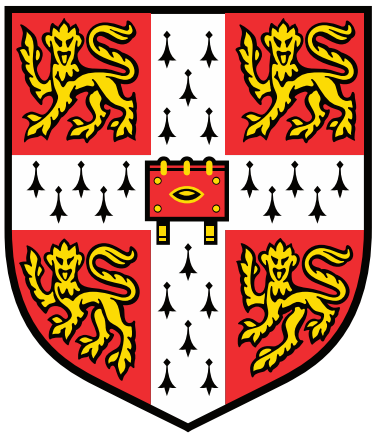
Genome Mining for Actinomycete Biosynthetic Gene Clusters



F. H. Dudbridge

Clare College

Department of Biochemistry

University of Cambridge

This dissertation is submitted for the degree of Doctor of Philosophy

September 2017

Deposit & Copying of Dissertation Declaration

UNIVERSITY OF CAMBRIDGE

Board of Graduate Studies

Please note that you will also need to bind a copy of this Declaration into your final, hardbound copy of thesis - this has to be the very first page of the hardbound thesis.

1 Surname (Family Name)	Forenames(s)	Title
Dudbridge	Frederic Henry	Mr

2 Title of Dissertation as approved by the Degree Committee
Genome mining for actinomycete biosynthetic gene clusters

In accordance with the University Regulations in *Statutes and Ordinances* for the PhD, MSc and MLitt Degrees, I agree to deposit one print copy of my dissertation entitled above and one print copy of the summary with the Secretary of the Board of Graduate Studies who shall deposit the dissertation and summary in the University Library under the following terms and conditions:

1. Dissertation Author Declaration

I am the author of this dissertation and hereby give the University the right to make my dissertation available in print form as described in 2. below.

My dissertation is my original work and a product of my own research endeavours and includes nothing which is the outcome of work done in collaboration with others except as declared in the Preface and specified in the text. I hereby assert my moral right to be identified as the author of the dissertation.

The deposit and dissemination of my dissertation by the University does not constitute a breach of any other agreement, publishing or otherwise, including any confidentiality or publication restriction provisions in sponsorship or collaboration agreements governing my research or work at the University or elsewhere.

2. Access to Dissertation

I understand that one print copy of my dissertation will be deposited in the University Library for archival and preservation purposes, and that, unless upon my application restricted access to my dissertation for a specified period of time has been granted by the Board of Graduate Studies prior to this deposit, the dissertation will be made available by the University Library for consultation by readers in accordance with University Library Regulations and copies of my dissertation may be provided to readers in accordance with applicable legislation.

3 Signature	Date
F. Dudbridge	02/06/2018

Corresponding Regulation

Before being admitted to a degree, a student shall deposit with the Secretary of the Board one copy of his or her hard-bound dissertation and one copy of the summary (bearing student's name and thesis title), both the dissertation and the summary in a form approved by the Board. The Secretary shall deposit the copy of the dissertation together with the copy of the summary in the University Library where, subject to restricted access to the dissertation for a specified period of time having been granted by the Board of Graduate Studies, they shall be made available for consultation by readers in accordance with University Library Regulations and copies of the dissertation provided to readers in accordance with applicable legislation.

Preface

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text

This thesis does not exceed the prescribed word limit.

Acknowledgements

I would first like to thank Professor Peter Leadlay for inviting me to study in his lab. It truly has been a fantastic and inspirational place to work. Thank you for your wise words, encouragement and patience over the four years that I have been here.

Throughout my time here in the lab I have been surrounded by people generous both with their time and knowledge. Many thanks to Dr Yongjun Zhou, Dr Anya Luhavaya, Dr Fanglu Huang (who put up with my mess daily), Dr Markiyan Samborskyy, Dr Annbel Murphy and Dr Hui Hong who have answered many questions and been there to guide me over the four years. I am grateful and have been very lucky enough to share a bench with so many fantastic people: Marie, Annabel, Anna, Anya, Nik, Kat, Constance, Jake, Rory, Oksana, Katha amongst many other. Particular thanks to Karen, who has been a great support throughout.

We have had some amazing overseas students studying with us in the Leadlay lab. Amongst these are Felix, Hannah, Vinzent Christian, Roel, Karen, Constance and Joachim. Not only have you become good friends, you have enhanced my language skills.

I would like to thank Shilo, John Lester and the Sequencing team and Clova and Janet, for keeping me fed throughout my PhD.

A special thanks to Adam Wurr for his friendship and support throughout my PhD. Cambridge would not have been the same without you, (or our grandfathers) . Steven Lowe, you have been a great friend on a similar PhD journey, we complement well (coys). Emma, Laura, Chrysa, Callum, Lois, Grace, Nick and many more.

Gavin Davey, if you were still with us, you would have been proud.

Finally, I would like to dedicate this work to my parents, Simon and Felicity Dudbridge. Thank you for all your support, and thank you for putting up with me. Thank you.

Genome mining for actinomycete biosynthetic gene clusters

Frederic Dudbridge

Whole-genome sequencing has shown that the large (8-12 Mbp) genomes of *Streptomyces* and allied genera of Gram-positive filamentous bacteria house a rich and previously underestimated repertoire of gene clusters for biosynthesis of specialised metabolites, including antibiotics, immunosuppressants and anticancer compounds. Many of these clusters remain uncharacterised because they are not expressed under the culture conditions used. Even for strains from which a specific compound has been identified, the challenge remains to link the compound to its gene cluster, and to develop procedures for analysing and manipulating the biosynthetic pathway. In this work, three strains have been studied that address different aspects of this challenge.

Streptomyces sp. DSM4137 is a genetically amenable strain and a notably prolific producer of diverse natural products, but also has multiple biosynthetic gene clusters that remain uncharacterised. In an attempt to differentiate those clusters where the genes are expressed from those that are essentially silent, the transcriptome of DSM4137 was analysed using total RNASeq and the results were used to inform analysis of HPLC/MS data of extracts under the same conditions. There was shown to be good correlation between the RNASeq results and the pattern of metabolites produced, suggesting that RNASeq may be a useful complement in the search for novel gene clusters.

In contrast, *Saccharopolyspora spinosa*, producer of the valuable insecticidal spinosyns, is not genetically amenable. A new technique has been developed for the mobilisation of an entire biosynthetic gene cluster and refactoring attempted to increase the production of spinosyns in a heterologous strain. Total transcriptome was analysed by RNAseq to give an insight into the regulation of the WT strain, helping identify future methods for strain manipulation for increase yields.

Albucidin is a remarkable insecticidal nucleoside antibiotic reported from the strain *Streptomyces albus* subsp. *chlorinus*, but produced in negligible amounts and by an unknown pathway. The sequenced genome was analysed to identify a candidate cluster, its identity was confirmed by mutation, and a mechanism is proposed for its production based on radical SAM-mediated ring contraction. The cluster was successfully heterologously expressed in the genetically amenable host *Steptomyces albus* J1074.

Table of Contents

Abbreviations

6-dEB 6-deoxyerythronolide B

A600 absorbance at 600 nm

ACP acyl carrier protein

AntiSMASH antibiotics & Secondary Metabolite Analysis SHell

AT acyltransferase

Amp ampicillin

Apr apramycin

BAC bacterial artificial chromosome

BLAST basic local alignment search tool bp base pair

C condensation domain

Cam chloramphenicol

Carb carbenicillin

CoASH coenzyme A

Da Dalton

DEBS 6-deoxyerythronolide B synthase

DH dehydratase

DMSO dimethyl sulphoxide

DNA deoxyribonucleic acid DQF

EDTA ethylenediaminetetraacetic acid

ER enoylreductase

ESAC E. coli-Streptomyces artificial chromosome

ESI electrospray ionisation

FA fatty acid

FA formic acid

FAS fatty acid synthase

gDNA genomic DNA

HEPES N-(2-hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid)

HPLC high performance liquid chromatography HRMS high resolution mass spectrometry

Int integrase

IPTG isopropyl-β-D-thiogalactopyranoside

Kan kanamycin

kbp kilo base pairs

KR ketoreductase

KS ketosynthase

LCHR linear plus circular homologous recombination LLHR linear plus linear homologous recombination

MalCoA malonyl coenzyme A

Mb mega base pairs

MDR multidrug-resistant

MIC minimum inhibitory concentration

mMalCoA methylmalonyl coenzyme A mon monensin

MS mass spectrometry

MT methyltransferase

m/v mass to volume ration

m/z mass to charge ratio

NADH nicotinamide adenine dinucleotide, reduced form

NADPH nicotinamide adenine dinucleotide phosphate, reduced form Nal nalidixic acid Ni-NRPS nonribosomal peptide synthetase

