	This work
KY5172-173	*S. formicae Δfor Δ6*	This work
KY5174-176	*S. formicae Δ8*	This work
KY5177-179	*S. formicae Δfor Δ8*	This work
KY5180-183	*S. formicae Δ27*	This work
KY5184-185	*S. formicae Δfor Δ27*	This work
KY5186-187	*S. formicae Δ29*	This work
KY5188-189	*S. formicae Δfor Δ29*	This work
KY5190	*S. formicae Δ30*	This work
KY5191	*S. formicae Δfor Δ30*	This work
KY5192-KY5193	*S. formicae Δ31*	This work
KY5194	*S. formicae Δfor Δ31*	This work
KY5195-KY5197	*S. formicae Δ34*	This work

8.2 Plasmids used and generated during this thesis

Plasmid	Description	Resistance	Reference
pUZ8002	RK2 derivative with a mutation in *oriT*	Kan^R	(Keiser *et al.*, 2000)
pMS82	*ori,* pUC18, *hyg, oriT,* RK2, int ΦBT1	Hyg^R	(Gregory & Smith, 2003)
pIJ773	- *aac(3)IV oriT bla*	Apr^R	(Gust *et al.*, 2004)
pIJ790	*araC-Parab,Y, β, exo, cat, repA1001ts, oriR101*	Cml^R	(Gust *et al.*, 2004)
pR9604	pUB307 derivative	$Carb^R$	(Piffaretti, Arini and Frey, 1988)
pIJ10257	*oriT*, ΦBT1 *attB-int*, Hygr, *ermEp**, pMS81 backbone	Hyg^R	(Hong *et al.*, 2005)
pESAC-13 215-G	*aph*II, *tsr*	Kan^R/Tsr^R	BioS&T and Qin *et al.*, 2017
pCRISPomyces-2	Apr^R, *oriT,* $rep^{pSG5(ts)}$, ori^{ColE1}, *sSpcas9,* synthetic guide RNA cassette	Apr^R	(Cobb *et al.*, 2015)
pIJ12900	pGM1190 with *kan* resistance marker	Kan^R	Professor Mervyn Bibb
BCG30	pCRISPomyces-2 BCG30 flanking DNA and gRNA	Apr^R	This work and Qin *et al.*, 2017
pGS-21a *forV*	pGS-21a with codon optimised *forV* cloned between NcoI and XhoI	Amp^R	GenScript®
pRD001	pCRISPomyces-2 *forV* flanking DNA and gRNA	Apr^R	This work and Qin *et al.*, 2017

pRD002	pCRISPomyces-2 *forX* flanking DNA and gRNA	Apr^R	This work
pRD003	pCRISPomyces-2 *forY* flanking DNA and gRNA	Apr^R	This work
pRD004	pMS82 *pforU forV*	Hyg^R	This work
pRD005	pMS82 *pforU forX*	Hyg^R	This work
pRD006	pMS82 *pforU forY*	Hyg^R	This work, GenScript®
pRD007	pCRISPomyces-2 *forD* flanking DNA and gRNA	Apr^R	This work
pRD008	pCRISPomyces-2 *forL* flanking DNA and gRNA	Apr^R	This work
pRD009	pCRISPomyces-2 *forR* flanking DNA and gRNA	Apr^R	This work
pRD010	pCRISPomyces-2 *forS* flanking DNA and gRNA	Apr^R	This work
pRD011	pCRISPomyces-2 f*orU* flanking DNA and gRNA	Apr^R	This work
pRD012	pMS82 *pforT* f*orD*	Hyg^R	This work
pRD013	pMS82 *pforM* f*orL*	Hyg^R	This work
pRD014	pMS82 *pforT forR*	Hyg^R	This work
pRD015	pIJ10257 f*orR-forA*	Hyg^R	This work
pRD016	pMS82 *pforT forS*	Hyg^R	This work

pRD017	pIJ10257 f*orS*	Hyg^R	This work
pRD018	pMS82 *pforU* forU	Hyg^R	This work
pRD019	pIJ10257 f*orU-forV*	Hyg^R	This work
pRD020	pCRISPomyces-2 f*orM* flanking DNA and gRNA	Apr^R	This work
pRD021	pCRISPomyces-2 f*orT* flanking DNA and gRNA	Apr^R	This work
pRD022	pCRISPomyces-2 f*orW* flanking DNA and gRNA	Apr^R	This work
pRD023	pMS82 *pforM* f*orM*	Hyg^R	This work
pRD024	pMS82 *pforT* f*orT*	Hyg^R	This work
pRD025	pMS82 *pforU* f*orW*	Hyg^R	This work
pRD026	pCRISPomyces-2 f*orJ* flanking DNA and gRNA	Apr^R	This work
pRD027	pCRISPomyces-2 f*orGF* flanking DNA and gRNA	Apr^R	This work
pRD028	pCRISPomyces-2 *ForZ* flanking DNA and gRNA	Apr^R	This work
pRD029	pCRISPomyces-2 f*orBBCC* flanking DNA and gRNA	Apr^R	This work
pRD030	pMS82 *pforM* f*orJ*	Hyg^R	This work
pRD031	pMS82 *pforG* f*orGF*	Hyg^R	This work

pRD032	pMS82 *pforZ* f*orZ*	Hyg^R	This work
pRD033	pMS82 *pfor*BB f*orBBCC*	Hyg^R	This work
pRD034	pMS82 *pforM* f*orJ 3x Flag*	Hyg^R	This work
pRD035	pMS82 *pforG* f*orGF 3x Flag*	Hyg^R	This work
pRD036	pMS82 *pforZ forZ 3x Flag*	Hyg^R	This work
pRD037	*pESAC-13* 215-G *1-4 aac(3)IV oriT*	Kan^R/Tsr^R	This work
pRD038	*pESAC-13* 215-G *1-7 aac(3)IV oriT*	Kan^R/Tsr^R	This work
pRD039	*pESAC-13* 215-G *32-43 aac(3)IV oriT*	Kan^R/Tsr^R	This work
pRD040	*pESAC-13* 215-G *36-43 aac(3)IV oriT*	Kan^R/Tsr^R	This work
pRD041	pCRISPomyces-2 *forK* flanking DNA and gRNA	Apr^R	This work
pRD042	pCRISPomyces-2 *forAA* flanking DNA and gRNA	Apr^R	This work
pRD043	pMS82 pForAA *forAA*	Hyg^R	This work
pRD044	pMS82 pForM *forK*	Hyg^R	This work
pRD045	pCRISPomyces-2 *Cluster 5* flanking DNA and gRNA	Apr^R	This work
pRD046	pCRISPomyces-2 *Cluster 6* flanking DNA and gRNA	Apr^R	This work

pRD047	pCRISPomyces-2 *Cluster 8* flanking DNA and gRNA	Apr^R	This work
pRD048	pCRISPomyces-2 *Cluster 27* flanking DNA and gRNA	Apr^R	This work
pRD049	pCRISPomyces-2 *Cluster 29* flanking DNA and gRNA	Apr^R	This work
pRD050	pCRISPomyces-2 *Cluster 30* flanking DNA and gRNA	Apr^R	This work
pRD051	pCRISPomyces-2 *Cluster 31* flanking DNA and gRNA	Apr^R	This work
pRD052	pCRISPomyces-2 *Cluster 34* flanking DNA and gRNA	Apr^R	This work
pRD053	pMF96 *pforM* GUS	Hyg^R	This work
pRD054	pMF96 *pforH-GF* GUS	Hyg^R	This work
pRD055	pMF96 *pforGF-H* GUS	Hyg^R	This work
pRD056	pMF96 *pforT-U* GUS	Hyg^R	This work
pRD057	pMF96 *pforU-T* GUS	Hyg^R	This work
pRD058	pMF96 *pforZ-AA* GUS	Hyg^R	This work
pRD059	pMF96 *pforAA-Z* GUS	Hyg^R	This work
pRD060	pCRISPomyces-2 ForJ S. Kanamyceticus flanking DNA and gRNA	Apr^R	This work
pRD061	pIJ12900 ForJ:Apr in S. Kanamyceticus	Kan^R, Apr^R	This work

8.3 Primers used and generated during this thesis

Name	Description	Sequence
pCRISP Test F	Test XbaI site pCRISP2 For	AGGCTAGTCCGTTATCAACTTGAAA
pCRISP Test R	Test XbaI site pCRISP2 Rev	TCGCCACCTCTGACTTGAGCGTCGA
Spacer test	Test BbsI site of pCRISP2	ATACGGCTGCCAGATAAGGC
pIJ10257 TEST F	Test MCS in pIJ10257 For	GATCTTGACGGCTGGCGAGAG
pIJ10257 TEST R	Test MCS in pIJ10257 Rev	GCGTCAGCATATCATCAGCGAGC
KY5001 35KOFor1	pCRISPomyces-2 left flank BGC Template For	ATCATCTAGAAAGGACATTCGC CTCGTCAGCCGCAAG
KY5002 35KORev1	pCRISPomyces-2 left flank BGC Template Rev	GCTGCTGCGACCAGGCGAGCTCGCGTCGAGACGCAACTCAGTGAAACCTTG
KY5003 35KOFor2	pCRISPomyces-2 right flank BGC Template For	GCGAGCTCGCCTGGTCGCAGCAGCGTACTGACAGACAATTTCTCACGTTCGGC
KY5004 35KORev2	pCRISPoyces-2 right flank BGC Template Rev	ACGTTCTAGAGAGGAACTCCTCATAGGTGATCAGATAACC
KY5005 gRNA35 For	BGC gRNA For	ACGCGCGCCATGAAGCTAAGG AGG
KY5006 gRNA35 Rev	BGC gRNA Rev	AAACCCTCCTTAGCTTCATGGC GC
KY5007 BGC_Left_Fwd	pESAC13 5' edge forward test primer for BGC	ACAGGTACGACGGGTCC
KY5008 BGC_Left_Rev	pESAC13 5' edge reverse test primer for BGC	CCCAACCAGTACGCGAAG

KY5009 BGC_Right_Fwd	pESAC13 3' edge forward test primer for BGC	GCATGGGATGTGAGCACC
KY5010 BGC_Right_Rev	pESAC13 3' edge reverse test primer for BGC	AAGAGGCGATGAGCGAGG
KY5011 BGC_Mid_Fwd	pESAC13 forward test primer for BGC centre	TACCACATCGGCGAGTCC
KY5012 BGC_Mid_Rev	pESAC13 reverse test primer for BGC centre	CGCTCCAGGTTGTACGAC
KY5013_6548	Confirmation of cluster deletion	ATCGGTGAGATCACCATGACTA CGG
KY5014_6507	Confirmation of cluster deletion	GTTCGACGGTGCCGA TGAAGC
KY5015_6505	Confirmation of cluster deletion	CTGTACGCTGACAGCCGGAAC
KY5016_6514	Detection of cluster edge	GGCGAAGAGGCGGGCGATCTCG
KY5017_6512	Detection of cluster edge	CACGACAGACCCCTCCCGCGT
KY5018_6538	Detection of cluster edge	CATTCCCGGGGCCCGGGTGT
KY5019_6537	Detection of cluster edge	AGCCGACGGCGTATCGGCTGA CG
RD125	ForV deletion For 1	gctcggttgccgccgggcgttttttaTCTAGAgacgagcacggatacgtgatcgg
RD126	ForV deletion Rev 1	GCGAGCTCGCCTGGTCGCAGCAGCgtgacggcacaggagcagagc
RD127	ForV deletion For 2	GCTGCTGCGACCAGGCGAGCTCGCtcactcatcgccgttcccttcc
RD128	ForV deletion Rev 2	gcaacgcggccttttacggttcctggccTCTAGAcgacgtgatgcggctcgactg
RD129	ForV gRNA F	ACGCacgaccagtccttgatggta
RD130	ForV gRNA R	AAACtaccatcaaggactggtcgt
RD037	ForV Test 1F	gcgcgaacgcctggagaacatgg
RD038	ForV Test 2R	gccgtcgccgagcgggatgaa
RD059	ForV Test 2F	gtcaaggcgcacgaggagctg
RD060	ForV Test 1R	gactcgttcagggcgaagtgcc
JTM431	ForX deletion For 1	gctcggttgccgccgggcgttttttatctagagccggtgacggcacaggagc
JTM432	ForX deletion Rev 1	GCTGCTGCGACCAGGCGAGCTCGCcatcgtgcttcctcactcctgggtg
RD019	ForX deletion For 2	GCGAGCTCGCCTGGTCGCAGCAGCtgacgagcccctctgttctcgtcgccc
RD020	ForX deletion Rev 2	gcaacgcggccttttacggttcctggccTCTAGAgccctcggcccgcaaggacttca
RD131	ForX gRNA For	ACGCctcgccctgttcgggccggta
RD132	ForX gRNA Rev	AAACtaccggcccgaacagggcgag
RD049	ForX Test 1F	ggtccgctggaacgtcgaacc
RD050	ForX Test 2R	gcagcgctcaggcgtgaagtc

RD053	ForX Test 2F	gcccgctgtctgcacaactg
RD054	ForX Test 1R	cgatccgtccgcctcggtac
RD133	ForY deletion For 1	gctcggttgccgccgggcgttttttaTCTAGAggtgccgaagcgttccacggcc
RD134	ForY deletion Rev 1	GCGAGCTCGCCTGGTCGCAGCAGCtcaccgctcgcacgccgcca
RD135	ForY deletion For 2	GCTGCTGCGACCAGGCGAGCTCGCggtcgtacacgccgctcgga
RD136	ForY deletion Rev 2	gcaacgcggccttttacggttcctggccTCTAGAcctggaagatggccgcggtcc
RD137	ForY gRNA For	ACGCggctcctcgcggcgccggta
RD138	ForY gRNA Rev	AAACtaccggcgccgcgaggagcc
RD139	ForY Test 1F	cgaagaaggtgagggagagcagcc
RD140	ForY Test 2R	cgcgatccaggacgcccacaacc
RD055	ForY Test 2F	gacgccttgcgcaggctgcac
RD056	ForY Test 1R	ctggccgagcggctctcgctc
RD079	ForJ deletion For1	gctcggttgccgccgggcgttttttaTCTAGAggtgtgcgcgaagaacggcc
RD080	ForJ deletion Rev 1	GCTGCTGCGACCAGGCGAGCTCGCcactgacgcggtcgttcccg
RD081	ForJ deletion For 2	GCGAGCTCGCCTGGTCGCAGCAGCtgacgtgcttcgagaccgcc
RD082	ForJ deletion Rev 2	gcaacgcggccttttacggttcctggccTCTAGAcctcttcatgttcctggtgggcc
RD167	ForJ gRNA For	ACGCtgccgacaccttctccatga
RD168	ForJ gRNA Rev	AAACtcatggagaaggtgtcggca
RD085	ForJ Test 1F	cctcttcggtgagcgcttcgagg
RD086	ForJ Test 2R	cctgttggacttcgcgcaggc
RD087	ForJ Test 2F	gtacgccaggaggacgtgcg
RD088	ForJ Test 1R	gccgacgcggcacttctatcc
RD089	ForZ deletion For1	gctcggttgccgccgggcgttttttaTCTAGAcgaacaggccgacgctgaacag
RD090	ForZ deletion Rev 1	GCTGCTGCGACCAGGCGAGCTCGCcatggcttgaagtccagcacgtcc
RD091	ForZ deletion For 2	GCGAGCTCGCCTGGTCGCAGCAGCtcatccgtacctggcagctcgtcg
RD092	ForZ deletion Rev 2	gcaacgcggccttttacggttcctggccTCTAGAccgaggcggacggatcgcgtcc
RD169	ForZ gRNA For	ACGCgtcggcggtcaactcgactg
RD170	ForZ gRNA Rev	AAACcagtcgagttgaccgccgac
RD095	ForZ Test 1F	gccggtgccgaaccggacgc
RD096	ForZ Test 2R	cgcacgccgccacgacgagc
RD097	ForZ Test 2F	cgcacgccgccacgacgagc
RD098	ForZ Test 1R	cgcacgccgccacgacgagc
RD099	ForGF deletion For1	gctcggttgccgccgggcgttttttaTCTAGAggagccggtcttggccatctgc