ORF, orf open reading frame

PAC P1-derived bacterial artificial chromosome

PAGE polyacrylamide gel electrophoresis

PCP peptidyl carrier protein

PCR polymerase chain reaction

PDB protein data bank

PEG polyethylene glycol

Phyre protein homology/analogy recognition engine

PK polyketide

PKS polyketide synthase

PMSF phenylmethanesulfonyl fluoride

PPTase 4'-phosphopantetheinyl transferase

SAM S-adenosyl methionine

SAR structure–activity relationship

SDS sodium dodecyl sulphate

ssDNA single-stranded DNA

TAR transformation-associated recombination

TE thioesterase TFA

Thio thiostrepton

TIC total ion current

TSB tryptic soya broth

UV ultra violet

v/v volume to volume ratio

WT wild type

YAC yeast artificial chromosome

Chapter 1. Introduction

1.1 Natural products and their cellular targets

The evolution of natural products as therapeutics and biocontrol agents

Natural products are organic chemicals, produced by a living organism, the vast majority of which are not essential for cell growth, development and division. This feature led to them being commonly referred to as "secondary metabolites" (Bu'Lock, 1965) but appropriately given their importance for life on earth the more helpful term "specialised metabolites" has replaced it. For thousands of years, natural products have been at the forefront of medicine and agriculture (Ji et al., 2009; Dias et al., 2012; Cragg and Newman, 2012). Clay tablets in Mesopotamia from 2600 BC record hundreds of natural products from plants, including *Cedrus* (cedar) *Commiphora* (myrrh) and *Papaver somniferum* (opium poppy), many of which are still used to this day (Brahmachari, 2012). Studies of ancient papyri from 2900 BC have likewise documented many Egyptian plant-derived medicines (Gallin, 2012), and the Chinese Materica Medica and the written accounts of Indian Ayurvedic medicine can also be traced back 2000 years (Gallin, 2012, Cragg and Newman, 2012).

Even in the modern era, leading causes of premature death included tuberculosis, cholera, and other infectious diseases. The ground-breaking work of Louis Pasteur, Joseph Lister and Robert Koch on (particularly) the microbial agents causing infectious disease paved the way for Alexander Fleming's discovery in 1928 of the antimicrobial activities of the fungal metabolite penicillin (Aldridge et al., 1999). In 1939, René Dubos discovered gramicidin, a secondary metabolite from *Bacillus brevis,* found to kill Gram-positive bacteria (Van Epps, 2006) and this became the first natural product antibiotic to be deliberately sought and clinically tested. Inspired by Dubos' work, Selman Waksman, and his graduate student Albert Schatz, isolated streptomycin from *Streptomyces griseus (*Ginsberg, 2005). Streptomycin had antibacterial activities and, importantly, was shown to be active against *Mycobacterium tuberculosis*. The discovery of such antibiotics changed the practice of medicine. Operations could be performed at lower risk, and life-threatening diseases, most notably at the time tuberculosis, could be treated.

The systematic screening of (principally) soil-borne actinomycetes for their antibiotic products that followed upon these early discoveries has been described as a "golden age" of antibiotic research:

between 1950 and 1960, one half of the currently-used therapeutics and most of the currently-known classes of antibiotics were discovered (Davies, 2006) (Figure 1.1).

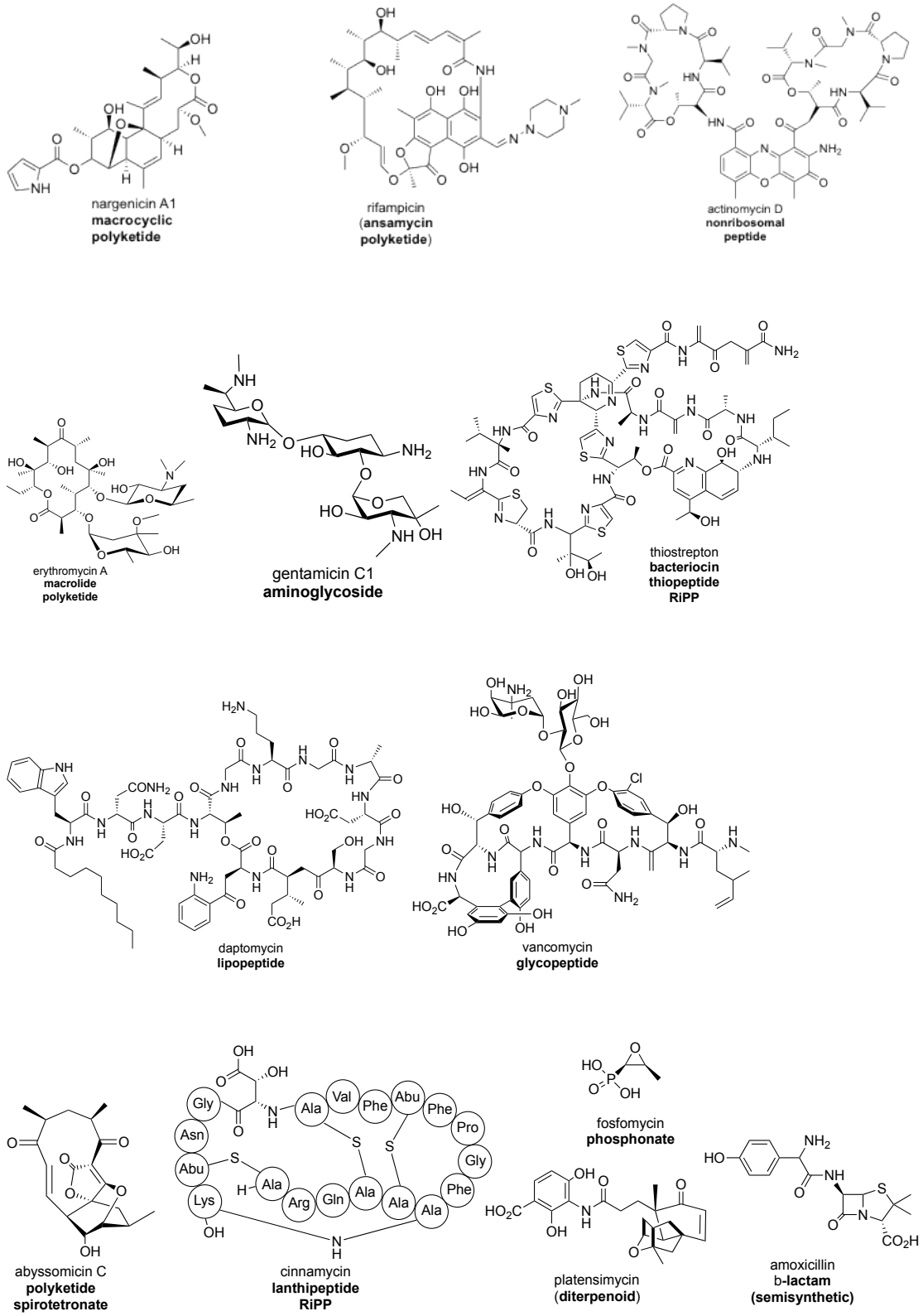


Figure 1.1. Natural product antibiotics acting on diverse bacterial cellular targets

The natural products obtained have also been tested for other therapeutic indications, and with success. For example, over half of cancer therapeutics in current use are based on, or inspired by, natural products (Cragg and Newman, 2012). Examples include taxol, vinblastine, and camptothecin from terrestrial plants; halichondrin B, and dolastatins from the marine environment; and bleomycin, doxorubicin, staurosporine and rapamycin from micro-organisms. Natural products have also played an important role in the chemistry of crop protection, either as the starting point for lead generation and optimisation or as active ingredients (Pachlatko, 1998; Cantrell et al., 2012; Lamberth, 2016). For example,To pyrethrin I from the daisy *Tanacetum cinerariaefolium* gave rise to the large family of synthetic pyrethroid insecticides (more stable to sunlight and with a wider activity spectrum than the natural product), and strobilurin A from the mushroom *Strobilurus tenacellus* led to the strobilurin fungicides (Sauter et al., 1999). Natural products marketed successfully as ingredients in biocontrol agents include the herbicide bialaphos, the fungicide blasticidin S, and the insecticidal spinosyns, as discussed in more detail in Chapter 5 of this thesis. These compounds are all produced by actinomycete bacteria.