RD100	FoGF deletion Rev 1	GCTGCTGCGACCAGGCGAGCTCGCggcagcctcgttcacagcagc
RD101	ForGF deletion For 2	GCGAGCTCGCCTGGTCGCAGCAGCtgaggctcaggcgggttcgatgg
RD102	ForGF deletion Rev 2	gcaacgcggccttttacggttcctggccTCTAGAcgagatcgtcatccacgcgcc
RD157	Cluster 8 deletion For 1	gctcggttgccgccgggcgtttttaTCTAGAggacatcgccaacgccatcgcc
RD158	Cluster 8 deletion Rev 1	GCTGCTGCGACCAGGCGAGCTCGCcacgcgcacgctcctcggtcg
RD159	Cluster 8 deletion For 2	GCGAGCTCGCCTGGTCGCAGCAGCggccgacgagagagggtttg
RD160	Cluster 8 deletion Rev 2	gcaacgcggccttttacggttcctggccTCTAGAcgaaggcttcgttgacctcg
RD161	Cluster 8 gRNA F	ACGCtcgccgtccgcagaccgtta
RD162	Cluster 8 gRNA R	AAACtaacggtctgcggacggcga
RD163	Cluster 8 Test 1F	ggacttcgaggagttccaggc
RD164	Cluster 8 Test 1R	ggtcctgctaccagtgagtagg
RD165	Cluster 8 Test 2F	ggtcgatggcggcatcttcc
RD166	Cluster 8 Test 2R	gccgtgcacgttgaccttctc
RD171	ForGF gRNA For	ACGCtggcgaagatgttgcgcaga
RD172	ForGF gRNA Rev	AAACtctgcgcaacatcttcgcca
RD105	ForGF Test 1F	gcagttcctggacgatgcgc
RD106	ForGF Test 1R	cgagggtctggagaacgcgc
RD107	ForGF Test 2F	cgtcggcaccttctaccaccg
RD108	ForGF Test 2R	gcctgcgtgattcatcggctg
RD109	ForBBCC deletion For1	gctcggttgccgccgggcgtttttaTCTAGAgcgcagcgagcggatcttgatcc
RD110	ForBBCC deletion Rev 1	GCTGCTGCGACCAGGCGAGCTCGCcacggacgggcgctcgacac
RD111	ForBBCC deletion For 2	GCGAGCTCGCCTGGTCGCAGCAGCtgacgccgggcgtcaccacc
RD112	ForBBCC deletion Rev 2	gcaacgcggccttttacggttcctggccTCTAGAgctcgtcggcgtcgactggcg
RD173	ForBBCC gRNA For	ACGCtgatcgactcctggatgata
RD174	ForBBCC gRNA Rev	AAACtatcatccaggagtcgatca
RD115	ForBBCC Test 1F	cacaccatcgacaccagctggac
RD116	ForBBCC Test 1R	cggcgtatcggctgacgtcg
RD117	ForBBCC Test 2F	cgagtccgaggagcatgcgc
RD118	ForBBCC Test 2R	ctggacggcgttcggactcg
RD187	pMS82-formicamycin promoter For	gccgagaaccTAGGATCCAAGCTTcatggtgaggtgctcctcctg
RD188	Formicamycin promoter-ForV Rev	ggcgctgctggtcactcatgtggagctgccctcactc
RD189	Formicamycin promoter-ForV For	gagtgagggcagctccacatgagtgaccagcagcgcc
RD190	ForV-pMS82 Rev	CTGGTACCATGCATAGATCTAAGCTTtcaccggccccgctccacgg

RD191	Formicamycin promoter-ForX Rev	gcacggatacgtgatcggtcatgtggagctgccctcactc
RD192	Formicamycin promoter-ForX For	gagtgagggcagctccacatgaccgatcacgtatccgtgc
RD193	ForX-pMS82 Rev	CTGGTACCATGCATAGATCTAAGCTTtcacggccgcctcccgttcc
pMS82 TEST For	Test HindIII site pMS82 For	gcaacagtgccgttgatcgtgctatg
pMS82 TEST Rev	Test HindIII site pMS82 Rev	gccagtggtatttatgtcaacaccgcc
RD175	ReDirect - oriT 1F	accacccgcaccgcccacctcaaggagagcagccccatgGCTCACGGTAACTGATGCCG
RD176	ReDirect 4R	ctgacccagcagtggcaggccgaccacaccgccttcatgTGTAGGCTGGAGCTGCTTC
RD177	ReDirect 8R	gctaagtcggtcgatctgcgcctcttggcgcggtgatcaTGTAGGCTGGAGCTGCTTC
RD178	ReDirect - oriT 34F	ggccgtacaaccagggtcgacacccgggccccgggaatgGCTCACGGTAACTGATGCCG
RD179	ReDirect – oriT 36F	tgtccgccgccgggcggcggggagtgaggagagagcgtgGCTCACGGTAACTGATGCCG
RD180	ReDirect 43R	ccgtcgtacctgtccgtcccgctgcgcgagcagtgatcaTGTAGGCTGGAGCTGCTTC
RD205	ForD deletion For 1	gctcggttgccgccgggcgtttttttaTCTAGAccagttcggccacggactgc
RD206	ForD deletion Rev 1	GCTGCTGCGACCAGGCGAGCTCGCcacggtcaggctcccttcg
RD207	ForD deletion For 2	GCGAGCTCGCCTGGTCGCAGCAGCgtatgagcgccaccgaggcc
RD208	ForD deletion Rev 2	gcaacgcggccttttttacggttcctggccTCTAGAccaacggccagacggcgcgtc
RD209	ForD gRNA For	ACGCtggatgcgcatgaacgcgaa
RD210	ForD gRNA Rev	ACGCtggatgcgcatgaacgcgaa
RD211	ForD Test 1F	ccacgctggtcgaacagtgc
RD212	ForD Test 1R	gcgagtctgaccaggcgctc
RD213	ForD Test 2F	cgtcctggtcaccgacgagg
RD214	ForD Test 2R	gcatggcgcaggtgcacacc
RD215	ForL deletion For 1	gctcggttgccgccgggcgtttttttaTCTAGAgcagcacgccgaccagcacg
RD216	ForL deletion Rev 1	GCTGCTGCGACCAGGCGAGCTCGCtagtcccgccatggcggacc
RD217	ForL deletion For 2	GCGAGCTCGCCTGGTCGCAGCAGCcatggacgaactccctttcgcc
RD218	ForL deletion Rev 2	gcaacgcggccttttttacggttcctggccTCTAGAgctaaggaggtggccgagg
RD219	ForL gRNA For	ACGCgagccccaactgccttggta
RD220	ForL gRNA Rev	AAACtaccaaggcagttggggctc
RD221	ForL Test 1F	cgagggcgagcagcaggcg
RD222	ForL Test 1R	cggcacgcgagttcggtggc
RD223	ForL Test 2F	cgctcggtcgccacggcc
RD224	ForL Test 2R	ccgcgtgatgacagatgcgcc
RD225	For M deletion For 1	gctcggttgccgccgggcgtttttttaTCTAGAggtcacgctcatggcgacgc
RD226	ForM deletion Rev 1	GCTGCTGCGACCAGGCGAGCTCGCtaacgcgagattccgcaaggcg

RD227	ForM deletion For 2	GCGAGCTCGCCTGGTCGCAGCAGCcacaccggctcccatcggttg
RD228	ForM deletion Rev 2	gcaacgcggccttttacggttcctggccTCTAGAcgaaggtgttgtcgagcctgatcc
RD229	ForM gRNA For	ACGCgtcgaactccccgtcttcgta
RD230	ForM gRNA Rev	AAACtacgaagacggggagttcgac
RD231	ForM Test 1F	cgtcgtggaaggtgaagagc
RD232	ForM Test 1R	ccgaggcgcgtcaatgagc
RD233	ForM Test 2F	cgagcaccgcgagtgatccg
RD234	ForM Test 2R	cacggtcgcgccgccgatcc
RD235	ForR deletion For 1	gctcggttgccgccgggcgttttttaTCTAGActtcggcaagcagttcttcg
RD236	ForR deletion Rev 1	GCTGCTGCGACCAGGCGAGCTCGCggtcatggttctccttgtcc
RD237	ForR deletion For 2	GCGAGCTCGCCTGGTCGCAGCAGCacatgacccggcaggtcgcc
RD238	ForR deletion Rev 2	gcaacgcggccttttacggttcctggccTCTAGAccgagccgtgcgcgttgacg
RD239	ForR gRNA For	ACGCcgtagaggaactcctcggagta
RD240	ForR gRNA Rev	AAACtactccgaggagttcctctacg
RD241	ForR Test 1F	gcacatacgccatcaggtcgc
RD242	ForR Test 1R	ggtccatgcgttgggcctcg
RD243	ForR Test 2F	cggcgagcgggtcttcgac
RD244	ForR Test 2R	cgtgccggtcgttctgcttgg
RD245	ForS deletion For 1	gctcggttgccgccgggcgttttttaTCTAGAgaggagcacctcaccatgacg
RD246	ForS deletion Rev 1	GCTGCTGCGACCAGGCGAGCTCGCgctcatacggtgccctccac
RD247	ForS deletion For 2	GCGAGCTCGCCTGGTCGCAGCAGCtgatccgacccaacggggac
RD248	ForS deletion Rev 2	gcaacgcggccttttacggttcctggccTCTAGAgaagcgccggtgatcatgacg
RD249	ForS gRNA For	ACGCgggtttgtggacgtaggtga
RD250	ForS gRNA Rev	AAACtcacctacgtccacaaaccc
RD251	ForS Test 1F	cctcgtcgaggacagcacgg
RD252	ForS Test 1R	cgatgcgtacgtcgacgttgc
RD253	ForS Test 2F	gcctcttgcgcggcttgagc
RD254	ForS Test 2R	cgcgtcgaagcaggagacgg
RD255	ForT deletion For 1	gctcggttgccgccgggcgttttttaTCTAGAcctggaagtacgcccagacg
RD256	ForT deletion Rev 1	GCTGCTGCGACCAGGCGAGCTCGCcatggtgaggtgctcctcctg
RD257	ForT deletion For 2	GCGAGCTCGCCTGGTCGCAGCAGCgtatgagccaggaggagccg
RD258	ForT deletion Rev 2	gcaacgcggccttttacggttcctggccTCTAGAcggatcgaagtcgcactcgg

RD259	ForT gRNA For	ACGCcgcagcaggcgctccaacgt
RD260	ForT gRNA Rev	AAACacgttggagcgcctgctgcg
RD261	ForT Test 1F	gctcctcgtgcgccttgacg
RD262	ForT Test 1R	cgctggtgatgagccattcg
RD263	ForT Test 2F	ctcgatcaggacgttgccgc
RD264	ForT Test 2R	ctgcacgtagcggtccatgcg
RD265	ForU deletion For 1	gctcggttgccgccgggcgtttttttaTCTAGAcgtagtactcgccgcggttc
RD266	ForU deletion Rev 1	GCTGCTGCGACCAGGCGAGCTCGCgatgagtgaccagcagcgcc
RD267	ForU deletion For 2	GCGAGCTCGCCTGGTCGCAGCAGCcatgtggagctgccctcactc
RD268	ForU deletion Rev 2	gcaacgcggcctttttacggttcctggccTCTAGAgcttcttgccgccgacgttg
RD269	ForU gRNA For	ACGCgtggccctcttgagctacgt
RD270	ForU gRNA Rev	AAACacgtagctcaagagggccac
RD271	ForU Test 1F	ggtccactgcggcaggtcg
RD272	ForU Test 1R	gcgcggttcctcgtcgagcg
RD273	ForU Test 2F	ggtcacgaagtcgacgtcgg
RD274	ForU Test 2R	cggtagccgtggtaggtgagg
RD275	ForW deletion For 1	gctcggttgccgccgggcgtttttttaTCTAGAcgtcggcggcatcacatgg
RD276	ForW deletion Rev 1	GCTGCTGCGACCAGGCGAGCTCGCtgaggaagcacgatgaccgatc
RD278	ForW deletion For 2	GCGAGCTCGCCTGGTCGCAGCAGCctcctgtgccgtcaccggcc
RD279	ForW deletion Rev 2	gcaacgcggcctttttacggttcctggccTCTAGAccttctccttcgccgacggc
RD279	ForW gRNA For	ACGCtggtcgaggtagtcgaagta
RD280	ForW gRNA Rev	AAACtacttcgactacctcgacca
RD281	ForW Test 1F	gcagcacttccatcgagcgg
RD282	ForW Test 1R	cgctacgaggagagctaccg
RD283	ForW Test 2F	ggccatcttccaggcgagg
RD284	ForW Test 2R	gctcgtgcggaagtacggc
RD297	ForJ compl pMS82 prom F	gccgagaaccTAGGATCCAAGCTTgatgccggtgagcagggcgag
RD298	ForJ compl prom-gene R	ggcgccgtggtcgtggtcataccggctcccatcggttgctg
RD299	ForJ compl prom-gene F	cagcaaccgatgggagccggtatgaccacgaccacggcgcc
RD300	ForJ compl gene-pMS82 R	CTGGTACCATGCATAGATCTAAGCTTgcggaggcggaccgtgcctag
RD301	ForZ compl pMS82 F	gccgagaaccTAGGATCCAAGCTTccggtcaccacccattggag
RD302	ForZ compl pMS82 R	CTGGTACCATGCATAGATCTAAGCTTaggagttgtgcgccctcgc
RD303	ForGF compl pMS82 F	gccgagaaccTAGGATCCAAGCTTcgtgtacccccctgtgcacg

RD304	ForGF compl pMS82 R	CTGGTACCATGCATAGATCTAAGCTTccgctgctcgccatcgaac
RD305	ForBBCC compl pMS82 F	gccgagaaccTAGGATCCAAGCTTtcgtcttgcccgcgccgttc
RD306	ForBBCC compl pMS82 R	CTGGTACCATGCATAGATCTAAGCTTtcagcggtccgcggagaccag
RD307	ForD pMS82 pF	gccgagaaccTAGGATCCAAGCTTgtggagctgccctcactctc
RD308	ForD pMS82 pR	cgttctcggtgagctcgggcacggtgaggtgctcctcctg
RD309	ForD pMS82 gF	caggaggagcacctcaccgtgcccgagctcaccgagaacg
RD310	ForD pMS82 gR	CTGGTACCATGCATAGATCTAAGCTTcctcggtggcgctcatacgg
RD311	ForL pMS82 pF	gccgagaaccTAGGATCCAAGCTTgattcttcggcgcacgacag
RD312	ForL pMS82 pR	cgacgatcaacgtggtgtgcataccggctcccatcggttgc
RD313	ForL pMS82 gF	gcaaccgatgggagccggtatgcacaccacgttgatcgtcg
RD314	ForL pMS82 gR	CTGGTACCATGCATAGATCTAAGCTTctactcgacggggactacgc
RD315	ForM pMS82 pF	gccgagaaccTAGGATCCAAGCTTgattcttcggcgcacgacag
RD316	ForM pMS82 gR	CTGGTACCATGCATAGATCTAAGCTTttatggaactcgggcgatcgc
RD317	ForR pMS82 pF	gccgagaaccTAGGATCCAAGCTTgtggagctgccctcactctc
RD318	ForR pMS82 pR	CGCACGGTGTGCGTGGTCATggtgaggtgctcctcctgt
RD319	ForR pMS82 gF	acaggaggagcacctcaccATGACCACGCACACCGTGCG
RD320	ForR pMS82 gR	CTGGTACCATGCATAGATCTAAGCTTTCATGTCCCGTACCGCTCTTCGG
RD321	ForS pMS82 pF	gccgagaaccTAGGATCCAAGCTTgtggagctgccctcactctc
RD322	ForS pMS82 pR	cggctcctcctggctcatggtgaggtgctcctcctgta
RD323	ForS pMS82 gF	tacaggaggagcacctcaccatgagccaggaggagccg
RD324	ForS pMS82 gR	CTGGTACCATGCATAGATCTAAGCTTcgttgggtcggatcagagg
RD325	ForT pMS82 pF	gccgagaaccTAGGATCCAAGCTTgtggagctgccctcactctc
RD326	ForT pMS82 pR	CTGGTACCATGCATAGATCTAAGCTTtcatacggtgccctccacc
RD327	ForU pMS82 pF	gccgagaaccTAGGATCCAAGCTTggtgaggtgctcctcctgtaaag
RD328	ForU pMS82 gR	CTGGTACCATGCATAGATCTAAGCTTtcactcatcgccgttcccttc
RD329	ForW pMS82 pF	gccgagaaccTAGGATCCAAGCTTggtgaggtgctcctcctgtaaag
RD330	ForW pMS82 pR	ctctgctcctgtgccgtcacgtggagctgccctcactc
RD331	ForW pMS82 gF	gagtgagggcagctccacgtgacggcacaggagcagag
RD332	ForW pMS82 gR	CTGGTACCATGCATAGATCTAAGCTTgcttcctcactcctgggtgac
RD337	ForJ-FLAG Rev	gcctgaaccgcctccaccgtgccccgcgggcacctg
RD338	ForJ-FLAG For	caggtgcccgcggggcacggtggaggcggttcaggc
RD339	FLAG-pMS82 R	CTGGTACCATGCATAGATCTAAGCTTtcaCTTGTCGTCATCGTCCTTG
RD340	pMS82 ForF prom F	gccgagaaccTAGGATCCAAGCTTtcgtgtacccccctgtgcacg