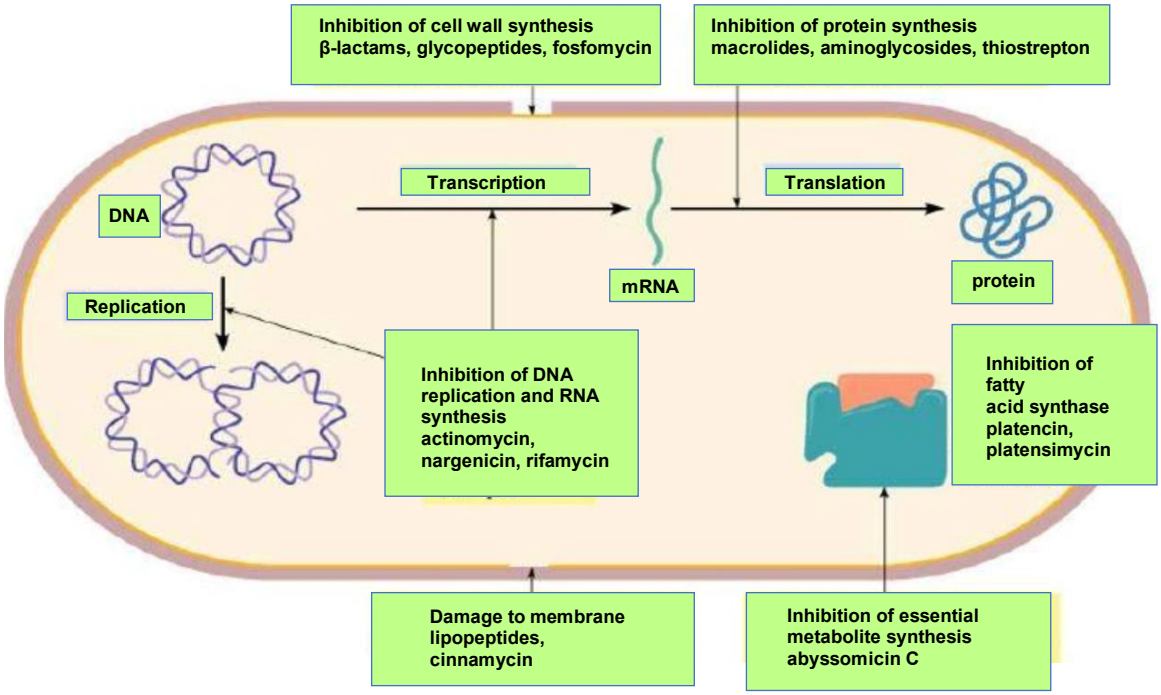


Figure 1.2. The cellular targets of antibiotics. (Adapted from Totora et al)

1.1.2 Natural product classification and modes of action

Based on their chemical structure, natural products can be divided into a number of major classes (Figure 1.1), based on common biosynthetic pathways. However, these pathways produce a huge chemical diversity, from relatively simple structures like the phosphonate fosfomycin to large and complex compounds with many stereocentres and/or complex ring structures, as in the thiopeptide thiostrepton or the terpenoid platencin (Figure 1.1). As illustrated in Figure 1.2, the principal mode of action of the antibacterial natural products in Figure 1.1 has been determined. Closely-related antibiotics tend to affect similar targets: macrolides and aminoglycosides, for example, typically inhibit one or more steps in protein synthesis. However, there are many exceptions to this generalisation.

1.1.3 Prokaryotic producers of natural products

The focus of the work in this thesis is on natural product biosynthetic pathways in bacteria, but it is nevertheless important to acknowledge the specialised metabolism of plants and fungi, given the intimate interactions between these and bacteria. Indeed, endophytic microbes (Kusari et al., 2012a; van der Meij et al., 2017) are responsible for the production of numerous host metabolites, including medicinally important taxol (Stierle et al., 1993) and azadirachtin (Kusari et al., 2012b), as well as toxins (Partida-Martinez and Hertweck, 2005). Plants themselves are estimated to be the source for over 200,000 natural products, including many valuable alkaloids, and terpenoid compounds such as artemisinin. Remarkably, this rich chemistry arises from only a few key scaffold-generating reactions followed by diverse enzyme-catalysed tailoring reactions (Anarat-Cappillino and Sattely, 2014) which are often characterised by relatively broad substrate specificity (Weng et al., 2012). Biosynthetic investigations of plant biosynthesis have until recently have been hindered by lack of detailed information about the genes and enzymes involved; and by the barriers to expression of plant genes in microbial cells (Nielsen and Keasling, 2016). Successful engineering of the plant pathway for artemisinin into yeast is estimated to have taken 150 person-years of work (Kwok, 2010). However, recent success in engineering *Nicotiana benthamia* for *in planta* gram-scale production of complex triterpenes, using the process known as agro-filtration, is likely to accelerate research in this area (Reed et al., 2017).

Filamentous fungi are also rich sources of natural products, particularly via polyketide, terpene and non-ribosomal peptide biosynthetic pathways (Cox, 2007; Macheleidt et al., 2016). Their biosynthesis has been studied not only because of the medical significance of penicillins and cephalosporins, or of the immunosuppressant cyclosporin, but also because

fungal toxins such as aflatoxins and ergot alkaloids offer significant challenges to health and to food security. As with plant-derived specialised metabolites, advances in genomics and molecular biology are bringing about rapid advances in understanding and manipulation of these pathways (Boecker et al., 2016; Motoyama and Osada, 2016).

Among prokaryotes, the ability to produce specialised metabolites is distributed very unequally, although more bacteria than previously thought have now been shown to produce such compounds. Established producers other than the actinomycetes include *Bacillus* spp. (Chen et al., 2006; Moldenhauer et al., 2007; Schneider et al., 2007; Kevany et al., 2008), *Pseudomonas* spp. (El-Sayed et al., 2003; Gross and Loper, 2009; Trippe et al., 2013) and gliding bacteria (*Myxobacteria*) (Herrmann et al., 2017). These have now been joined by the Gram-negative, plant-associated *Pantoea* spp., which produce for example pantocin (Brady et al., 1999), dapdiamides (Dawlaty et al., 2010), andrimid (Jin et al., 2006) and agglomerins (Kanchanabanca et al., 2013); and by *Photorhabdus* and *Xenorhabdus* spp. (Bode et al., 2009). *Burkholderia* spp. are particularly prolific producers, producing for example enacyloxin (Mahenthiralingam et al., 2011), rhizoxin (Partida-Martinez and Hertweck, 2005), and jagaricin (Graupner et al., 2012). The first examples have also been found among anaerobic *Clostridia* spp. (Behnken et al., 2012). As discussed in section 1.3.5, the vast majority of bacteria not culturable under conventional laboratory conditions may represent an additional reservoir of novel chemical diversity. *Streptomyces* and allied genera remain, however, the principal sources of bacterial chemical diversity so far uncovered (Katz and Baltz, 2016).

1.1.4 The life cycle of *Streptomyces* and related actinomycetes

The filamentous actinomycetes are Gram-positive bacteria of characteristically high G+C content. Although they have colonised many ecological niches they are predominant in soil. They produce, like fungi, thread-like branching hyphae that eventually form a vegetative mycelial mat, and produce extracellular degradative enzymes that allow them to exploit insoluble polymers such as dead plant tissue (Chater, 2006). Phylogenetic comparisons suggest that the last common ancestor of actinomycetes appeared 450 Ma, coinciding with the colonisation of the land by plants (Embley and Stackebrandt, 1994).

After initial growth of the vegetative mycelium, and as nutrients become depleted, a developmental change involving several tens of regulatory genes (Chandra and Chater, 2012) leads to the formation in *Streptomyces* of aerial hyphae perpendicular to the surface. As in the vegetative mycelium, the growing tips consist of cellular compartments containing multiple

copies of the genome. When their growth ceases, multiple septa form, and the individual cells round off to form heat-labile but desiccation-resistant spores (Angert, 2005) each containing a single genome copy.

Remarkably, at the onset of aerial mycelium formation the vegetative or substrate mycelium is broken down by a mixture of degradative enzymes to provide a temporary food supply for sporogenesis (Bibb, 2005; Liu et al., 2013). This transitional phase of the life cycle coincides in many cases with the production of antibiotics and other specialised metabolites, consistent with a major role for antibiotics in protecting this nutrient supply from competing fungi and bacteria (Barka et al., 2016). The orchestration of these events is under complex regulation which is only partly understood (Rigali et al., 2006; Hesketh et al., 2007; den Hengst et al., 2010). For the model organism *Streptomyces coelicolor*, transposon mutagenesis has revealed nearly 350 genes that modulate the production of the antibiotic prodigiosin (Xu et al., 2017).

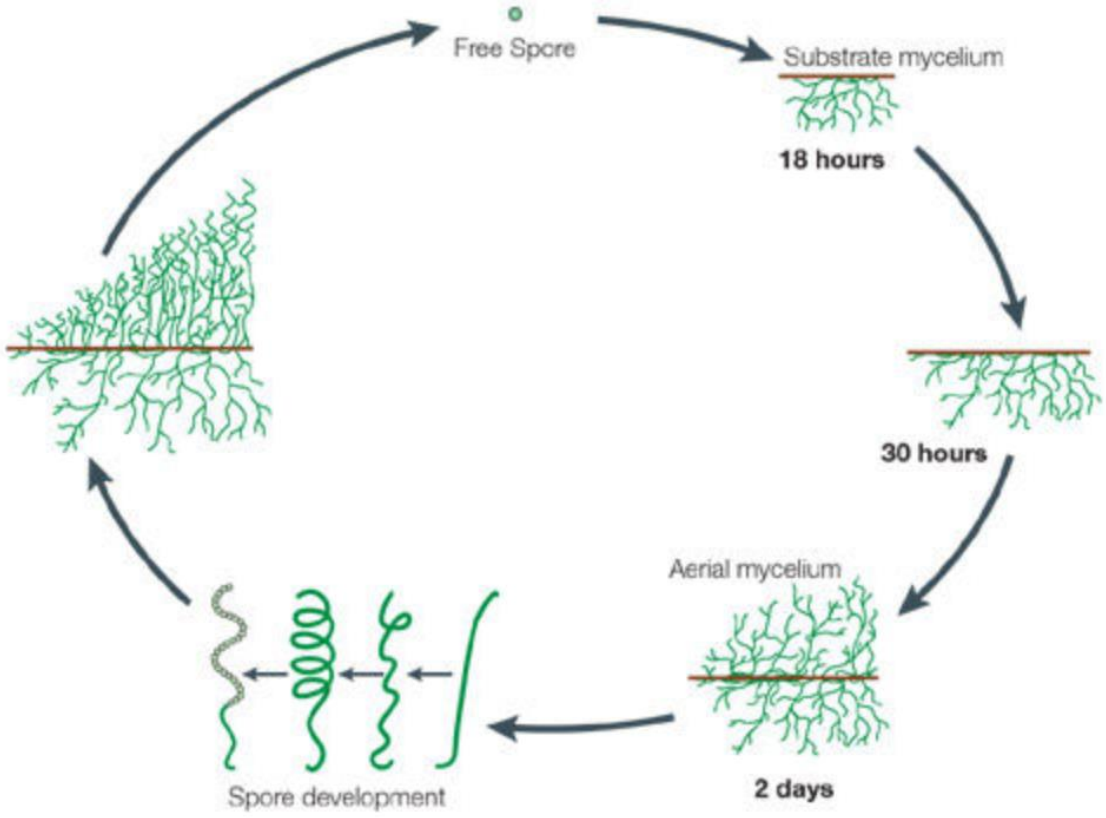


Figure 1.3. The life cycle of *Streptomyces*. Adapted from Angert (2005).

1.1.5 Function and evolution of actinomycete natural products

Given the extraordinary range of specialised metabolites produced by *Streptomyces* and allied bacteria, there has been considerable debate over their evolution, their function and how they contribute to fitness (Challis and Hopwood, 2003). Their complex, highly regulated biosynthetic pathways (see section 1.3) require both energy and precursors (Haslam, 1994), and the retention and curation of the corresponding gene sets also places significant demands on the organism. In support of the idea that antibiotic production prevents competition from other bacteria and fungi, *Streptomyces* typically produce extracellular cellulase, chitinase and glucanase enzymes that can actively attack fungal cells (Zhao et al., 2013) and there is evidence that *Streptomyces* spp. are at least facultatively predatory towards other bacteria (Kumbhar et al., 2014). Wellington and colleagues (Laskaris et al., 2010) have shown, for closely-related *Streptomyces griseus* strains in soil, that antibiotic resistance co-evolves in the presence of antibiotic biosynthetic capability, also consistent with the idea of antibiosis during competition for resources.

However, any general theory must take into account that many specialised metabolites are not antibiotic at realistic concentrations (Yim et al., 2006), and many biosynthetic gene clusters appear not to be active under laboratory conditions. Certainly, antibiotics have been shown at low concentration to regulate gene expression (Goh et al., 2002) and some specialised metabolites may act as signalling molecules within mycelia or between a microbial community (Yim et al., 2007). Davies has proposed that the origin and evolution of specialised metabolite biosynthesis lies in their original catalytic roles in the RNA world (Davies et al., 1992). In contrast, Firn and Jones have argued that possession of the biosynthetic machinery rather than the evolution of a particular compound confers the real selective advantage to the producer (Firn and Jones, 2000). They propose that maintenance of the biosynthetic genes is selected for to provide the potential for evolution and the ability of the producer to adapt to the changing environment by rapidly producing a novel metabolite. Natural products are often found as mixtures of closely structurally-related compounds, which rather than being evidence for a 'primitive', less selective enzymology, may instead hint at a flexibility that provides material for further evolution while maintaining production of a useful molecule (Firn and Jones, 2003). Jensen (Jensen, 2016), based on analysis of the distribution of natural product biosynthetic genes in 75 strains of *Salinospora* spp., expresses a similar view in terms of the positive fitness value of maintaining a large and dynamic repertoire of such pathway genes within a population, to meet fluctuating needs, as opposed to within every member of that population. Clearly, though, much remains to be learned about the ecological roles of specialised metabolites.

1.1.6 Organization and regulation of biosynthetic gene clusters

In both bacteria and filamentous fungi, the biosynthetic genes governing the production of a given specialised metabolite are normally found clustered together on the bacterial chromosome (Osbourn, 2010). In general, the initial phase of biosynthesis involves the assembly of a scaffold or template structure, which is then acted upon by tailoring enzymes (often including glycosyltransferases, hydroxylases, and methyltransferases) to create diversity and confer bioactivity. In addition to the enzymes catalysing each step, specific mechanisms are needed to transport the end-products out of the cell; to regulate the timing and extent of production; and crucially, for antibiotic biosynthetic clusters, genes are present conferring self-resistance to the antibiotic product of the pathway. Stone and Williams (Stone and Williams, 1992) have argued that this integrated nature of biosynthetic gene clusters is conclusive evidence for evolutionary advantage of their acquisition by the bacterium. At a practical level, the co-location of resistance genes with biosynthetic gene clusters has greatly assisted experimental analysis of these pathways.

For many years it has been known that the entire biosynthetic potential of a typical *Streptomyces* is not expressed under a single set of conditions, and striking changes in the specialised metabolites produced can be observed by screening a range of media and growth conditions - the One Strain Many Compounds (OSMAC) approach (Bode et al., 2002). Nevertheless, whole-genome DNA sequencing of actinomycetes has transformed our view of the organisation, distribution and number of gene clusters for specialised metabolites that are present in these large bacterial chromosomes (typically between 7-11 Mbp). The chromosome of typical *Streptomyces* spp. is linear (Bentley et al., 2002, Ikeda et al., 2003) while that of other genera such as *Micromonospora* spp. (Trujillo et al., 2014) and *Saccharopolyspora* spp. (Oliynyk et al., 2007) is circular, but in each case genes encoding essential primary metabolism are located in a core region of roughly half of the chromosome centred on the origin of replication. In contrast, genes encoding specialised metabolite biosynthesis are located in the region away from the origin of replication. Certain strains house additional biosynthetic gene clusters on giant plasmids (Medema et al., 2010).