RD341	ForF-prom R	ggtcaccacggtctgcatagcagcctccccggttcg
RD342	ForF F	cgaaccggggaggctgctatgcagaccgtggtgacc
RD343	ForF-FLAG R	gcctgaaccgcctccaccgccccggtcgccctgcg
RD344	FLAG-pMS82 For	cgcagggcgaccggggcggtggaggcggttcaggc
RD345	ForZ-FLAG Rev	gcctgaaccgcctccaccccgctcgcacgccgccacg
RD346	ForZ-FLAG For	cgtggcggcgtgcgagcggggtggaggcggttcaggc
RD347	Prom-ForA Rev	cgttcacgccttgggcacccatgtggagctgccctcactc
RD350	ForAA deletion For 1	gctcggttgccgccgggcgttttttaTCTAGAgctgttgcgcaggctgtgg
RD351	ForAA deletion Rev 1	GCTGCTGCGACCAGGCGAGCTCGCcatgacggcctcccggattgc
RD352	ForAA deletion For 2	GCGAGCTCGCCTGGTCGCAGCAGCtgacgcccggcgtcagcg
RD353	ForAA deletion Rev 2	gcaacgcggccttttacggttcctggccTCTAGAcgacctcgacaggatcgcc
RD354	ForAA gRNA F	ACGCccggtaactcgttcacgcctt
RD355	ForAA gRNA R	AAACaaggcgtgaacgagttaccgg
RD356	ForAA Test 1F	cggcggcaggctcaccgtcg
RD357	ForAA Test 1R	cgagtacgtctatgccgcgc
RD358	ForAA Test 2F	cgagctgactggtggtcgtcg
RD359	ForAA Test 2R	gtcagccgatacgccgtcgg
RD360	Cluster 31 deletion For 1	gctcggttgccgccgggcgttttttaTCTAGAgctcgaccgcggctatctgg
RD361	Cluster 31 deletion Rev 1	GCTGCTGCGACCAGGCGAGCTCGCcatgactaatccgtcgcatggc
RD362	Cluster 31 deletion For 2	GCGAGCTCGCCTGGTCGCAGCAGCcatgtgcggatcatcgatggaagc
RD363	Cluster 31 deletion Rev 2	gcaacgcggccttttacggttcctggccTCTAGAcgtcctggtaccagaccgtcc
RD364	Cluster 31 gRNA F	ACGCggtgaggacggaacgggtta
RD365	Cluster 31 gRNA R	AAACtaacccgttccgtcctcacc
RD366	Cluster 31 Test 1F	cgctacctcgacctggtcgc
RD367	Cluster 31 Test 1R	gtgtgtgtcgaagcggcatg
RD368	Cluster 31 Test 2F	cgacgaggacttcctcgacg
RD369	Cluster 31 Test 2R	cgatcagcgctggtacgagc
RD381	Cluster 5 deletion For 1	gctcggttgccgccgggcgttttttaTCTAGAggacctggatcgccgagatcg
RD382	Cluster 5 deletion Rev 1	GCTGCTGCGACCAGGCGAGCTCGCgacaaggagagcacaggccg
RD383	Cluster 5 deletion For 2	GCGAGCTCGCCTGGTCGCAGCAGCccaggtgccgaacgtgctg
RD384	Cluster 5 deletion Rev 2	gcaacgcggccttttacggttcctggccTCTAGActctcgtgcttgctccgcac
RD385	Cluster 5 gRNA F	ACGCacacaccaccaccacacgaa
RD386	Cluster 5 gRNA R	AAACttcgtgtggtggtggtgtgt

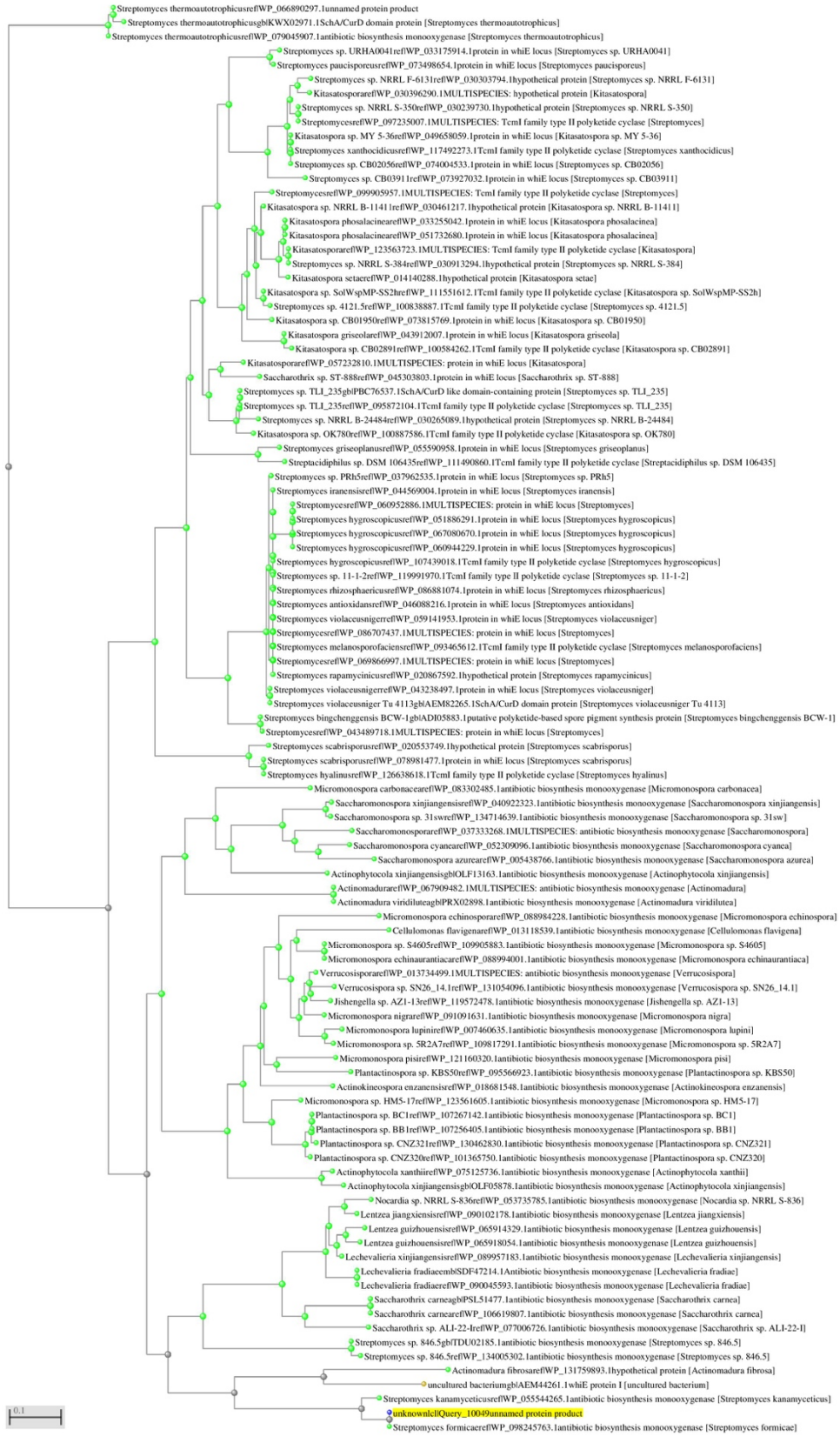
RD387	Cluster 5 Test 1F	gacgaccgcgatcatcaggag
RD388	Cluster 5 Test 1R	gatcgtcgaccgcaacctcgtc
RD389	Cluster 5 Test 2F	gtggtgcaagcggtggtgac
RD390	Cluster 5 Test 2R	ggtaccggtccgctcactcc
RD391	Cluster 6 deletion For 1	gctcggttgccgccgggcgttttttaTCTAGAgcaggtccgtctcaccgacc
RD392	Cluster 6 deletion Rev 1	GCTGCTGCGACCAGGCGAGCTCGCcacatcgagcggtacctctagc
RD393	Cluster 6 deletion For 2	GCGAGCTCGCCTGGTCGCAGCAGCgacgcgggcatgaactgaag
RD394	Cluster 6 deletion Rev 2	gcaacgcggccttttttacggttcctggccTCTAGAggcggtgagcagcacgagg
RD395	Cluster 6 gRNA F	ACGCcggagccggacatcgtggtaa
RD396	Cluster 6 gRNA R	AAACttaccacgatgtccggctccg
RD397	Cluster 6 Test 1F	ccggcgtggccgagaagg
RD398	Cluster 6 Test 1R	ccacctccgcggaggtcag
RD399	Cluster 6 Test 2F	ccacgctcaccgtcgacgg
RD400	Cluster 6 Test 2R	cggcagctcctgcaccagg
RD441	Cluster 27 deletion For 1	gctcggttgccgccgggcgttttttaTCTAGAtcgccgaggcacaggacgtg
RD442	Cluster 27 deletion Rev 1	GCTGCTGCGACCAGGCGAGCTCGCcgtggcgaagtcctggatgg
RD443	Cluster 27 deletion For 2	GCGAGCTCGCCTGGTCGCAGCAGCtgacgccttgatcaccgcgc
RD444	Cluster 27 deletion Rev 2	gcaacgcggccttttttacggttcctggccTCTAGAggagcgcatgttgatcgtgc
RD445	Cluster 27 gRNA F	ACGCaggcgtcgccgagggaataa
RD446	Cluster 27 gRNA R	AAACttattccctcggcgacgcct
RD447	Cluster 27 Test 1F	ggtgccgcacgtcaagacg
RD448	Cluster 27 Test 1R	gtgcgaatgtcgatctgcacg
RD449	Cluster 27 Test 2F	gcactcgacgtggccgacg
RD450	Cluster 27 Test 2R	cgaccaggtcaggcgcatc
RD451	Cluster 29 deletion For 1	gctcggttgccgccgggcgttttttaTCTAGAcgtacacggcggtgacctcc
RD452	Cluster 29 deletion Rev 1	GCTGCTGCGACCAGGCGAGCTCGCcggctcgaaggcttccgtg
RD453	Cluster 29 deletion For 2	GCGAGCTCGCCTGGTCGCAGCAGCgacggcgatgacggtgtgg
RD454	Cluster 29 deletion Rev 2	gcaacgcggccttttttacggttcctggccTCTAGAcgacttcctcggctgcgacg
RD455	Cluster 29 gRNA F	ACGCgcgcctcggatcaggcttaa
RD456	Cluster 29 gRNA R	AAACttaagcctgatccgaggcg
RD457	Cluster 29 Test 1F	ggtagtcgtcgtgctcggtgg
RD458	Cluster 29 Test 1R	ccttgacctcaccatcgagcg
RD459	Cluster 29 Test 2F	cgacgacggtcacgtcgtacg

RD460	Cluster 29 Test 2R	gcacatctacctcctgcacgc
RD461	Cluster 30 deletion For 1	gctcggttgccgccgggcgtttttttaTCTAGAtgccatcggcatcggcag
RD462	Cluster 30 deletion Rev 1	GCTGCTGCGACCAGGCGAGCTCGCcggagtccgctgagcctgtg
RD463	Cluster 30 deletion For 2	GCGAGCTCGCCTGGTCGCAGCAGCcgaatgcattctcgtccttcaatgg
RD464	Cluster 30 deletion Rev 2	gcaacgcggccttttttacggttcctggccTCTAGAccacacggtcacgtgcacg
RD465	Cluster 30 gRNA F	ACGCtgaacaacggccatgccctaa
RD466	Cluster 30 gRNA R	AAACttagggcatggccgttgttca
RD467	Cluster 30 Test 1F	ggacgcggtccgaaccagg
RD468	Cluster 30 Test 1R	ggccgtattggagcaatccgg
RD469	Cluster 30 Test 2F	cggagtccacttcggtgtgc
RD480	Cluster 30 Test 2R	cgtacgtggagtcgctgcc
RD481	Cluster 34 deletion For 1	gctcggttgccgccgggcgtttttttaTCTAGAggtgtcgctctcggtgagc
RD482	Cluster 34 deletion Rev 1	GCTGCTGCGACCAGGCGAGCTCGCccgtctccgtcgtccgaac
RD483	Cluster 34 deletion For 2	GCGAGCTCGCCTGGTCGCAGCAGCcatgcggcgtcgacggatac
RD484	Cluster 34 deletion Rev 2	gcaacgcggccttttttacggttcctggccTCTAGAgcgatgctgcacgtgctcg
RD485	Cluster 34 gRNA F	ACGCgcgttccgccacgagagaat
RD486	Cluster 34 gRNA R	AAACattctctcgtggcggaacgc
RD487	Cluster 34 Test 1F	ccgtcggcagcccgttgag
RD488	Cluster 34 Test 1R	gcccgacgtgacgtcactg
RD489	Cluster 34 Test 2F	cgacagatcgccgacaccg
RD490	Cluster 34 Test 2R	cgttgcggttcgccgcctc
RD491	ForJ F	gcaaggcggcgcagagcg
RD492	ForJ R	gccgacaccttctccatgagg
RD493	ForZ F	gaaccggacgcagccgcag
RD394	ForZ R	cctcgacgcgtgccacgag
RD495	ForF F	cctcgacgcgtgccacgag
RD496	ForF R	gcgaccagggtcatgacctcg
RD497	ForK deletion For 1	gctcggttgccgccgggcgtttttttaTCTAGAgcatccggtcaccgagatgg
RD498	ForK deletion Rev 1	GCTGCTGCGACCAGGCGAGCTCGCcactgacttctgcaatgattgacc

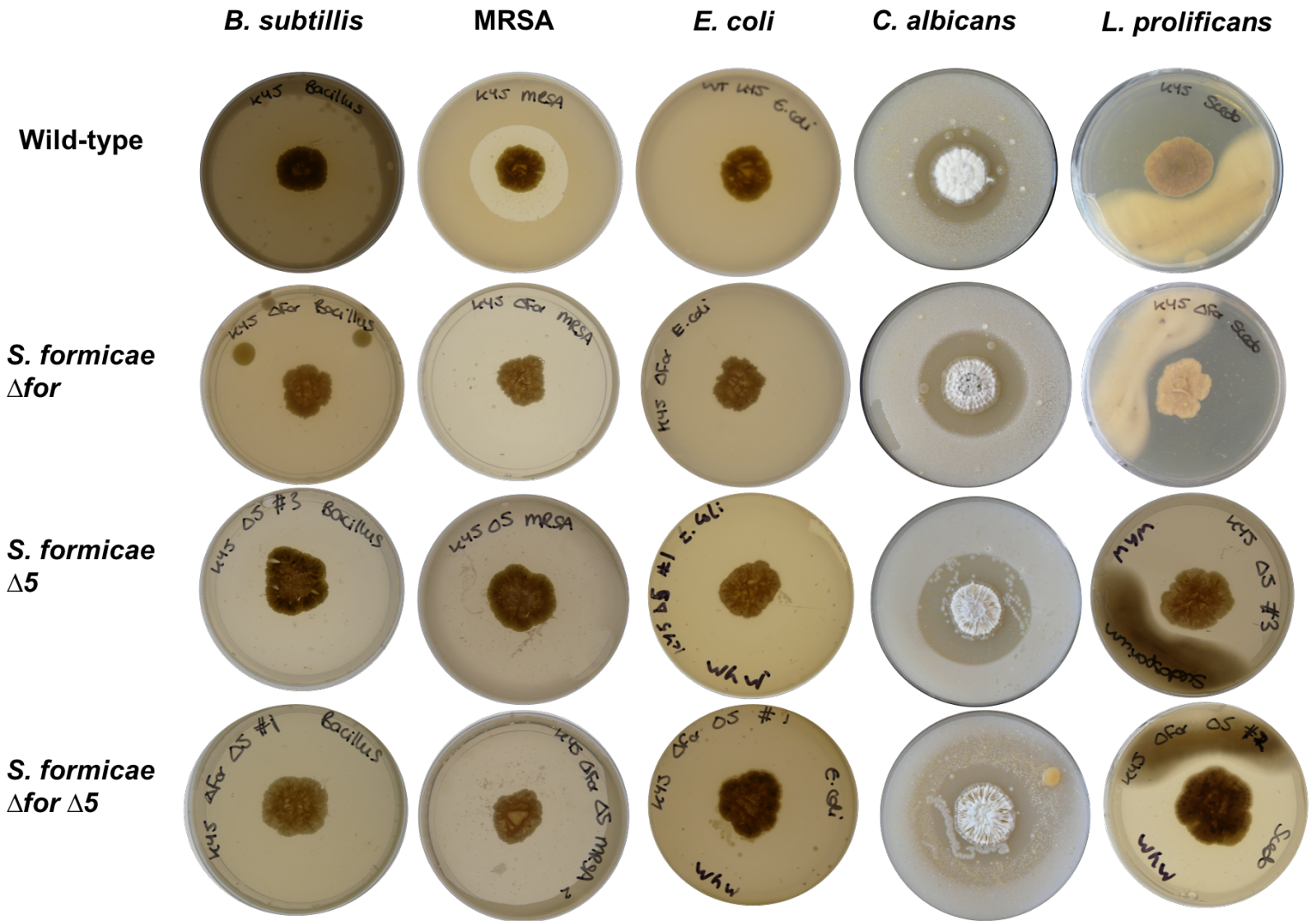
RD499	ForK deletion For 2	GCGAGCTCGCCTGGTCGCAGCAGCcatggcgggactactcgacg
RD500	ForK deletion Rev 2	gcaacgcggcctttttacggttcctggccTCTAGAccagttccagtcgtccttcgagg
RD501	ForK gRNA F	ACGCtccgcatccggcatcccttg
RD502	ForK gRNA R	AAACcaagggatgccggatgcgga
RD503	ForK Test 1F	gaggtcaactgccgtatccagg
RD504	ForK Test 1R	cgtcggcgacgcagttctacc
RD505	ForK Test 2F	ctgtgacgtgcttcgagaccg
RD506	ForK Test 2R	gccggttatgggaccaaggc
RD522	ForAA permE* pIJ10257 F	AAAAAcatatgcgtccggtgctgtactgcaag
RD508	ForAA permE* pIJ10257 R	AAAAAaagctttcaccacccttgcgcatgttcacg
RD509	ForK permE* pIJ10257 F	AAAAAcatatgatggcggaccacaagggatgc
RD510	ForK permE* pIJ10257 R	AAAAAaagctttcagtggtggatggatagaagtgcc
RD512	ForR-ForA ErmE* F	AAAAAcatatgatgaccacgcacaccgtgc
RD513	ForR-ForA ErmE* R	AAAAAaagctttcacctcagctccctccggtc
RD514	ForT ErmE* F	AAAAAcatatgatgacgtcaactcccttgcg
RD515	ForT ErmE* R	AAAAAaagctttcatacggtgccctccacc
RD516	ForS ErmE* F	AAAAAcatatgatgagccaggaggagccgc
RD517	ForS ErmE* R	AAAAAaagctttcagaggggtgctgccgtg
RD518	ForU-ForV ErmE* F	AAAAAcatatgatggccgagatctccgccg
RD519	ForU-ForV ErmE* R	AAAAAaagcttcgctctgctcctgtgccgtc
RD520	ForW ErmE* F	AAAAAcatatggtgacggcacaggagcagag
RD521	ForW ErmE* R	AAAAAaagcttcgtgcttcctcactcctggg
RD523	S. Kan ForJ deletion For 1	gctcggttgccgccgggcgttttttaTCTAGAGCGATGTACCGGACGCTCGC
RD524	S. Kan ForJ deletion Rev 1	GCTGCTGCGACCAGGCGAGCTCGCCATGTTCGCTCACCTCTGCTGTG
RD525	S. Kan ForJ deletion For 2	GCGAGCTCGCCTGGTCGCAGCAGCCGCTGACGCGGTCGTTCC
RD526	S. Kan ForJ deletion Rev 2	gcaacgcggcctttttacggttcctggccTCTAGACGAGCGGCTCGGCGACAAGC
RD527	S. Kan ForJ gRNA F	ACGCccaccttcgtgtgcatgcga
RD528	S. Kan ForJ gRNA R	AAACtcgcatgcacacgaaggtgg
RD529	S. Kan ForJ Test 1F	CTCGGCATCGGGCTCGGCTG
RD530	S. Kan ForJ Test 1R	GACGCGTCCGTTCCTGCTCC
RD531	S. Kan ForJ Test 2F	GCGGACGCGGCACTTCTATCC
RD532	S. Kan ForJ Test 2R	CGAAGAACGGCCTCGCCTTCG
RD543	pMF96 HindIII Test For	GCTCAATCAATCACCGGATCC

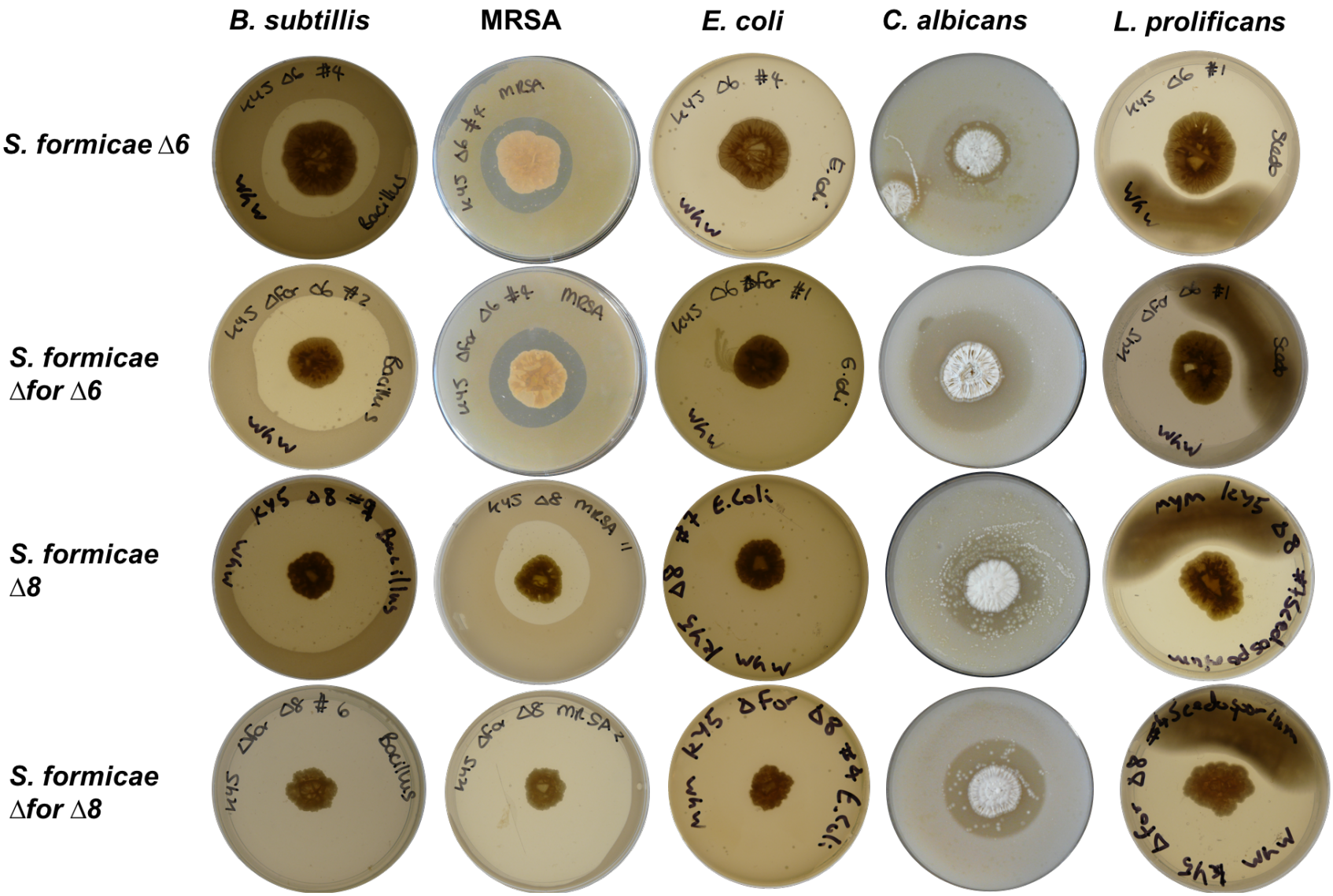
RD544	pMF96 HindIII Test Rev	CATGTCCGTACCTCCGTTG
RD545	pforM For	AAAAAcatatgtcttcggcgcacgacagacc
RD546	pforM Rev	AAAAActcgagaccggctcccatcggttgc
RD547	pforH-GF For	AAAAAcatatggcgctgctcacggtcatcg
RD548	pforH-GF Rev	AAAAActcgaggcagcctcgttcacagcag
RD549	pforGF-H For	AAAAAcatatggcagcctcgttcacagcagc
RD550	pforGF-H Rev	AAAAActcgaggcgctgctcacggtcatcg
RD551	pforT-U For	AAAAAcatatgggcgaagccgaggctcatgc
RD552	pforT-U Rev	AAAAActcgagcagatcgaccagcttctgctggtcc
RD553	pforU-T For	AAAAAcatatgcagatcgaccagcttctgctggtcc
RD554	pforU-T Rev	AAAAActcgagggcgaagccgaggctcatgc
RD559	pforZ-AA For	AAAAAcatatggaatccctgacgcgccgcg
RD560	pforZ-AA Rev	AAAAActcgaggacgatggtggtgtcgagcac
RD561	pforAA-Z For	AAAAAcatatggacgatggtggtgtcgagcacc
RD562	pforAA-Z Rev	AAAAActcgaggaatccctgacgcgccgcg
RD575	ForK pForM R	catcccttgtggtccgccataccggctcccatcggttgc
RD576	ForK F	gcaaccgatgggagccggtatggcggaccacaagggatg
RD577	ForK R pMS82	CTGGTACCATGCATAGATCTAAGCTTtcagtggtggatggatagaagtgcc
RD578	pForAA-ForAA F pMS82	gccgagaaccTAGGATCCAAGCTTggcttgaagtccagcacgtcc
RD579	pForAA-ForAA R pMS82	CTGGTACCATGCATAGATCTAAGCTTtcaccacccttgcgcatgttc
RD596	S. Kan ForJ 1.1 For	gaagggagcggacatatgaagcttGTGATGACGGATGCGCCATG
RD597	S. Kan ForJ 1.2 Rev	CGGCATCAGTTACCGTGAGCCATGTTCGCTCACCTCTGCTGTG
RD598	S. Kan ForJ Apr For	CACAGCAGAGGTGAGCGAACATGGCTCACGGTAACTGATGCCG
RD599	S. Kan ForJ Apr Rev	GACACCTTCTCCAGGAGCGCGTGTAGGCTGGAGCTGCTTC
RD600	S. Kan ForJ 2.1 For	GAAGCAGCTCCAGCCTACACGCGCTCCTGGAGAAGGTGTC
RD601	S. Kan ForJ 2.2 Rev	ctcggtacccggggatcctctagaGCAGATGGCCAAGACCGGCTC
RD602	pGM1190 Apr ForJ Test For	CCTGACCGAACTGGTCATCC
RD603	pGM1190 Apr ForJ Test Rev	CGGATCCGTGCTTCAGGTCG
RD607	Apr-oriT ForM For	atgaacaccctgcaataccgctttgacgggccggaagacGCTCACGGTAACTGATGCCG
RD608	Apr-oriT ForX Rev	ctcctgggtgacgggggcggcttcgacgaggtggagccgTGTAGGCTGGAGCTGCTTC
RD611	ForM-X Apr Test For	catcagtcacgagatgtcac
RD612	ForM-X Apr Test Rev	cgatgaccgatcacgtatcc

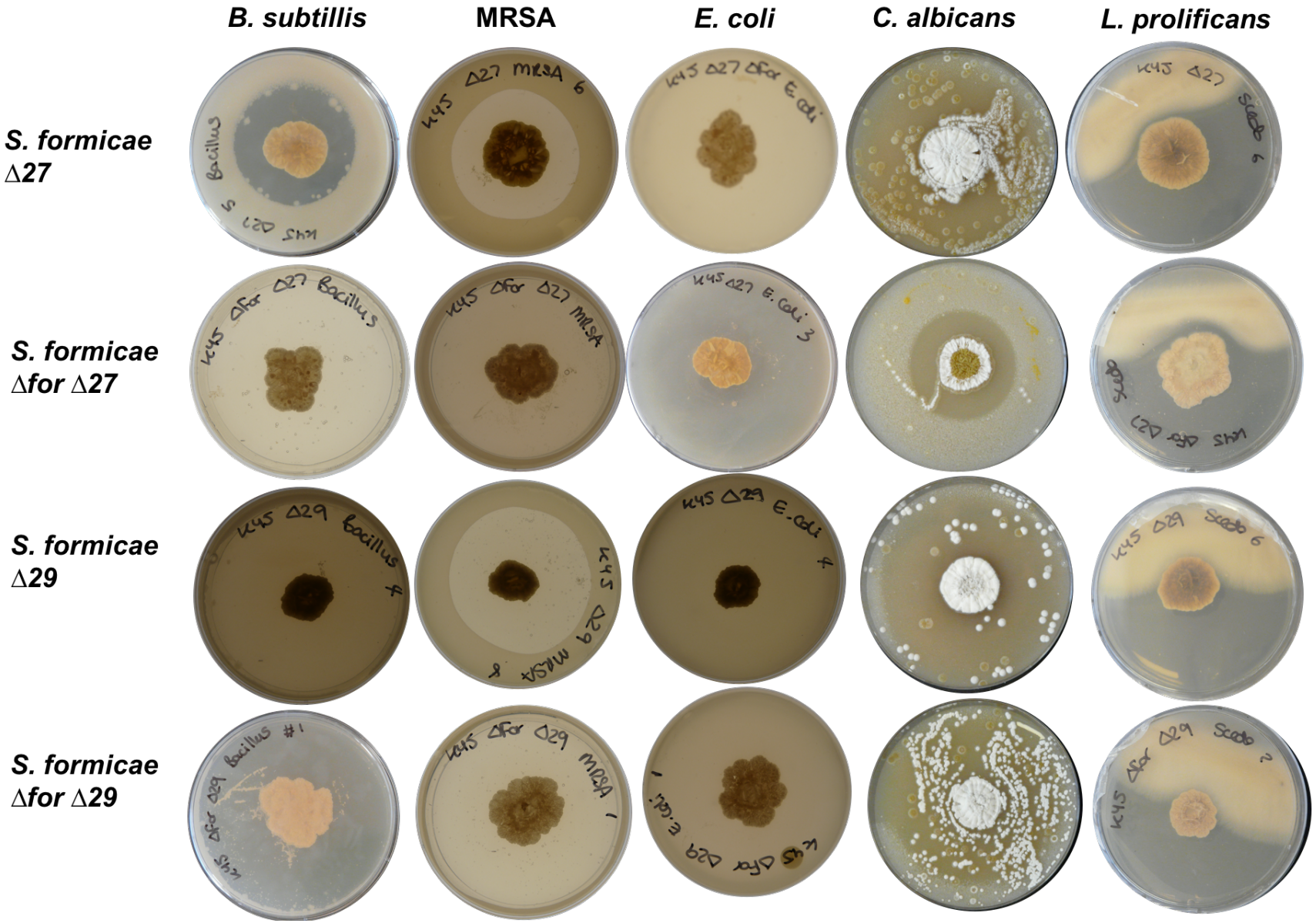
8.4 Phylogenetic analysis of ForS and similar ABM domain containing proteins