In *Streptomyces*, the timing and level of production of a given specialised metabolite are determined by the interplay of a complex set of regulatory proteins, in turn responding to environmental and nutritional signals (van Wezel and McDowall, 2011; Martín and Liras, 2013). Pathway-specific regulators, or cluster-situated regulators (CSRs) (Huang et al., 2005) control the transcription of cluster genes, while global or pleiotropic regulators such as the γ-butryolactones control multiple clusters, and may mediate crosstalk between them (Vicente et

al., 2015; Niu et al., 2016). There may also be direct crosstalk between clusters: for example, a CSR for the polyene candicidin has been shown to activate a cluster for antimycin (McLean et al., 2016). Among the most prominent of CSRs are the (normally activating) *Streptomyces* antibiotic regulatory proteins (SARPs) (Wietzorrek and Bibb, 1997); members of the TetR superfamily of transcriptional repressors (Ramos et al., 2005); and ATP-dependent activators of the LuxR family (De Schrijver and De Mot, 1999).

1.1.7 Antibiotic self-resistance

A producing strain must also possess at least one mechanism to protect itself against its own antibiotic. A common strategy is modification of the target of the antibiotic. For antibiotics affecting protein synthesis, this is often achieved by specific methylation of rRNA residues by *S*-adenosylmethionine-(SAM)-dependent methyltransferases, as for example in the methylation of A2058 in the 16S rRNA of several macrolide-producing actinomycetes (Zalacain and Cundliffe, 1990). For glycopeptides and lantibiotics that target lipid II in cell wall biosynthesis, the strategy is to synthesise modified peptidoglycan (Stegmann et al., 2015). For other antibiotics, like the fatty acid synthase inhibitor thiolactomycin (Tang et al., 2015; Tao et al., 2016) and the aminoacyl tRNA synthetase inhibitor borrelidin (Olano et al., 2004), the biosynthetic gene cluster houses an isoenzyme of the target enzyme that is impervious to the action of the inhibitor.

In some strains that produce inhibitors of protein synthesis, the ribosomes remain unmethylated and sensitive to the antibiotic. Instead, the antibiotic itself is inactivated by phosphorylation, as in streptomycin biosynthesis in *Streptomyces griseus* (Sugiyama et al., 1983), or by acetylation (Sugiyama, 2015). The protective group is enzymatically removed after export. Even more commonly, so-called ABC transporters (Mendéz and Salas, 2001) catalyse the active, ATP-driven efflux of antibiotic molecules (Martín et al., 2005), with an inactive precursor often inducing synthesis of the pump protein (Hopwood, 2007). In contrast, a different group of efflux systems, the major facilitator superfamily exporters, are driven by chemiosmotic ion gradients (Martín et al., 2005).

1.1.8 Resistance development to natural product therapeutics and biocontrol agents

Unfortunately, the widespread and often injudicious use of broad-spectrum antibiotics has led to an inexorable rise of pathogen resistance, causing arguably one of the biggest problems of the 21st century (Brown and Wright, 2016). For example, *Staphylococcus aureus* infections cause painful skin and soft tissue conditions or life-threatening systemic infection infections. In 2002, a strain of methicillin-resistant *S. aureus* (MRSA) was found resistant to vancomycin, an antibiotic used as a last resort leading to what is now known as VRSA (vancomycin-resistant *S. aureus*) (Gardete and Tomasz, 2014).

The emergence of resistance has been so rapid that since 2009 several pathogens have been isolated that are resistant to essentially all front-line antibiotics; these include Gram-negative pathogens such as polymyxin-resistant and/or carbapenem-resistant *Pseudomonas aeruginosa* (Walsh and Amyes, 2004), carbapenem-resistant *Klebsiella pneumoniae* (Fair and Tor, 2014), and multi drug-resistant (MDR) *Mycobacterium tuberculosis* (Velayati et al., 2009).

Mechanisms of resistance to antibiotics mirror three of those involved in self-protection of producing strains: modification of the antibiotic target (cell wall, cell membrane, or protein), acquisition of efflux pumps capable of removing the antibiotic from the organism, and antibiotic-inactivating enzymes. The latter are especially serious, as the enzymes are normally encoded by a single gene and can be easily transferred between bacterial pathogens. The origin of antibiotic-modifying proteins is often the producers themselves, and often such genes are located on mobile genetic elements, contributing to the spread of resistance by horizontal gene transfer.

Microbial resistance is driven by four main factors: antibiotic overuse, overdose, use of broad-spectrum antibiotics, and non-compliance. Increasing quantities of antibiotics used for non-human applications (estimated 80% of all antibiotics sold in the USA), such as farming and fishing, has also led to antibiotic-contaminated waste being released into the environment. This creates evolutionary pressure for bacteria to evolve and acquire more resistance genes, and highlights the importance of discovering new antibiotics as well as curbing the overuse of existing compounds.

Similar issues surround the use of specialised metabolites, or compounds based on them, in crop protection. Pyrethroids target the sodium channel in insect cells (Vais et al., 1991) and there are now instances of emerging resistance among malaria-carrying mosquitoes (*Anopheles gambiae*) to these compounds (Hien et al., 2017). An *N*-acetyltransferase, the product of the *bar* gene, is responsible for resistance to bialaphos in the

producing *Streptomyces viridochromogenes* strain and in other *streptomycetes* (Bedford et al., 1991). The *bar* gene has actually been used in plant engineering to create plants resistant to bialaphos (D'Halluin et al., 1992), but the role of bialaphos in agriculture has been limited by the overwhelming success of the synthetic compound phosphonylmethylglycine (glyphosate), the basis of major herbicide-resistant crops. Unfortunately, the resistance of weed species to glyphosate (which targets enoylpyruvylshikimate (EPSP) synthase in the shikimate pathway leading to aromatic amino acids) is growing rapidly (Owen and Zelaya, 2005; Casida and Durkin, 2017), as is the resistance of fungi to the strobilurins, which target cytochrome bc1 in the electron transport chain (Casida and Durkin, 2017).

Resistance has also been seen with the more recently introduced insecticidal spinosyns (marketed as a mixture known as Spinosad): deletion of one of the subunits of the acetylcholine receptor have led to Western flower thrips becoming 180,000-fold more resistant to the insecticide than unexposed populations (Hou et al., 2014). Perhaps the most serious current threat to food security in Europe is the fact that since the 1980s the serious fungal disease of wheat, *Zymoseptoria tritici* (formerly *Mycosphaerella graminicola*), septoria leaf blotch, has developed widespread resistance to two synthetic pesticide classes (Omrane et al., 2015), and its resistance to azoles, the remaining useful class, is progressing. Since development of resistance is inexorable, there is an urgent need for new biocontrol agents.

1.2 Genome-level analysis of natural product biosynthesis

1.2.1 The evolution of DNA sequence analysis in actinomycetes.

The first (portions of) biosynthetic gene clusters from actinomycetes to be cloned and characterised by sequencing were parts of the relatively compact clusters for the isochromanequinones actinorhodin (5 kbp) from the model organism *Streptomyces coelicolor* A3(2) (Malpartida and Hopwood, 1986; and granaticin (6.5 kbp) from *Streptomyces violaceoruber* (Sherman et al., 1989). The completion of the sequencing of the entire *act* cluster was reported (~ 25 kbp) a little later (Fernández-Moreno et al., 1992). The sequencing of the erythromycin gene cluster was the first for a complex reduced polyketide, (Cortés et al., 1990; Donadio et al., 1991; Bevitt et al., 1992). These projects used manual DNA sequencing by the Sanger method, for which special measures were needed to sequence "difficult" dsDNA from actinomycete genomes (>70% G+C). Gradually, equipment to automate the Sanger

method was introduced and this remained the mainstay of biosynthetic cluster sequencing projects until the introduction in 2005-2006 of massively-parallel sequencing instruments that could maximise the advantage of shotgun analysis of large pieces of genomic DNA, either pyrosequencing in pico-wells (454 instruments) Margulies et al., 2005) or solid-phase fluorescence-based sequencing on lawns of immobilised oligonucleotides (Solexa/Illumina instruments) (Voelkerding et al., 2009).

The further development of (particularly) the Illumina methodology has fuelled the current explosive increasing in whole-genome sequencing of bacterial and fungal organisms producing specialised metabolites. Elegant new third-generation technologies including single-molecule real time (SMRT) sequencing (Rhoads and Au, 2015) and nanopore sequencing (Goodwin et al., 2015; Heather and Chain, 2016) allow far superior readlengths, but the high cost of SMRT sequencing and the high error rate of nanopore sequencing prevents them - as yet - supplanting Illumina sequencing in most research laboratories. For a critical review of the state of the art in uncovering biosynthetic gene clusters by sequencing actinobacteria, see (Gomez-Escribano et al., 2016).