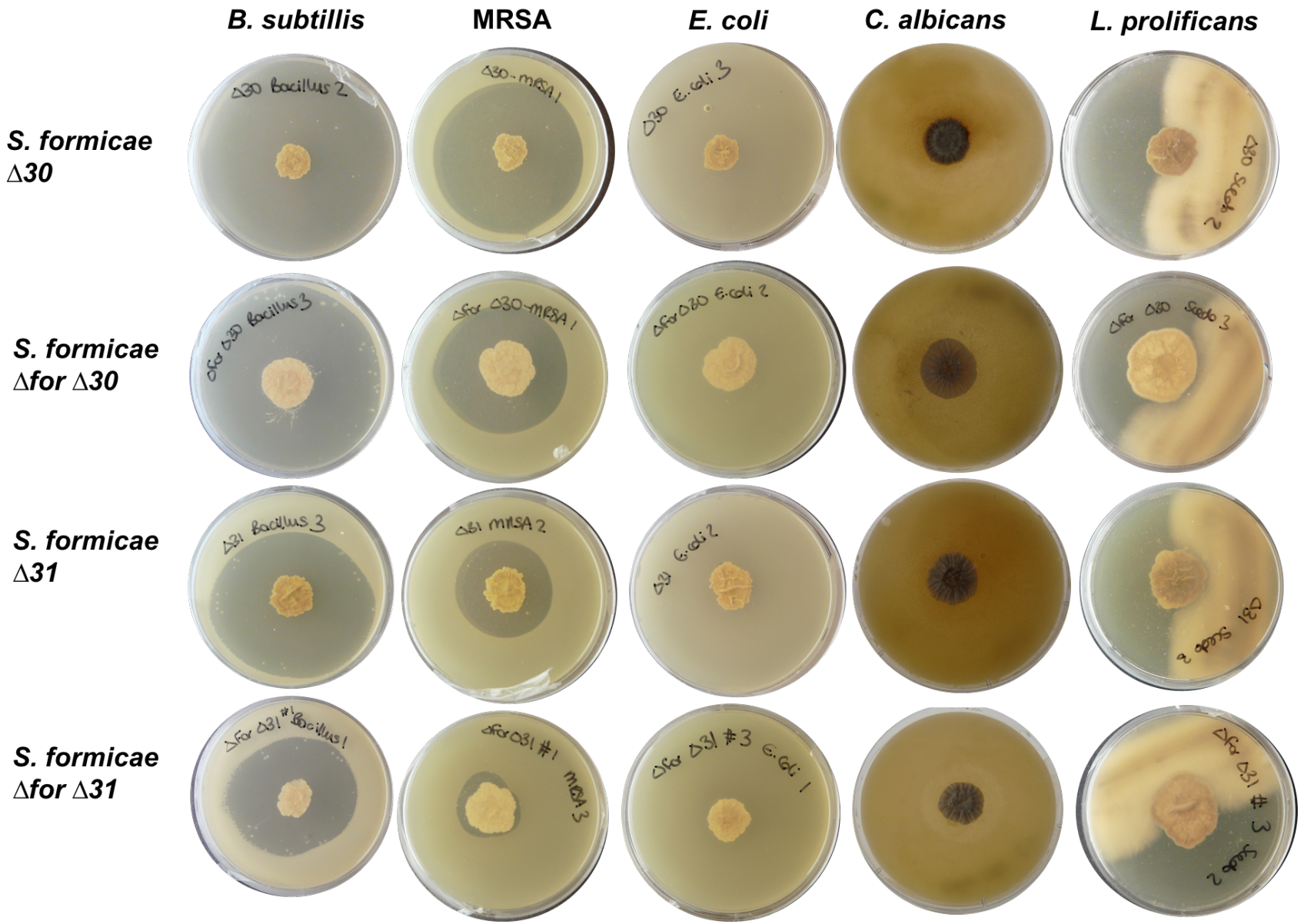


8.5 Bioassay images from CRISPR/Cas9 genome mining project




B. subtillis
MRSA
E. coli
C. albicans
L. prolificans
S. formicae Δ6
S. formicae Δfor Δ6
S. formicae Δ8
S. formicae Δfor Δ8


B. subtillis
MRSA
E. coli
C. albicans
L. prolificans
S. formicae Δ27
S. formicae Δfor Δ27
S. formicae Δ29
S. formicae Δfor Δ29


B. subtillis
MRSA
E. coli
C. albicans
L. prolificans
S. formicae Δ30
S. formicae Δfor Δ30
S. formicae Δ31
S. formicae Δfor Δ31

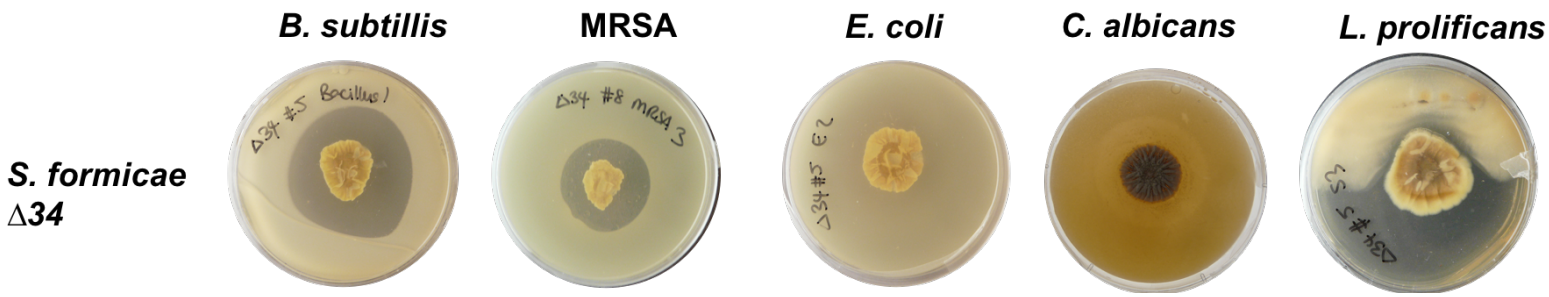


Figure 8.1: Bioassays of whole-cluster deletion mutants were set up against *B. subtilis*, MRSA, *E. coli, C. albicans* and *L. prolificans*. At least three replicates f each assay were set up. The images presented here are a representative summary of the result for each strain.

9 References

Abdelmohsen, UR, Grkovic, T., Balasubramanian, S., Kamel, MS, Quinn, RJ and Hentschel, U. (2015) 'Elicitation of secondary metabolism in actinomycetes', *Biotechnology Advances*, 33(6), pp. 798–811.

Agudelo, Daniel, Bourassa, Philippe, Bérubé, Gervais and Tajmir-Riahi, Heidar-Ali (2014) 'Intercalation of antitumor drug doxorubicin and its analogue by DNA duplex: Structural features and biological implications', *International Journal of Biological Macromolecules*, 66, pp. 144–150.

Aigle, B. and Corre, C. (2012) 'Waking up Streptomyces Secondary Metabolism by Constitutive Expression of Activators or Genetic Disruption of Repressors', *Methods in Enzymology*, 517, pp. 343–366.

Alanis, AJ (2005) 'Resistance to Antibiotics: Are We in the Post-Antibiotic Era?', *Archives of Medical Research*, 36(6), pp. 697–705.

Alanjary, M., Steinke, K. and Ziemert, N. (2019) 'AutoMLST: an automated web server for generating multi-locus species trees highlighting natural product potential', *Nucleic Acids Research*, online, pp. 1–7.

Alberti, F. and Corre, C. (2019) 'Editing streptomycete genomes in the CRISPR/Cas9 age', *Nat. Prod. Rep*, Advance.

Alberti, F., Leng, DJ, Wilkening, I., Song, L., Tosin, M. and Corre, C. (2019) 'Triggering the expression of a silent gene cluster from genetically intractable bacteria results in scleric acid discovery', *Chemical Science*, 10(2), pp. 453–463.

Aldred, KJ, Kerns, RJ and Osheroff, N. (2014) 'Mechanism of quinolone action and resistance.', *Biochemistry*, 53(10), pp. 1565–1574.

Altschup, SF, Gish, W., Miller, W., Myers, EW and Lipman, J. (1990) 'Basic Local Alignment Search Tool', *J. Mol. Biol*, 215, pp. 403–410.

Ames, B. D., Korman, T. P., Zhang, W., Smith, P., Vu, T., Tang, Y. and Tsai, S. C. (2008) 'Crystal structure and functional analysis of tetracenomycin ARO/CYC: Implications for cyclization specificity of aromatic polyketides', *Proceedings of the National Academy of Sciences*, 105(14), pp. 5349–5354.

Anantharaman, V. and Aravind, L. (2003) 'Evolutionary history, structural features and biochemical diversity of the NlpC/P60 superfamily of enzymes.', *Genome biology*, 4(2), pp. 1–11.

Andorfer, MC and Lewis, JC (2018) 'Understanding and Improving the Activity of Flavin-Dependent Halogenases via Random and Targeted Mutagenesis.', *Annual review of biochemistry*, 87, pp. 159–185.

Bailey, T. L., Johnson, J., Grant, C. E. and Noble, W. S. (2015) 'The MEME Suite.', *Nucleic acids research*, 43(1), pp. 39–49.

Baltz, RH (2015) 'Genetic manipulation of secondary metabolite biosynthesis for improved production in Streptomyces and other actinomycetes', *Journal of Industrial Microbiology and Biotechnology*, 43(2), pp. 343–370.

Baltz, RH (2018) 'Natural product drug discovery in the genomic era: realities, conjectures,

misconceptions, and opportunities', *Journal of Industrial Microbiology & Biotechnology*, pp. 1–19.

Barke, J., Seipke, RF, Grüschow, S., Heavens, D., Drou, N., Bibb, M., Goss, RJM, Yu, DW and Hutchings, M. I. (2010) 'A mixed community of actinomycetes produce multiple antibiotics for the fungus farming ant Acromyrmex octospinosus', *BMC Biology*, 8, pp. 1–10.

Beam, M. P., Bosserman, M. A., Noinaj, N., Wehenkel, M. and Rohr, J. (2009) 'Crystal structure of Baeyer-Villiger monooxygenase MtmOIV, the key enzyme of the mithramycin biosynthetic pathway .', *Biochemistry*, 48(21), pp. 4476–4487.

Bentley, S. D., Chater, K. F., Cerdeño-Tárraga, A. M., Challis, G. L., Thomson, N. R., James, K. D., Harris, D. E., Quail, M. a, Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C. W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C. H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O'Neil, S., Rabbinowitsch, E., Rajandream, M-a, Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B. G., Parkhill, J. and Hopwood, D. A. (2002) 'Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2).', *Nature*, 417(6885), pp. 141–147.

Bérdy, J. (2005) 'Bioactive microbial metabolites.', *The Journal of antibiotics*, 58(1), pp. 1–26.

Bibb, MJ (2005) 'Regulation of secondary metabolism in Streptomyces', *Current Opinion in Microbiology*, 8, pp. 208–215.

Bradford, M. M. (1976) 'A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding', *Analytical Biochemistry*, 72, pp. 248–254.

Buedenbender, S., Rachid, S., Müller, R. and Schulz, GE. (2009) 'Structure and Action of the Myxobacterial Chondrochloren Halogenase CndH: A New Variant of FAD-dependent Halogenases', *Journal of Molecular Biology*, 385(2), pp. 520–530.

Busarakam, K., Bull, AT, Girard, G., Labeda, DP, van Wezel, GP and Goodfellow, M. (2014) 'Streptomyces leeuwenhoekii sp. nov., the producer of chaxalactins and chaxamycins, forms a distinct branch in Streptomyces gene trees', *Antonie van Leeuwenhoek*, 105(5), pp. 849–861.

Busby, S. J. W. (2019) 'Transcription activation in bacteria: ancient and modern', *Microbiology*, 165, pp. 386–395.

Bush, MJ., Chandra, G., Bibb, M. J., Findlay, K. C. and Buttner, M. J. (2016) 'Genome-Wide Chromatin Immunoprecipitation Sequencing Analysis Shows that WhiB Is a Transcription Factor That Cocontrols Its Regulon with WhiA To Initiate Developmental Cell Division in Streptomyces', *mBio*, 7(2), pp. 516–523.

Bush, MJ, Tschowri, N., Schlimpert, S., Flärdh, K. and Buttner, MJ (2015) 'c-di-GMP signalling and the regulation of developmental transitions in streptomycetes', *Nature Reviews Microbiology*, 13(12), pp. 749–760.

Butler, MS. and Buss, AD. (2006) 'Natural products — The future scaffolds for novel antibiotics?', *Biochemical Pharmacology*, 71(7), pp. 919–929.

Butler, MS, Blaskovich, MA and Cooper, MA (2013) 'Antibiotics in the clinical pipeline in

2013', *The Journal of Antibiotics*, 66, pp. 571–591.

Cantillo, D. and Kappe, C. .. (2017) 'Halogenation of organic compounds using continuous flow and microreactor technology', *Reaction Chemistry & Engineering*, 2(1), pp. 7–19.

Challis, GL. (2014) 'Exploitation of the Streptomyces coelicolor A3(2) genome sequence for discovery of new natural products and biosynthetic pathways', *Journal of Industrial Microbiology & Biotechnology*, 41, pp. 219–232.

Challis, GL and Hopwood, DA (2003) 'Synergy and contingency as driving forces for the evolution of multiple secondary metabolite production by Streptomyces species.', *Proceedings of the National Academy of Sciences of the United States of America*, 100, pp. 14555–14561.

Chan, Yolande A., Podevels, Angela M., Kevany, Brian M. and Thomas, Michael G. (2009) 'Biosynthesis of polyketide synthase extender units.', *Natural product reports*. NIH Public Access, 26(1), pp. 90–114.

Chater, KF (2016) 'Recent advances in understanding Streptomyces.', *F1000Research*, 5, pp. 2795–2811.

Chater, KF and Chandra, G. (2006) 'The evolution of development in Streptomyces analysed by genome comparisons', *FEMS Microbiology Reviews*, 30, pp. 651–672.

Chen, A., Re, RN and Burkart, MiD (2018) 'Type II fatty acid and polyketide synthases: deciphering protein-protein and protein-substrate interactions Type II fatty acid and polyketide synthases: deciphering protein-protein and protein-substrate interactions', *Natural product reports*, 35, pp. 1029–1045.

Chen, S., Zheng, G., Zhu, H., He, H., Chen, L., Zhang, W., Jiang, W. and Lu, Y. (2016) 'Roles of two-component system AfsQ1/Q2 in regulating biosynthesis of the yellow-pigmented coelimycin P2 in Streptomyces coelicolor', *FEMS Microbiology Letters*. Edited by David Clarke, 363(15), pp. 160–168.

Chevrette, M. G., Carlson, C. M., Ortega, H. E., Thomas, C., Ananiev, G. E., Barns, K. J., Book, A. J., Cagnazzo, J., Carlos, C., Flanigan, W., Grubbs, K. J., Horn, H. A., Hoffmann, F. M., Klassen, J. L., Knack, J. J., Lewin, G. R., McDonald, B. R., Muller, L., Melo, W. G. P., Pinto-Tomás, A. A., Schmitz, A., Wendt-Pienkowski, E., Wildman, S., Zhao, M., Zhang, F., Bugni, T. S., Andes, D. R., Pupo, M. T. and Currie, C. R. (2019) 'The antimicrobial potential of Streptomyces from insect microbiomes', *Nature Communications*, 10(1), pp. 1–11.

Chopra, I. and Roberts, M. (2001) 'Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance.', *Microbiology and molecular biology reviews : MMBR*, 65(2), pp. 232–260.

Cobb, RE, Wang, Y. and Zhao, H. (2015) 'High-E ffi ciency Multiplex Genome Editing of Streptomyces Species Using an Engineered CRISPR/Cas System', *ACS Synthetic Biology*, 4, pp. 723–728.

Collin, F., Karkare, S. and Maxwell, A. (2011) 'Exploiting bacterial DNA gyrase as a drug target: current state and perspectives', *Appl Microbiol Biotechnol (2011)*, 92, pp. 479–497.

Corre, C. and Challis, GL (2007) 'Heavy Tools for Genome Mining', *Chemistry & Biology*, 14, pp. 7–9.

Corre, C. and Challis, GL (2009) ‘New natural product biosynthetic chemistry discovered by genome mining’, *Natural product reports*, 26, pp. 977–986.

Crone, WJK, Leeper, FJ and Truman, AW (2012) ‘Identification and characterisation of the gene cluster for the anti-MRSA antibiotic bottromycin: expanding the biosynthetic diversity of ribosomal peptides’, *Chemical Science*, 3(12), pp. 3516–3521.

Currie, C. R., Bot, A. N. M. and Boomsma, J. J. (2003) ‘Experimental evidence of a tripartite mutualism: bacteria protect ant fungus gardens from specialized parasites’, *Oikos*, 101(1), pp. 91–102.

Currie, CR., Scott, JA., Summerbell, RC. and Malloch, D. (1999) ‘Fungus-growing ants use antibiotic-producing bacteria to control garden parasites’, *Nature*, 398(6729), pp. 701–704.

Currie, CR, Poulsen, M., Mendenhall, J., Boomsma, JJ and Billen, J. (2006) ‘Coevolved crypts and exocrine glands support mutualistic bacteria in fungus-growing ants.’, *Science (New York, N.Y.)*, 311(5757), pp. 81–83.

Cuthbertson, L. and Nodwell, J. R. (2013) ‘The TetR family of regulators.’, *Microbiology and molecular biology reviews : MMBR*, 77(3), pp. 440–475.

Dangel, V., Harle, J., Goerke, C., Wolz, C., Gust, B., Pernodet, J. L. and Heide, L. (2009) ‘Transcriptional regulation of the novobiocin biosynthetic gene cluster’, *Microbiology*, 155(12), pp. 4025–4035.

Datsenko, K. A. and Wanner, B. L. (2000) ‘One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products.’, *Proceedings of the National Academy of Sciences of the United States of America*, 97(12), pp. 6640–6645.

Davies, J. and Davies, D. (2010) ‘Origins and evolution of antibiotic resistance.’, *Microbiology and molecular biology reviews : MMBR*, 74(3), pp. 417–433.

Demain, AL. (2009) ‘Antibiotics: Natural Products Essential to Human Health’, *Medical Research Reviews*, 29(6), pp. 821–842.

Devine, R., Hutchings, MI and Holmes, NA (2017) ‘Future directions for the discovery of antibiotics from actinomycete bacteria’, *Emerging Topics in Life Scienes*, 1, pp. 1–12.

Dodge, G. J., Maloney, F. P. and Smith, J. L. (2018) ‘Protein–protein interactions in “ cis -AT” polyketide synthases’, *Natural Product Reports*, 35(10), pp. 1082–1096.

Dong, C., Flecks, S., Unversucht, S., Haupt, C., Van Pée, KH and Naismith, JH (2005) ‘Structural biology: Tryptophan 7-halogenase (PrnA) structure suggests a mechanism for regioselective chlorination’, *Science*, 309(5744), pp. 2216–2219.

Dorrestein, PC, Yeh, E., Garneau-Tsodikova, S., Kelleher, NL and Walsh, CT (2005) ‘Dichlorination of a pyrrolyl-S-carrier protein by FADH2-dependent halogenase PltA during pyoluteorin biosynthesis.’, *Proceedings of the National Academy of Sciences of the United States of America*, 102(39), pp. 13843–8.

Du, D., van Veen, H. W., Murakami, S., Pos, K. M. and Luisi, B. F. (2015) ‘Structure, mechanism and cooperation of bacterial multidrug transporters’, *Current Opinion in Structural Biology*, 33, pp. 76–91.

Dutta, S., Whicher, JR, Hansen, DA, Hale, WA, Chemler, JA, Congdon, GR, Narayan, ARH,

Håkansson, K., Sherman, DH, Smith, JL and Skiniotis, G. (2014) 'Structure of a modular polyketide synthase.', *Nature*, 510(7506), pp. 512–517.

Ettwiller, L., Buswell, J., Yigit, E. and Schildkraut, I. (2016) 'A novel enrichment strategy reveals unprecedented number of novel transcription start sites at single base resolution in a model prokaryote and the gut microbiome', *BMC Genomics*, 17(1), pp. 199–213.

Felnagle, E. A., Jackson, E. E., Chan, Y. A., Podevels, A. M., Berti, A. D., McMahon, M. D. and Thomas, M. G. (2008) 'Nonribosomal peptide synthetases involved in the production of medically relevant natural products.', *Molecular pharmaceutics*, 5(2), pp. 191–211.

Feng, Z., Chakraborty, D., Dewell, S., Reddy, BVB and Brady, SF (2012) 'Environmental DNA-encoded antibiotics fasamycins A and B inhibit FabF in type II fatty acid biosynthesis.', *Journal of the American Chemical Society*, 134(6), pp. 2981–2987.

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

Fischbach, MA and Walsh, CT (2006) 'Assembly-Line Enzymology for Polyketide and Nonribosomal Peptide Antibiotics: Logic, Machinery, and Mechanisms', *Chemical Reviews*, 106, pp. 3468–3496.

Fleming, A. (1929) 'On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae', *British Jounral of Experimental Pathology*, 10(3), pp. 226–236.

Fukutake, M., Takahashi, M., Ishida, K., Kawamura, H., Sugimura, T. and Wakabayashi, K. (1996) 'Quantification of genistein and genistin in soybeans and soybean products.', *Food and chemical toxicology : an international journal published for the British Industrial Biological Research Association*, 34(5), pp. 457–461.

Gao, R., Mack, T. R. and Stock, A. M. (2007) 'Bacterial response regulators: versatile regulatory strategies from common domains.', *Trends in biochemical sciences*, 32(5), pp. 225–234.

Gomez-Escribano, JP and Bibb, MJ (2011) 'Engineering Streptomyces coelicolor for heterologous expression of secondary metabolite gene clusters', *Microbial Biotechnology*, 4(2), pp. 207–215.

Gomez-Escribano, JP and Bibb, MJ (2014) 'Heterologous expression of natural product biosynthetic gene clusters in Streptomyces coelicolor: from genome mining to manipulation of biosynthetic pathways', *Journal of Industrial Microbiology & Biotechnology*, 41(2), pp. 425–431.

Gomez-Escribano, JP, Song, L., Bibb, MJ and Challis, GL (2012) 'Posttranslational β-methylation and macrolactamidination in the biosynthesis of the bottromycin complex of ribosomal peptide antibiotics', *Chemical Science*, 3(12), pp. 3522–3525.

Goodfellow, M., Busarakam, K., Idris, H., Labeda, DP, Nouioui, I., Brown, R., Kim, BY, Del Carmen Montero-Calasanz, M., Andrews, BA and Bull, AT (2017) 'Streptomyces asenjonii sp. nov., isolated from hyper-arid Atacama Desert soils and emended description of Streptomyces viridosporus Pridham et al. 1958.', *Antonie van Leeuwenhoek*, 110(9), pp. 1133–1148.

Grove, A. (2017) 'Regulation of Metabolic Pathways by MarR Family Transcription Factors.', *Computational and structural biotechnology journal*, 15, pp. 366–371.

Guo, C., Mandalapu, D., Ji, X., Gao, Ji and Zhang, Q. (2018) 'Chemistry and Biology of Teixobactin', *Chemistry - A European Journal*, 24(21), pp. 5406–5422.

Guo, J., Zhang, X., Lu, X., Liu, W., Chen, Z., Li, J., Deng, L. and Wen, Y. (2018) 'SAV4189, a MarR-Family Regulator in Streptomyces avermitilis, Activates Avermectin Biosynthesis.', *Frontiers in microbiology*, 9, pp. 1–15.

Gust, B., Chandra, G., Jakimowicz, D., Yuqing, T., Bruton, CJ. and Chater, K. F. (2004) 'λ Red-Mediated Genetic Manipulation of Antibiotic-Producing Streptomyces', *Advances in Applied Microbiology*, 54, pp. 107–128.

Heine, D., Holmes, NA., Worsley, S. F., Santos, ACA., Innocent, TM, Scherlach, K., Patrick, ElH, Yu, DW, Murrell, JC, Vieria, PC, Boomsma, JJ, Hertweck, C., Hutchings, MI and Wilkinson, B. (2018) 'Chemical warfare between leafcutter ant symbionts and a co-evolved pathogen', *Nature Communications*, 9(1), pp. 2208–2219.

Hengst, CD, Tran, NT, Bibb, MJ, Chandra, G., Leskiw, BK and Buttner, MJ (2010) 'Genes essential for morphological development and antibiotic production in Streptomyces coelicolor are targets of BldD during vegetative growth', *Molecular Microbiology*, 78(2), pp. 361–379.

Hertweck, C., Luzhetskyy, A., Rebets, Yu and Bechthold, A. (2007) 'Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork', *Nat. Prod. Rep*, 24, pp. 162–190.

van der Heul, HU, Bilyk, BL, McDowall, KJ, Seipke, RF and van Wezel, GP (2018) 'Regulation of antibiotic production in Actinobacteria: new perspectives from the post-genomic era', *Natural Product Reports*, 35(6), pp. 575–604.

Hofeditz, T., Unsin, C. E. M., Wiese, J., Imhoff, J. F., Wohlleben, W., Grond, S. and Weber, T. (2018) 'Lysoquinone-TH1, a New Polyphenolic Tridecaketide Produced by Expressing the Lysolipin Minimal PKS II in Streptomyces albus.', *Antibiotics*, 7(53), pp. 1–11.

Holmes, NA., Devine, R., Qin, Zh, Seipke, RF., Wilkinson, B. and Hutchings, MI (2018) 'Complete genome sequence of Streptomyces formicae KY5, the formicamycin producer', *Journal of Biotechnology*, 265, pp. 116–118.

Hopwood, D. A. (2007) 'How do antibiotic-producing bacteria ensure their self-resistance before antibiotic biosynthesis incapacitates them?', *Molecular Microbiology*, 63(4), pp. 937–940.

Hopwood, DA (2006) 'Soil To Genomics: The Streptomyces Chromosome', *Annual Review of Genetics*, 40(1), pp. 1–23.

Hopwood, DA (2007) *Streptomyces in Nature and Medicine: The Antibiotic Makers*. Oxford University Press.

Hosaka, T., Ohnishi-Kameyama, M., Muramatsu, H., Murakami, K., Tsurumi, Y., Kodani, S., Yoshida, M., Fujie, A. and Ochi, K. (2009) 'Antibacterial discovery in actinomycetes strains with mutations in RNA polymerase or ribosomal protein S12', *Nature Biotechnology*. Nature Publishing Group, 27(5), pp. 462–464.

Hover, BM, Kim, SH, Katz, M., Charlop-Powers, Z., Owen, JG, Ternei, MA, Maniko, J., Estrela, AB, Molina, H., Park, S., Perlin, DS and Brady, SF (2018) 'Culture-independent discovery of the malacidins as calcium-dependent antibiotics with activity against multidrug-resistant Gram-positive pathogens', *Nature Microbiology*, 3(4), pp. 415–422.

Hudson, GA and Mitchell, DA (2018) 'RiPP antibiotics: biosynthesis and engineering potential', *Current Opinion in Microbiology*, 45, pp. 61–69.

Hur, GH, Vickery, CR and Burkart, MD (2012) 'Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology.', *Natural product reports*, 29(10), pp. 1074–98.

Hutchings, M. I., Hoskisson, Paul A., Chandra, Govind and Buttner, Mark J. (2004) 'Sensing and responding to diverse extracellular signals? Analysis of the sensor kinases and response regulators of Streptomyces coelicolor A3(2)', *Microbiology*, 150(9), pp. 2795–2806.

Jacob-Dubuisson, F., Mechaly, A., Betton, JM and Antoine, R. (2018) 'Structural insights into the signalling mechanisms of two-component systems', *Nature Reviews Microbiology*, 16, pp. 585–593.

Jansen, R., Kunze, B., Reichenbach, H. and Höfle, G. (2003) 'Chondrochloren A and B, New β-Amino Styrenes from Chondromyces crocatus (Myxobacteria)', *European Journal of Organic Chemistry*, 2003(14), pp. 2684–2689.

Jiang, J., Guiza Beltran, D., Schacht, A., Wright, S., Zhang, L. and Du, L. (2018) 'Functional and Structural Analysis of Phenazine O -Methyltransferase LaPhzM from Lysobacter antibioticus OH13 and One-Pot Enzymatic Synthesis of the Antibiotic Myxin', *ACS Chemical Biology*, 13(4), pp. 1003–1012.

Jiang, W., Bikard, D., Cox, D., Zhang, F. and Marraffini, LA (2013) 'RNA-guided editing of bacterial genomes using CRISPR-Cas systems', *Nature Biotechnology*. Nature Publishing Group, 31(3), pp. 233–239.

Kapoor, G., Saigal, S. and Elongavan, A. (2017) 'Action and resistance mechanisms of antibiotics: A guide for clinicians.', *Journal of anaesthesiology, clinical pharmacology*, 33(3), pp. 300–305.

Kemung, HM, Tan, LTH, Khan, TM, Chan, KG, Pusparajah, P., Goh, BH and Lee, LH (2018) 'Streptomyces as a Prominent Resource of Future Anti-MRSA Drugs.', *Frontiers in microbiology*, 9, pp. 1–26.

Kepplinger, B., Morton-Laing, S., Seistrup, KH, Marrs, ECL, Hopkins, AP, Perry, JD, Strahl, H., Hall, MJ, Errington, J. and Ellis Allenby, NE (2018) 'Mode of Action and Heterologous Expression of the Natural Product Antibiotic Vancoresmycin', *ACS Chemical Biology*, 13(1), pp. 207–214.

van Keulen, G. and Dyson, PJ (2014) 'Production of Specialized Metabolites by Streptomyces coelicolor A3(2)', *Advances in Applied Microbiology*, 89, pp. 217–266.

Khan, S., Komaki, H., Motohashi, K., Kozone, I., Mukai, A., Takagi, M. and Shin-Ya, K. (2011) 'Streptomyces associated with a marine sponge Haliclona sp.; biosynthetic genes for secondary metabolites and products', *Environmental Microbiology*, 13(2), pp. 391–403.

Kim, JY, Inaoka, T., Hirooka, K., Matsuoka, H., Murata, M., Ohki, R., Adachi, Y., Fujita, Y. and Ochi, K. (2009) 'Identification and characterization of a novel multidrug resistance operon, mdtRP (yusOP), of Bacillus subtilis.', *Journal of bacteriology*, 191(10), pp. 3273–3281.

Komatsu, M., Tsuda, M., Omura, S., Oikawa, H. and Ikeda, H. (2008) 'Identification and functional analysis of genes controlling biosynthesis of 2-methylisoborneol.', *Proceedings of the National Academy of Sciences of the United States of America*, 105(21), pp. 7422–7427.

Kosol, S., Jenner, M., Lewandowski, Jf R. and Challis, G. L. (2018) 'Protein–protein interactions in trans -AT polyketide synthases', *Natural Product Reports*, 35(10), pp. 1097–1109.

Labeda, D. P., Dunlap, C. A., Rong, X., Huang, Y., Doroghazi, J. R., Ju, K. and Metcalf, W. W. (2017) 'Phylogenetic relationships in the family Streptomycetaceae using multi-locus sequence analysis', *Antonie van Leeuwenhoek*, 110(4), pp. 563–583.

Latham, J., Brandenburger, E., Shepherd, SA, Melon, BRK and Micklefield, J. (2017) 'Development of Halogenase Enzymes for Use in Synthesis', *Chemical Reviews*, 118, pp. 232–269.

Laureti, L., Song, L., Huang, S., Corre, C., Leblond, P., Challis, G. and Aigle, B. (2011) 'Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in Streptomyces ambofaciens.', *Proceedings of the National Academy of Sciences of the United States of America*, 108(15), pp. 6258–6263.