The first actinomycete to be fully sequenced was *Mycobacterium tuberculosis* (Cole et al., 1998) in the expectation that this would lead to the identification of multiple new targets for anti-TB therapies (Lechartier et al., 2014). Soon after, the complete genome sequence for the model actinomycete *Streptomyces* coelicolor (A3)2 was published (Bentley et al. 2002). It is a linear genome and at the time was the largest bacterial genome sequenced at 8,667,507 bp. In the following six years the genome sequence of a handful of other antibiotic-producing actinomycetes was completed, including that of avermectin-producing *Streptomyces avermitilis* (Ikeda et al., 2003), erythromycin-producing *Saccharopolyspora erythraea* (Oliynyk et al., 2007), salinosporamide-producing Salinispora tropica (Udwary et al., 2007); and streptomycin-producing *Streptomyces griseus* (Ohnishi et al., 2008). Surprisingly*, S. erythraea* turned out to have a circular, not a linear genome (Oliynyk et al. 2007), but all these genomes were large (>8 Mbp) and demonstrated the same unexpected features: first, biosynthetic clusters were found to be mostly positioned outside a central core region containing primary metabolic genes and centred on the origin of replication. Secondly, many more clusters were found encoded than there were specialised metabolites known to be produced by the organism, so that on average biosynthetic clusters accounted for 10% of the genome sequence (Nett et al., 2009). The ability to sequence these genomes, and the revelation of the existence of so many "cryptic" clusters has led to a complete change in the way that the biosynthesis of specialised metabolites is studied. There are presently nearly 500 genome sequences for bacterial producers of natural products deposited in public databases. The

majority are draft sequences of variable quality, but approximately 100 of them are complete or near-complete.

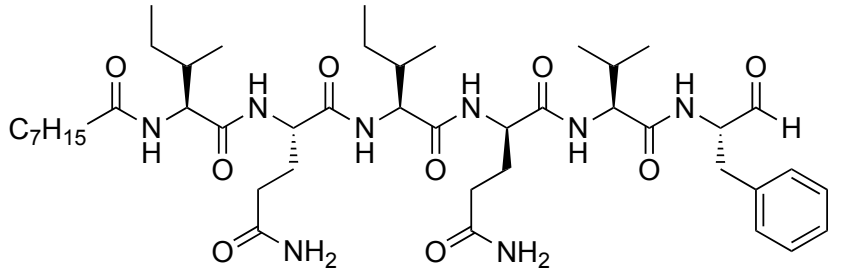
1.2.2 Prediction of biosynthetic gene clusters in actinomycete genomes

In step with this progress in DNA sequence analysis, software has been developed for automated recognition of ORFs (Delcher et al., 2007) which yield a basic annotation, when combined with BLAST searches (Altschul et al., 1990) that identify relationships with proteins in the rapidly-expanding public databases. Increasingly, the accumulated experimental insights into biosynthetic gene clusters, especially the recognition of characteristic protein sequence motifs, have provided the basis for automated and systematic annotation of biosynthetic gene clusters, using tools such as Clustscan (Starcevic et al., 2008), clusterfinder (Cimermancic et al., 2014), and SMURF (for fungi) (Khaldi et al., 2010). The program Natural Product Domain Seeker (NaPDoS) (Ziemert et al., 2012) has proved particularly useful for analysis of fragmented genome assemblies, being based on recognition of short sequence tags. Boddy (2014) has reviewed tools specific for assembly-line polyketide synthases and nonribosomal peptide synthetases (section 1.3.3). In 2011 several research groups pooled their efforts and co-developed a versatile genome mining platform they named AntiSMASH (Medema et al., 2011), the most recent update (AntiSMASH 4.0) appearing in 2017 (Blin et al., 2017). AntiSMASH uses various algorithms to predict the locations and potential products of biosynthetic gene clusters. It is an extremely useful tool to highlight possible clusters in actinomycetes, although in practice manual curation is essential to obtain an accurate assessment of an individual cluster. Clustermine360 is a database specific for polyketide synthases and nonribosomal peptide synthetases based on AntiSMASH notation that facilitates gene cluster searching and phylogenetic analysis (Tremblay et al., 2016)

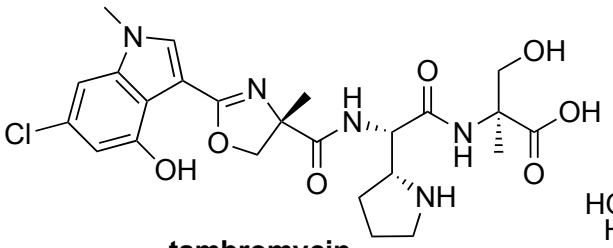
1.2.3 Genome mining of actinomycetes

Traditionally, natural products discovery has followed a paradigm that might be described as "bioactivity first", in which fermentation extracts or purified compounds were tested for bioactivity and only then were the compound and its biosynthesis studied in more detail. The rapid developments in DNA sequencing at genome scale have made possible an alternative approach. An early example of this, based on the genome sequence of *S. coelicolor*, was the prediction of the structure of the siderophore coelichelin (Challis and Ravel, 2000) although its

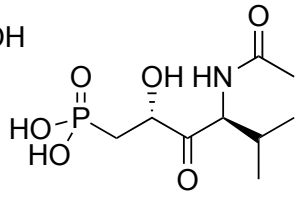
ensuing structure determination (Lautru et al., 2005) showed that the prediction was not quite right. At the same time, Marahiel and colleagues predicted and then confirmed the production of a lasso peptide (section 1.3.2) from *Burkholderia thailandensis* (Knappe et al., 2008). Since then, there have been many successful examples of structure prediction and ensuing isolation of the relevant metabolite. An extension of this idea (Ju et al., 2015) was to screen 10,000 draft-sequenced *Streptomyces* strains for the presence of a key biosynthetic gene (*pepM*) for a specific class of specialised metabolite (phosphonates); classify the gene clusters discovered into known and unknown; screen the strains predicted to produce unknowns for phosphonate products using ^{31}P NMR; and characterise the compounds detected. This led to the discovery of the novel phosphonate argolaphos (Figure 1.4) from *Streptomyces monomycini*.