Leisch, H., Morley, K. and Lau, P. C. K. (2011) 'Baeyer−Villiger Monooxygenases: More Than Just Green Chemistry', *Chemical Reviews*, 111(7), pp. 4165–4222.

Levine, C., Hiasa, H. and Marians, KJ (1998) 'DNA gyrase and topoisomerase IV: biochemical activities, physiological roles during chromosome replication, and drug sensitivities', *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, 1400, pp. 29–43.

Li, K., Li, QL, Ji, NY, Liu, B., Zhang, W. and Cao, XP (2011) 'Deoxyuridines from the marine sponge associated actinomycete Streptomyces microflavus.', *Marine drugs*, 9(5), pp. 690–695.

Ling, LL, Schneider, T., Peoples, AJ, Spoering, AL, Engels, I., Conlon, BP, Mueller, A., Schäberle, TF, Hughes, DE, Epstein, S., Jones, M., Lazarides, L., Steadman, VA, Cohen, DR, Felix, CR, Fetterman, KA, Millett, WP, Nitti, AG, Zullo, AM, Chen, C. and Lewis, K. (2015) 'A new antibiotic kills pathogens without detectable resistance', *Nature*, 517(7535), pp. 455–459.

Liu, YH, Belcheva, A., Konermann, L. and Golemi-Kotra, D. (2009) 'Phosphorylation-Induced Activation of the Response Regulator VraR from Staphylococcus aureus: Insights from Hydrogen Exchange Mass Spectrometry', *Journal of Molecular Biology*, 391(1), pp. 149–163.

Liu, YY, Wang, Y., Walsh, TR, Yi, LX, Zhang, R., Spencer, J., Doi, Y., Tian, G., Dong, B., Huang, X., Yu, LF, Gu, Da, Ren, H., Chen, X., Lv, L., He, D., Zhou, H., Liang, Z., Liu, JH and Shen, J. (2016) 'Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study', *The Lancet Infectious Diseases*, 16(2), pp. 161–168.

Livermore, D. M. (2004) 'The need for new antibiotics', *Clinical Microbiology and Infection*, 10, pp. 1–9.

Low, Z. J., Pang, L. M., Ding, Y., Cheang, Q. W., Le Mai Hoang, K., Thi Tran, H., Li, J., Liu, X. W., Kanagasundaram, Y., Yang, L. and Liang, Z. X. (2018) 'Identification of a biosynthetic gene cluster for the polyene macrolactam sceliphrolactam in a Streptomyces strain isolated

from mangrove sediment', *Scientific Reports*, 8(1), pp. 1–13.

Lu, F., Hou, Y., Zhang, H., Chu, Y., Xia, H. and Tian, Y. (2017) 'Regulatory genes and their roles for improvement of antibiotic biosynthesis in Streptomyces.', *3 Biotech*, 7(4), pp. 250–265.

MacNeil, D. J., Gewain, K. M., Ruby, C. L., Dezeny, G., Gibbons, P. H. and Maeneil, T. (1992) 'Analysis of Streptomyces avermitilis genes required for avermectin biosynthesis utilizing a novel integration vector', *Gene*, 111, pp. 61–68.

Mak, S., Xu, Y. and Nodwell, J. R. (2014) 'The expression of antibiotic resistance genes in antibiotic-producing bacteria', *Molecular Microbiology*, 93(3), pp. 391–402.

Malpartida, F. and Hopwood, D. A. (1984) 'Molecular cloning of the whole biosynthetic pathway of a Streptomyces antibiotic and its expression in a heterologous host.', *Nature*, 309(5967), pp. 462–4.

Manteca, Á. and Yagüe, P. (2018) 'Streptomyces Differentiation in Liquid Cultures as a Trigger of Secondary Metabolism', *Antibiotics*, 7(2), pp. 41–54.

Marinello, J., Delcuratolo, M., Capranico, G., Marinello, J., Delcuratolo, M. and Capranico, G. (2018) 'Anthracyclines as Topoisomerase II Poisons: From Early Studies to New Perspectives', *International Journal of Molecular Sciences*, 19(11), pp. 3480–3497.

Martín, JF (2004) 'Phosphate control of the biosynthesis of antibiotics and other secondary metabolites is mediated by the PhoR-PhoP system: an unfinished story.', *Journal of bacteriology*, 186(16), pp. 5197–5201.

Martín, JF, Sola-Landa, A., Santos-Beneit, F., Fernández-Martínez, LT, Prieto, C. and Rodríguez-García, A. (2010) 'Minireview Cross-talk of global nutritional regulators in the control of primary and secondary metabolism in Streptomyces', *Microbial Biotechnology*, 4(2), pp. 165–174.

Martínez-Núñez, Mario Alberto and López, Víctor Eric López y (2016) 'Nonribosomal peptides synthetases and their applications in industry', *Sustainable Chemical Processes*. SpringerOpen, 4(1), p. 13.

Mclean, T. C., Wilkinson, B., Hutchings, M. I. and Devine, R. (2019) 'Dissolution of the Disparate: Co-ordinate Regulation in Antibiotic Biosynthesis', *Antibiotics*, 8, pp. 21–32.

van der Meij, A., Worsley, SF, Hutchings, MI and van Wezel, GP (2017) 'Chemical ecology of antibiotic production by actinomycetes', *FEMS Microbiology Reviews*, 41(3), pp. 392–416.

Menon, BRK, Brandenburger, E., Sharif, HH, Klemstein, U., Shepherd, SA, Greaney, MF and Micklefield, J. (2017) 'RadH: A Versatile Halogenase for Integration into Synthetic Pathways.', *Angewandte Chemie (International ed. in English)*, 56(39), pp. 11841–11845.

Michael, CA, Dominey-Howes, D. and Labbate, M. (2014) 'The antimicrobial resistance crisis: causes, consequences, and management.', *Frontiers in public health*, 2, pp. 1–8.

Miles, A. A., Misra, S. S. and Irwin, J. O. (1938) 'The estimation of the bactericidal power of the blood.', *The Journal of hygiene*, 38(6), pp. 732–749.

Milshteyn, A., Schneider, JS and Brady, SF (2014) 'Mining the Metabiome: Identifying Novel Natural Products from Microbial Communities', *Chemistry and Biology*, 18, pp. 1211–1223.

Misumi, M. and Tanaka, N. (1980) 'Mechanism of inhibition of translocation by kanamycin and viomycin: a comparative study with fusidic acid.', *Biochemical and biophysical research communications*, 92(2), pp. 647–654.

Mitscher, LA (2005) 'Bacterial Topoisomerase Inhibitors: Quinolone and Pyridone Antibacterial Agents', *Chemical Reviews*, 105, pp. 559–592.

Mo, J., Wang, S., Zhang, W., Li, C., Deng, Z., Zhang, L. and Qu, X. (2019) 'Efficient editing DNA regions with high sequence identity in actinomycetal genomes by a CRISPR-Cas9 system', *Synthetic and Systems Biotechnology*, 4(2), pp. 86–91.

Moore, JM, Bradshaw, E., Seipke, R., Hutchings, MI and McArthur, M. (2012) 'Use and Discovery of Chemical Elicitors That Stimulate Biosynthetic Gene Clusters in Streptomyces Bacteria', *Methods in Enzymology*, 517, pp. 367–385.

Muheim, C., Götzke, H., Eriksson, A. U., Lindberg, S., Lauritsen, I., Nørholm, M. H. H. and Daley, D. O. (2017) 'Increasing the permeability of Escherichia coli using MAC13243.', *Scientific Reports*, 7(1), pp. 1–11.

Myronovskyi, M., Welle, E., Fedorenko, V. and Luzhetskyy, A. (2011) 'B-Glucuronidase as a Sensitive and Versatile Reporter in Actinomycetes', *Applied and Environmental Microbiology*, 77(15), pp. 5370–5383.

Nah, HJ, Pyeon, HR, Kang, SH, Choi, SS and Kim, ES (2017) 'Cloning and Heterologous Expression of a Large-sized Natural Product Biosynthetic Gene Cluster in Streptomyces Species.', *Frontiers in microbiology*, 8, pp. 394–403.

Nature Biotechnology (2018) 'Wanted: a reward for antibiotic development', *Nature Biotechnology*, 36(7), pp. 555–555.

Nechitaylo, TY, Westermann, M. and Kaltenpoth, M. (2014) 'Cultivation reveals physiological diversity among defensive "Streptomyces philanthi" symbionts of beewolf digger wasps (Hymenoptera, Crabronidae)', *BMC Microbiology*, 14, pp. 202–218.

Nesme, J., Cécillon, S., Delmont, T. .., Monier, J. .., Vogel, T. .. and Simonet, P. (2014) 'Large-Scale Metagenomic-Based Study of Antibiotic Resistance in the Environment', *Current Biology*, 24(10), pp. 1096–1100.

Newmister, S. A., Romminger, S., Schmidt, J. J., Williams, R. M., Smith, J. L., Berlinck, R. G. S. and Sherman, D. H. (2018) 'Unveiling sequential late-stage methyltransferase reactions in the meleagrin/oxaline biosynthetic pathway', *Organic & Biomolecular Chemistry*, 16(35), pp. 6450–6459.

Nichols, D., Cahoon, N., Trakhtenberg, E. M., Pham, L., Mehta, A., Belanger, A., Kanigan, T., Lewis, K. and Epstein, S. S. (2010) 'Use of ichip for high-throughput in situ cultivation of "uncultivable microbial species???', *Applied and Environmental Microbiology*, 76(8), pp. 2445–2450.

Notredame, C., Higgins, D. G. and Heringa, J. (2000) 'T-Coffee: A novel method for fast and accurate multiple sequence alignment.', *Journal of molecular biology*, 302(1), pp. 205–17.

O'Neil, J. (2016) *Tackling drug-resistant infections globally: Final report and recommendations. The Review on antimicrobial resistance chaired by Jim O'Neil*.

O'Rourke, Sean, Widdick, David and Bibb, Mervyn (2017) 'A novel mechanism of immunity

controls the onset of cinnamycin biosynthesis in Streptomyces cinnamoneus DSM 40646', *Journal of Industrial Microbiology & Biotechnology*, 44(4), pp. 563–572.

O Bachmann, B., Van Lanen, SG and Baltz, RH (2014) 'Microbial gemone mining for accelerated natural products discovery: is a renaissance in the making?', *Journal of Industrial Microbiology and Biotechnology*, 41(2), pp. 1384–1399.

Ochi, K. and Hosaka, T. (2013) 'New strategies for drug discovery: activation of silent or weakly expressed microbial gene clusters.', *Applied microbiology and biotechnology*, 97(1), pp. 87–98.

Ohnishi, Y., Ishikawa, J., Hara, H., Suzuki, H., Ikenoya, M., Ikeda, H., Yamashita, A., Hattori, M. and Horinouchi, S. (2008) 'Genome sequence of the streptomycin-producing microorganism Streptomyces griseus IFO 13350.', *Journal of bacteriology*, 190(11), pp. 4050–4060.

Okamoto, S., Taguchi, T., Ochi, K. and Ichinose, K. (2009) 'Biosynthesis of Actinorhodin and Related Antibiotics: Discovery of Alternative Routes for Quinone Formation Encoded in the act Gene Cluster', *Chemistry & Biology*, 16, pp. 226–236.

Onaka, H., Mori, Y., Igarashi, Y. and Furumai, T. (2011) 'Mycolic acid-containing bacteria induce natural-product biosynthesis in Streptomyces species', *Applied and Environmental Microbiology*, 77(2), pp. 400–406.

Ortega, MA and Van Der Donk, WA (2016) 'New Insights into the Biosynthetic Logic of Ribosomally Synthesized and Post-translationally Modified Peptide Natural Products', *Cell Chemical Biology*, 23, pp. 31–44.

Pang, AH, Garneau-Tsodikova, S. and Tsodikov, OV (2015) 'Crystal structure of halogenase PltA from the pyoluteorin biosynthetic pathway', *Journal of Structural Biology*, 192(3), pp. 349–357.

Panter, F., Krug, D., Baumann, S. and Müller, R. (2018) 'Self-resistance guided genome mining uncovers new topoisomerase inhibitors from myxobacteria', *Chemical Science*, 9(21), pp. 4898–4908.

Paulsen, I. T., Brown, M. H. and Skurray, R. A. (1996) 'Proton-Dependent Multidrug Efflux Systems', *Microbiological Reviews*, 60(4), pp. 575–608.

van Pée, K. H., Milbredt, D., Patallo, E. P., Weichold, V. and Gajewi, M. (2016) 'Application and Modification of Flavin-Dependent Halogenases', *Methods in Enzymology*, 575, pp. 65–92.

Peláez, F. (2006) 'The historical delivery of antibiotics from microbial natural products - Can history repeat?', *Biochemical Pharmacology*, 71(7), pp. 981–990.

Perera, I. C. and Grove, A. (2010) 'Molecular Mechanisms of Ligand-Mediated Attenuation of DNA Binding by MarR Family Transcriptional Regulators', *Journal of Molecular Cell Biology*, 2(5), pp. 243–254.

Pérez, J., Muñoz-Dorado, J., Braña, AF, Shimkets, LJ, Sevillano, L. and Santamaría, RI (2011) 'Myxococcus xanthus induces actinorhodin overproduction and aerial mycelium formation by Streptomyces coelicolor.', *Microbial biotechnology*, 4(2), pp. 175–183.

Pickens, L. B. and Tang, Y. (2010) 'Oxytetracycline biosynthesis.', *The Journal of biological*

chemistry, 285(36), pp. 27509–27515.

Piddock, LJV (2015) 'Teixobactin, the first of a new class of antibiotics discovered by ichip technology?', *Journal of Antimicrobial Chemotherapy*, 70(10), pp. 2679–2680.

Piffaretti, JC, Arini, A. and Frey, J. (1988) 'pUB307 mobilizes resistance plasmids from Escherichia coli into Neisseria gonorrhoeae', *MGG Molecular & General Genetics*, 212(2), pp. 215–218.

Pongkittiphan, V., Theodorakis, EA and Chavasiri, W. (2009) 'Hexachloroethane: a highly efficient reagent for the synthesis of chlorosilanes from hydrosilanes.', *Tetrahedron letters*, 50(36), pp. 5080–5082.

Porter, C. M. and Miller, B. G. (2012) 'Cooperativity in Monomeric Enzymes with Single Ligand-Binding Sites', *Bioorganic chemistry*, 43, pp. 44–63.

Poulsen, M., Oh, DC, Clardy, J. and Currie, CR (2011) 'Chemical Analyses of Wasp-Associated Streptomyces Bacteria Reveal a Prolific Potential for Natural Products Discovery', *PLoS ONE*, 6(2), pp. 1–8.

Qin, Z., Devine, R., Hutchings, M. I. and Wilkinson, B. (2019) 'Aromatic polyketide biosynthesis: fidelity, evolution and engineering', *bioRxiv*, online.

Qin, Z., Munnoch, JT, Devine, R., Holmes, NA, Seipke, RF, Wilkinson, KA, Wilkinson, B. and Hutchings, MI (2017) 'Formicamycins, antibacterial polyketides produced by Streptomyces formicae isolated from African Tetraponera plant-ants †', *Chemical Science*, 8, pp. 3218–3227.

Rachid, S., Scharfe, M., Blöcker, H., Weissman, KJ. and Müller, R. (2009) 'Unusual Chemistry in the Biosynthesis of the Antibiotic Chondrochlorens', *Chemistry & Biology*, 16(1), pp. 70–81.

Ray, L. and Moore, BS (2002) 'Recent advances in the biosynthesis of unusual polyketide synthase substrates', *Nat. Prod. Rep*, 19(2), pp. 150–161.

Rayner, DM and Cutts, SM (2014) 'Anthracyclines', *Side Effects of Drugs Annual*, 36, pp. 683–694.

Reddy, BVB, Kallifidas, D., Kim, JH, Charlop-Powers, Z., Feng, Z. and Brady, SF (2012) 'Natural Product Biosynthetic Gene Diversity in Geographically Distinct Soil Microbiomes', *Applied and Environmental Microbiology*, 78, pp. 3744–3752.

Reece, RJ and Maxwell, A. (1991) 'DNA Gyrase: Structure and Function', *Critical Reviews in Biochemistry and Molecular Biology*, 26(3–4), pp. 335–375.

Reynolds, K. .., Luhavaya, H., Li, J., Dahesh, S., Nizet, V., Yamanaka, K. and Moore, BS (2018) 'Isolation and structure elucidation of lipopeptide antibiotic taromycin B from the activated taromycin biosynthetic gene cluster', *The Journal of Antibiotics*, 71(2), pp. 333–338.

Robbins, T., Liu, YC, Cane, D. E. and Khosla, C. (2016) 'Structure and mechanism of assembly line polyketide synthases.', *Current opinion in structural biology*, 41, pp. 10–18.

Rodríguez, H., Rico, S., Díaz, M. and Santamaría, Ri (2013) 'Two-component systems in Streptomyces: key regulators of antibiotic complex pathways', *Microbial Cell Factories*, 12, pp. 127–137.

Romero-Rodríguez, Al, Rocha, D., Ruiz-Villafán, B., Guzmán-Trampe, S., Maldonado-Carmona, N., Vázquez-Hernández, M., Zelarayán, A., Rodríguez-Sanoja, R. and Sánchez, S. (2017) 'Carbon catabolite regulation in Streptomyces: new insights and lessons learned', *World Journal of Microbiology and Biotechnology*, 33(9), p. 162.

Romero, DA, Hasan, AH, Lin, YFi, Kime, L., Ruiz-Larrabeiti, O., Urem, M., Bucca, G., Mamanova, L., Laing, E. E., van Wezel, G. P., Smith, C. P., Kaberdin, V. R. and McDowall, K. J. (2014) 'A comparison of key aspects of gene regulation in Streptomyces coelicolor and Escherichia coli using nucleotide-resolution transcription maps produced in parallel by global and differential RNA sequencing.', *Molecular microbiology*, 94(5), pp. 963–987.

Roy, A., Kucukural, A. and Zhang, Y. (2010) 'I-TASSER: a unified platform for automated protein structure and function prediction.', *Nature protocols*, 5(4), pp. 725–38.

Roy, K., Gabunilas, J., Gillespie, A., Ngo, D. and Chanfreau, G. F. (2016) 'Common genomic elements promote transcriptional and DNA replication roadblocks.', *Genome research*, 26(10), pp. 1363–1375.

Ruiz, N., Kahne, D. and Silhavy, T. J. (2006) 'Advances in understanding bacterial outer-membrane biogenesis', *Nature Reviews Microbiology*, 4(1), pp. 57–66.

Rutledge, PJ and Challis, GL (2015) 'Discovery of microbial natural products by activation of silent biosynthetic gene clusters', *Nature Reviews Microbiology*, 13, pp. 509–523.

Sanchez, S. and Demain, AL (2008) 'Metabolic regulation and overproduction of primary metabolites.', *Microbial biotechnology*, 1(4), pp. 283–319.

Schnepel, C. and Sewald, N. (2017) 'Enzymatic Halogenation: A Timely Strategy for Regioselective C−H Activation', *Chemistry - A European Journal*, 23(50), pp. 12064–12086.

Sciara, G., Kendrew, SG., Miele, A. E., Marsh, NG., Federici, L., Malatesta, F., Schimperna, G., Savino, C. and Vallone, B. (2003) 'The structure of ActVA-Orf6, a novel type of monooxygenase involved in actinorhodin biosynthesis', *The EMBO Journal*, 22(2), pp. 205–215.

Scott, JJ, Oh, DC, Yuceer, MC, Klepzig, KD, Clardy, J. and Currie, CR (2008) 'Bacterial protection of beetle-fungus mutualism.', *Science*, 322.

Seipke, RF, Barke, J., Heavens, D., Yu, DW and Hutchings, MI (2013) 'Analysis of the bacterial communities associated with two ant-plant symbioses', *MicrobiologyOpen*, 2(2), pp. 276–283.

Seipke, RF, Grüschow, S., Goss, RJM and Hutchings, MI (2012) 'Isolating antifungals from fungus-growing ant symbionts using a genome-guided chemistry approach.', *Methods in enzymology*, 517, pp. 47–70.

Seipke, Ryan F., Crossman, Lisa, Drou, Nizar, Heavens, Darren, Bibb, Mervyn J., Caccamo, Mario and Hutchings, Matthew I. (2011) 'Draft genome sequence of Streptomyces strain S4, a symbiont of the leaf-cutting ant Acromyrmex octospinosus.', *Journal of bacteriology*, 193(16), pp. 4270–4271.

Seipke, Ryan F., Patrick, Elaine and Hutchings, Matthew I. (2014) 'Regulation of antimycin biosynthesis by the orphan ECF RNA polymerase sigma factor σ (AntA.).', *PeerJ*, 2, pp. 253–271.

Sharif, EU and O'Doherty, G. A. (2012) 'Biosynthesis and Total Synthesis Studies on The Jadomycin Family of Natural Products.', *European journal of organic chemistry*, 2012(11), pp. 1–46.

Silvennoinen, L., Sandalova, T. and Schneider, G. (2009) 'The polyketide cyclase RemF from Streptomyces resistomycificus contains an unusual octahedral zinc binding site', *FEBS Letters*, 583(17), pp. 2917–2921.

Singer, AC, Shaw, H., Rhodes, V. and Hart, A. (2016) 'Review of Antimicrobial Resistance in the Environment and Its Relevance to Environmental Regulators', *Frontiers in Microbiology*, 7, pp. 1–22.

Singh, M., Chaudhary, S. and Sareen, D. (2017) 'Non-ribosomal peptide synthetases: Identifying the cryptic gene clusters and decoding the natural product', *J. Biosci*, 42, pp. 175–187.

Som, NF, Heine, Da, Holmes, NA, Munnoch, JT, Chandra, G., Seipke, RF, Hoskisson, PA, Wilkinson, B. and Hutchings, MI (2017) 'The Conserved Actinobacterial Two-Component System MtrAB Coordinates Chloramphenicol Production with Sporulation in Streptomyces venezuelae NRRL B-65442', *Frontiers in Microbiology*, 8, pp. 1–11.

Studier, F. William and Moffattf, Barbara A. (1986) 'Use of Bacteriophage T7 RNA Polymerase to Direct Selective High-level Expression of Cloned Genes', *J. MoZ. Biol*, 189, pp. 113–130.

Sultan, I., Rahman, S., Jan, AT, Siddiqui, MT, Mondal, AH and Haq, QMR (2018) 'Antibiotics, Resistome and Resistance Mechanisms: A Bacterial Perspective.', *Frontiers in microbiology*. Frontiers Media SA, 9, pp. 1–16.

Tao, W., Yang, A., Deng, Z. and Sun, Y. (2018) 'CRISPR/Cas9-Based Editing of Streptomyces for Discovery, Characterization, and Production of Natural Products.', *Frontiers in microbiology*, 9, p. 1660.

Thompson, T. B., Katayama, K., Watanabe, K., Hutchinson, C. R. and Rayment, I. (2004) 'Structural and Functional Analysis of Tetracenomycin F2 Cyclase from Streptomyces glaucescens A TYPE II POLYKETIDE CYCLASE*', *The Journal of biological chemistry*, 279(36), pp. 37956–37963.

Tolmie, C., Smit, M. S. and Opperman, D. J. (2019) 'Native roles of Baeyer–Villiger monooxygenases in the microbial metabolism of natural compounds', *Natural Product Reports*, 36(2), pp. 326–353.

Tong, Y., Weber, T. and Lee, SY (2018) 'CRISPR/Cas-based genome engineering in natural product discovery', *Natural Product Reports*, Advance.

Tran, L., Broadhurst, R. W., Tosin, M., Cavalli, A. and Weissman, K. J. (2010) 'Insights into Protein-Protein and Enzyme-Substrate Interactions in Modular Polyketide Synthases', *Chemistry & Biology*, 17(7), pp. 705–716.

Umezawa, H., Ueda, M., Maeda, K., Yagishita, K., Kondo, S., Okami, Y., Utahara, R., Osato, Y., Nitta, K. and Takeuchi, T. (1957) 'Production and isolation of a new antibiotic: kanamycin.', *The Journal of antibiotics*, 10(5), pp. 181–188.

Ventola, CL (2015) 'The antibiotic resistance crisis: part 1: causes and threats.', *P & T : a peer-reviewed journal for formulary management*, 40(4), pp. 277–83.

Vurukonda, SSKP, Giovanardi, D. and Stefani, E. (2018) 'Plant Growth Promoting and Biocontrol Activity of Streptomyces spp. as Endophytes.', *International journal of molecular sciences*, 19(4), pp. 952–977.

Wagner, A., Segler, L., Kleinsteuber, S., Sawers, G., Smidt, H. and Lechner, U. (2013) 'Regulation of reductive dehalogenase gene transcription in Dehalococcoides mccartyi.', *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 368(1616), pp. 201–210.

Wang, G., Hosaka, T. and Ochi, K. (2008) 'Dramatic activation of antibiotic production in Streptomyces coelicolor by cumulative drug resistance mutations', *Applied and Environmental Microbiology*, 74(9), pp. 2834–2840.

Wang, Lei, Yuan, Meijuan and Zheng, Jianting (2019) 'Crystal structure of the condensation domain from lovastatin polyketide synthase.', *Synthetic and systems biotechnology*, 4(1), pp. 10–15.

Wang, M., Carver, JJ, Bandeira, N., *et al.* (2016) 'Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking', *Nature Biotechnology*, 34(8), pp. 828–837.

Ward, AC and Allenby, NEE (2018) 'Genome mining for the search and discovery of bioactive compounds: The Streptomyces paradigm', *FEMS Microbiology Letters*, 365(24), pp. 1–42.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, FT, de Beer, TAP, Rempfer, C., Bordoli, L., Lepore, R. and Schwede, T. (2018) 'SWISS-MODEL: homology modelling of protein structures and complexes', *Nucleic Acids Research*, 46(1), pp. 296–303.

Watve, M., Tickoo, R., Jog, M. and Bhole, B. (2001) 'How many antibiotics are produced by the genus Streptomyces ?', *Archives of Microbiology*, 176(5), pp. 386–390.

Weber, T., Blin, K., Duddela, S., Krug, D., Kim, HU, Bruccoleri, R., Lee, SY, Fischbach, MA, Müller, R., Wohlleben, W., Breitling, R., Takano, E. and Medema, MH (2015) 'antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters', *Nucleic Acids Research*, 43(1), pp. 237–243.

Westman, EL, Canova, MJ, Radhi, IJ, Koteva, K., Kireeva, I., Waglechner, N. and Wright, GD (2012) 'Bacterial Inactivation of the Anticancer Drug Doxorubicin', *Chemistry & Biology*, 19(10), pp. 1255–1264.

Wiederhold, NP (2017) 'Antifungal resistance: current trends and future strategies to combat.', *Infection and drug resistance*, 10, pp. 249–259.

Wilkinson, S. P. and Grove, A. (2006) 'Ligand-responsive Transcriptional Regulation by Members of the MarR Family of Winged Helix Proteins', *Current Issues in Molceular Biology*, 8, pp. 51–62.

Williams, PG, Buchanan, GO, Feling, RH, Kauffman, CA, Jensen, PR and Fenical, W. (2005) 'New Cytotoxic Salinosporamides from the Marine Actinomycete Salinispora tropica', *Journal of organic chemistry*, 70, pp. 6196–6203.

Wittmann, M., Linne, U., Pohlmann, V. and Marahiel, MA (2008) 'Role of DptE and DptF in the lipidation reaction of daptomycin', *FEBS Journal*, 275(21), pp. 5343–5354.

Wlodek, A., Kendrew, SG, Coates, NJ, Hold, A., Pogwizd, J., Rudder, Steven, Sheehan, LS, Higginbotham, SJ, Stanley-Smith, AE, Warneck, T., Nur-E-Alam, M., Radzom, M., Martin, CJ, Overvoorde, L., Samborskyy, M., Alt, S., Heine, D., Carter, GT, Graziani, EI, Koehn, FE, McDonald, L., Alanine, A., Rodríguez Sarmiento, RM, Chao, SK, Ratni, H., Steward, L., Norville, IH, Sarkar-Tyson, M., Moss, SJ, Leadlay, PF, Wilkinson, B. and Gregory, MA (2017) 'Diversity oriented biosynthesis via accelerated evolution of modular gene clusters', *Nature Communications*, 8(1), pp. 1206–1215.

Xiang, WS, Zhang, J. I., Wang, JiD, Jiang, L., Jiang, B., Xiang, ZD and Wang, XJ (2010) 'Isolation and Identification of Chlorinated Genistein from Actinoplanes sp. HBDN08 with Antioxidant and Antitumor Activities', *J. Agric. Food Chem*, 58, pp. 1933–1938.

Xu, G., Wang, J., Wang, L., Tian, X., Yang, H., Fan, K., Yang, K. and Tan, H. (2010) '"Pseudo" gamma-butyrolactone receptors respond to antibiotic signals to coordinate antibiotic biosynthesis.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 285(35), pp. 27440–8.

Xu, Z., Schenk, L. and Hertweck, C. (2007) 'Molecular Analysis of the Benastatin Biosynthetic Pathway and Genetic Engineering of Altered Fatty Acid-Polyketide Hybrids', *Journal of the American Chemical Society*, 129, pp. 6022–6030.

Xue, Y., Zhao, L., Liu, H. and Sherman, D. H. (1998) 'A gene cluster for macrolide antibiotic biosynthesis in Streptomyces venezuelae: Architecture of metabolic diversity', *Proceedings of the National Academy of Sciences of the United States of America*, 95(21), pp. 12111–12121.

Yanai, K. and Murakami, T. (2004) 'The Kanamycin Biosynthetic Gene Cluster from Streptomyces kanamyceticus', *The Journal of Antibiotics*, 57(5), pp. 351–354.

Yang, SC, Lin, CH, Sung, CT and Fang, JY (2014) 'Antibacterial activities of bacteriocins: application in foods and pharmaceuticals', *Frontiers in Microbiology*, 5, pp. 1–10.

Yeh, E., Cole, LJ, Barr, EW, Bollinger, JM Jr., Ballou, DP and Walsh, CT (2006) 'Flavin Redox Chemistry Precedes Substrate Chlorination during the Reaction of the Flavin-Dependent Halogenase RebH', *Biochemistry*, 45(25), pp. 7904–7912.

Yim, G., Thaker, MN, Koteva, K. and Wright, G. (2014) 'Glycopeptide antibiotic biosynthesis', *The Journal of Antibiotics*, 67(1), pp. 31–41.

Zhan, J., Watanabe, K. and Tang, Y. (2008) 'Synergistic Actions of a Monooxygenase and Cyclases in Aromatic Polyketide Biosynthesis', *ChemBioChem*, 9, pp. 1710–1715.

Zhang, S., Chen, T., Jia, J., Guo, L., Zhang, H., Li, C. and Qiao, R. (2018) 'Establishment of a highly efficient conjugation protocol for Streptomyces kanamyceticus ATCC12853', *MicrobiologyOpen*, pp. 747–755.

Zhang, W., Wang, L., Kong, L., Wang, T., Chu, Y., Deng, Z. and You, D. (2012) 'Unveiling the Post-PKS Redox Tailoring Steps in Biosynthesis of the Type II Polyketide Antitumor Antibiotic Xantholipin', *Chemistry & Biology*, 19(3), pp. 422–432.

Zhang, Z., Pan, HX and Tang, GL (2017) 'New insights into bacterial type II polyketide biosynthesis.', *F1000Research*, 6, pp. 172–184.

Zheng, X., Hu, GQ, She, ZS and Zhu, H. (2011) 'Leaderless genes in bacteria: clue to the evolution of translation initiation mechanisms in prokaryotes', *BMC Genomics*. BioMed

Central, 12(1), pp. 361–373.

Zhou, M., Jing, X., Xie, P., Chen, W., Wang, T., Xia, H. and Qin, Z. (2012) ‘Sequential deletion of all the polyketide synthase and nonribosomal peptide synthetase biosynthetic gene clusters and a 900-kb subtelomeric sequence of the linear chromosome of Streptomyces coelicolor’, *FEMS Microbiology Letters*, 333(2), pp. 169–179.