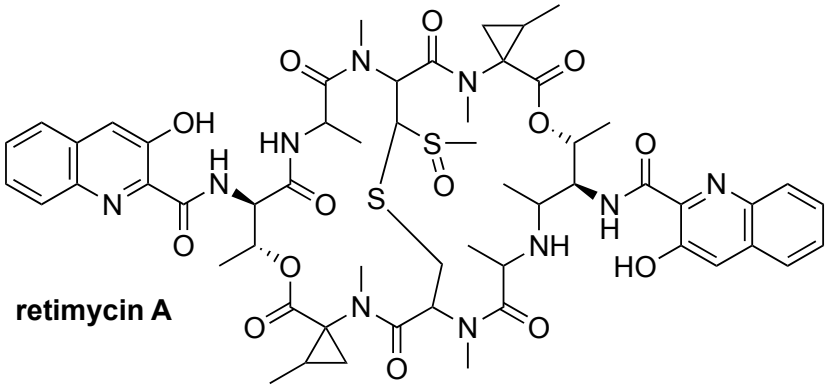
flavopeptin



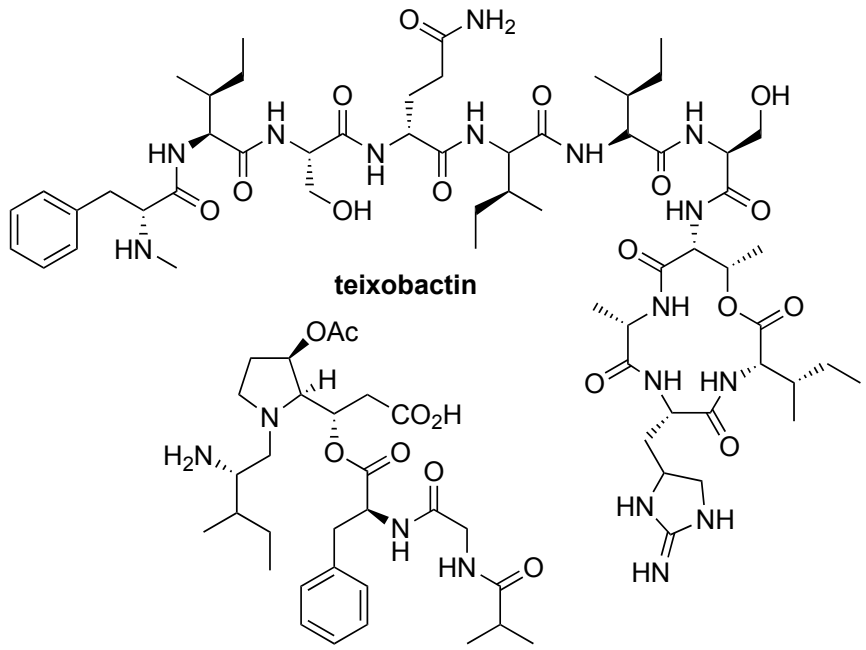
tambromycin



valinophos



retimycin A



teixobactin

rimosamide A

Figure 1.4. Structures of novel natural products characterised by genome mining

One of the first attempts to survey systematically the pattern of specialised metabolites from isolates of the same strain was the study of Müller and colleagues (Krug et al., 2008) on the metabolome of *Myxocoocus xanthus,* previously known only for a few compounds. They showed that 98 *M. xanthus* strains between them produced 37 non-ubiquitous metabolites, a strikingly high level of intraspecific diversity. In a later review, they stressed the importance of high-resolution mass spectrometry for what they called "secondary metabolomics" (Krug and Müller, 2014), linking gene clusters and pathways to metabolites. The use of untargeted MS analysis had led for example to the discovery of a novel family of hydrophilic ferrioxamines in the model organism *Streptomyces coelicolor* (Sidebottom et al., 2013). Other authors, too, had realised the value of high-resolution mass spectrometry (HRMS) and hyphenated mass spectrometry (MS-MS) as tools to establish the connection between gene cluster and specialised metabolite (Bumpus and Kelleher, 2008; Nguyen et al., 2013; Doroghazi et al., 2014). Kelleher and colleagues have coined the word metabologenomics for the large-scale MS profiling of 830 *Streptomyces* strains, coupled with draft genome sequencing, as a way to characterise orphan clusters, find new members of known structural classes of metabolite, and even find new classes (Doroghazi et al., 2014). They have demonstrated the approach

by correlating the appearance of a distinctive MS signature with the presence of a specific gene cluster, leading to the discovery of rimosamide A (McClure et al., 2016) (Figure 1.4) and of the novel chlorinated metabolite tambromycin (Goering et al., 2016) (Figure 1.4) related to the known marine-derived metabolites JBIR-34 and -35 (Muliandi et al., 2014). Jensen and colleagues have used a similar approach of pattern-based genome mining and molecular networking to link a specific gene cluster for a nonribosomal peptide synthetase to the quinone-type depsipeptide retimycin A (Duncan et al., 2015) (Figure 1.4) (Molecular networks are visual displays of the chemical space present in mass spectrometry experiments).

Several research groups have co-opted proteomics technologies to help connect biosynthetic gene clusters to their products. An advantage of a "protein first" approach is that it identifies those gene clusters that are actually expressed. For example, activity-based probes directed against the thiol groups of the 4'-phosphopantetheinyl prosthetic groups of nonribosomal peptide synthetases (NRPSs) (section 1.3.2) and polyketide synthases (PKSs) (section 1.3.3) have been used in *Bacillus subtilis* to identify these multienzymes (Meier et al., 2009). A very similar study (Bumpus et al., 2009) separated high molecular weight assembly-line multienzymes (>200 kDa) from extracts of *Bacillus* that synthesise nonribosomal peptides and polyketides, by SDS-PAGE of cell-free extracts followed by in-gel tryptic digestion and nano-LC-MS-MS, to detect phosphopantetheinylated peptides by their MS-MS signature. This approach is most powerful when the genome sequence of the bacterium is known, as in the discovery of the antiproliferative aldehyde flavopeptin (Figure 1.4) as the product of an orphan cluster in *Streptomyces flavogriseus* (Chen et al., 2013) and of the antitumor cyclic depsipeptide rakicidin D (Figure 1.4) as the product of an orphan cluster in *Streptomyces lilacinus* (Albright et al., 2014).

1.2.4 Metagenome mining

It has been estimated that only about 1% of microorganisms in a complex environment such as soil are readily culturable in standard laboratory media (Daniel, 2005). In an effort to tap the potential chemical diversity represented by these uncultured organisms, metagenomic DNA has been sequenced and analysed for the presence of biosynthetic gene clusters (Banik and Brady, 2010; Piel, 2011; Wilson and Piel, 2013). This has been used to identify new members of known classes of natural product, for example aromatic polyketides (Feng et al., 2011), indolocarbazoles and indolotryptolines (Chang and Brady, 2013), and epoxyketone inhibitors of the proteasome (Owen et al., 2015). The software tool eSNaPD, analogous to

NaPDoS, has been developed to aid the search for new scaffolds, or for new members of known classes (Reddy et al., 2014).

Metagenomic analysis has been particularly useful in marine environments, where symbionts are frequently responsible for the specialised metabolites isolated from sponges or other marine animals (Piel, 2011; Wilson and Piel, 2013; Barone et al., 2014; Wilson et al., 2014; Trindade et al., 2015; Lackner et al., 2017). Importantly, "unculturable" soil organisms have proved amenable to isolation by using special techniques (see, e.g., Kaeberlein et al., 2002) and this has led to the characterisation of the potent cell-wall directed antibiotic teixobactin (Ling et al., 2015) (Figure 1.4).

1.1 The genes and enzymes of actinomycete natural product biosynthesis

Impressive advances have been made over the last 20 years in our understanding of the enzymology of the biosynthesis pathways to the main classes of natural product found in filamentous actinomycetes, thanks to our increasing ability to link the isolation and structural characterisation of a particular natural product to the identity of the biosynthetic gene cluster that governs its production. Central to these efforts has been the use of recombinant purified enzymes expressed from pathway genes to interrogate defined substrates; as well as input from X-ray crystal structure determination of (portions of) the (multi)enzymes involved. As the biosynthetic logic of more and more pathways has been worked out, this in turn has allowed more accurate deduction of gene structure starting from the structure of the metabolite, and *vice versa*. In the sections that follow, a brief overview is given of the state of our understanding for several of the major classes of actinomycete natural product, to provide context for the work discussed in Chapter 3, probing the value of transcriptomics in guiding the genome mining of a fully-sequenced and prolific *Streptomyces* strain. In particular, an introductory account is given here of the pathways that lead to reduced complex polyketides and to nucleosides, since these pathways are the subject of the work to be described in, respectively, Chapters 4 and 5 of this thesis.

1.3.1 Terpenes and terpene synthases

Terpenes are the most abundant single class of natural products on earth, with over 65,000 different compounds known (http://dnp.chemnetbase.com/). Classical work in numerous laboratories established the origin of terpenes as composed of C5 isoprene units, and formed from two (mono-), three (sesqui-), four (diterpenes) or six (triterpenes) such units; that the biologically relevant monomers are isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), and that geranyl diphosphate (GPP), farnesyl diphosphate (FPP), geranylgeranyl diphosphate and squalene are the unique linear intermediates to the corresponding types of terpene.

Perhaps the single most striking achievement was the recognition of the *biogenetic isoprene rule* by Ruzicka's students Eschenmoser and Arigoni at ETH Zürich, which accounts, using chemical insights into the alternative folded conformations available to the linear intermediates, for the (stereo) structure of every sesqui-, di-, and triterpene characterised in the following years. This set the pattern for the idea that biosynthesis involves a common pathway for chain assembly, followed by extensive possibilities for diversification. Terpenes have traditionally been considered products of plants and fungi (Dewick, 2002; Dewick, 2009)

but in recent years it has become clear that bacteria, and specifically filamentous actinomycetes, are rich sources of both known and novel terpenes. Arigoni and his colleagues (Eisenreich et al., 1998) were also co-discoverers, with Rohmer (Rohmer, 1999) of the alternative 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway to IPP, in both microorganisms and higher plants. Most (but not all (Dairi, 2005)) *Streptomyces* and allied genera exclusively use the MEP pathway rather than the mevalonate pathway.

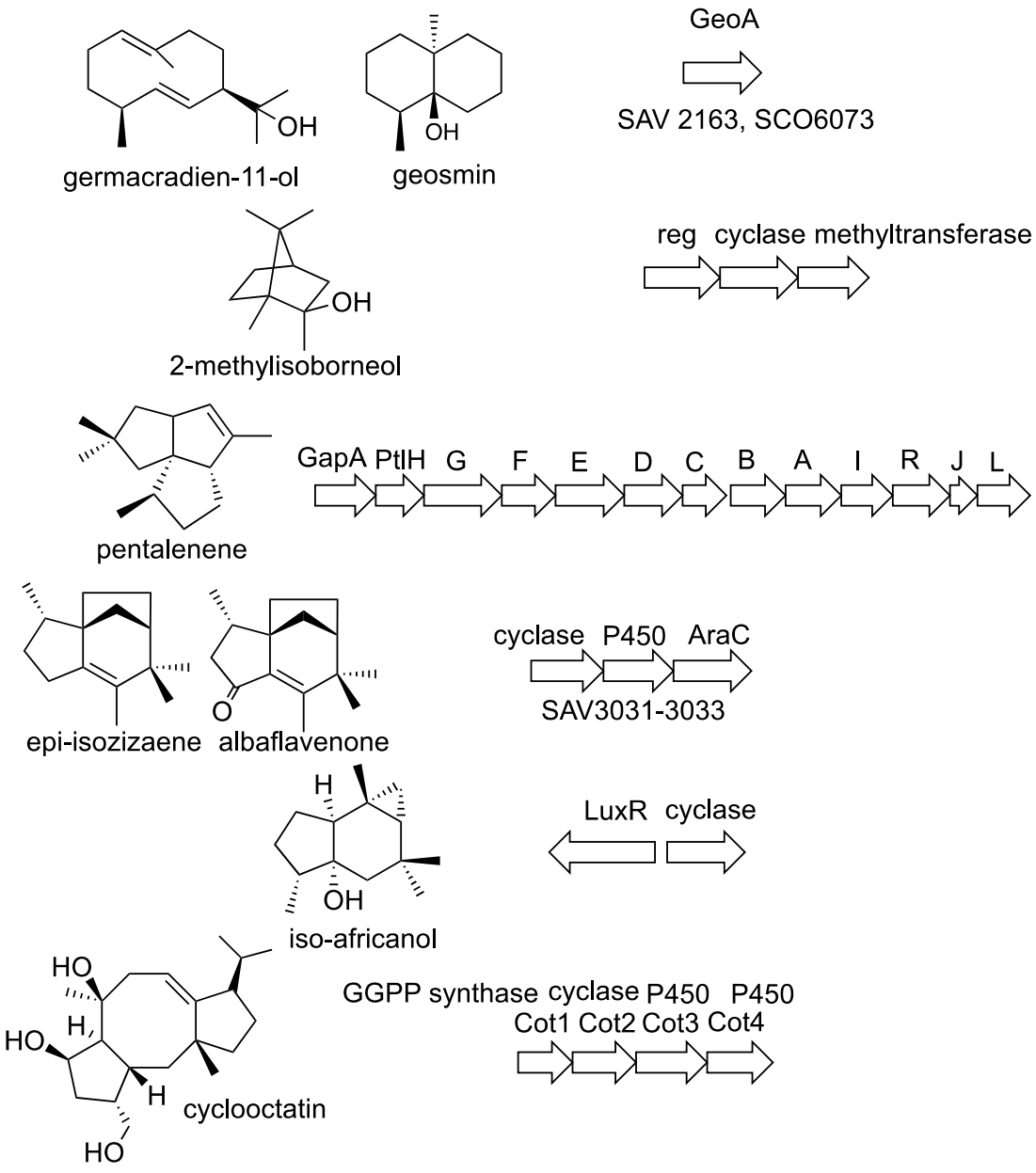


Figure 1.5. Biosynthetic gene clusters for typical *Streptomyces* terpenes.

After a prenyltransferase enzyme has added the appropriate number of IPP monomers to DMAPP to give GPP, FPP or GGPP, the pyrophosphate group is released to give an allyl cation, leading to rearrangement via a series of cationic intermediates, that is terminated either by proton loss or by quenching with water to give an alcohol. The bacterial enzymes have the same conserved active site, containing sequence motifs as in previously-characterised plant GPP, FPP, GGPP synthases and terpene cyclases (Dickschat, 2016), as first shown for the sesquiterpene cyclase for pentalenene synthase from *Streptomyces exfoliatus* (Lesburg et al., 1997). A similar enzyme is involved in neopentalenolactone biosynthesis in *Streptomyces avermitilis* (Cane and Ikeda, 2012) (Figure 1.5). The gene cluster for this product (Figure 1.5) contains 13 ORFs: *pltB* encodes the FPP synthase, *pltA* encodes the terpene cyclase, *ptlR* is a regulatory gene, and *gapA* is an altered glyceraldehyde dehydrogenase gene conferring self-resistance, and pltG encodes a membrane-associated exporter. The other eight genes all encode redox enzymes for the extensive oxidative tailoring to give neopentalenolactone.

Geosmin (Figure 1.5) is the terpene most commonly produced by *Streptomyces*. In contrast to neopentalenolactone, its biosynthetic pathway consists of a single distinctive bifunctional enzyme (GeoA in *Streptomyces coelicolor*). Presumably FPP is furnished by a synthase encoded elsewhere in the genome. The *N*-terminal domain of GeoA houses the terpene cyclase that forms germacradien-11-ol, and the C-terminal domain catalyses the fragmentation of this intermediate to geosmin (DIckschat, 2016). Likewise, biosynthesis of another widespread sesquiterpene in *Streptomyces*, *epi*-isozizaene (Figure 1.5), requires only a single terpene cyclase (encoded by SCO5222 in *S. coelicolor*). The gene is apparently coupled to the downstream SCO5223, encoding a cytochrome P450, and it has been shown, for the recombinant enzyme cloned from the cryptic cluster in *S. avermitilis,* that it catalyses the 2-step allylic oxidation of *epi*-isozizaene to albaflavenone (Figure 1.5) (Cane and Ikeda, 2012). Inspection of the published sequences shows that the neighbouring ORF in *S. coelicolor* SCO5224, which is predicted to encode an AraC-like regulatory gene (Figure 1.5), is missing in *S. avermitilis*, but its possible role in albaflavenone biosynthesis has not been investigated. As for geosmin synthase, the small, distinctive gene cluster for biosynthesis of 2-methylisoborneol is widespread in *Streptomyces* (Figure 1.5) (Komatsu et al., 2008). It consists of three clustered ORFs: a regulatory gene, and genes encoding the terpene cyclase, and an S-adenosylmethionine (SAM)-dependent methyltransferase. Again, the biochemistry has been confirmed using purified recombinant enzymes cloned and expressed from *S. coelicolor* (Cane and Ikeda, 2012), but the native regulation remains obscure.

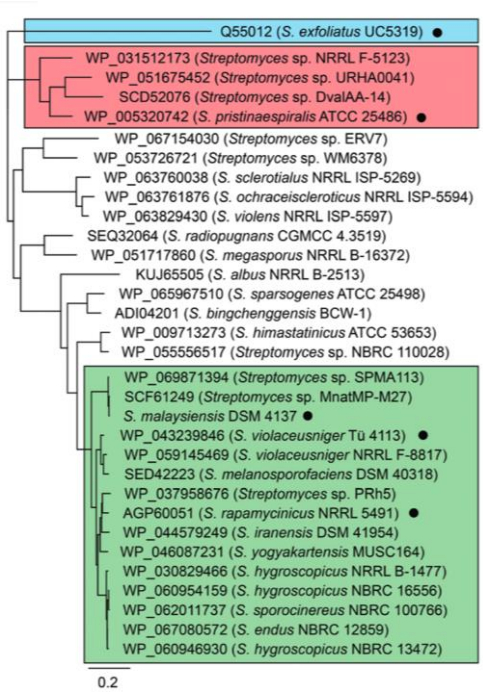


Figure 1.6 Phylogenetic tree of bacterial terpene cylcases.

In contrast, individual actinomycete genomes frequently house 'orphan' genes encoding putative terpene cyclases, whose amino acid sequence differs sufficiently from known enzymes that it does not immediately identify their likely function (Figure 1.6) (Rabe et al., 2017). The gene cluster for the unusual sesquiterpene iso-africanol (Figure 1.5) again contains only two adjacent ORFs, one encoding a typical terpene cyclase originally mis-annotated as a pentalenene synthase, and the other a LuxR-family regulator. Iso-africanol has been detected as being produced from several *Streptomyces* spp. (Riclea et al., 2014), and the pathway has been biochemically confirmed *in vitro* using the cloned and expressed gene from *Streptomyces malaysiensis* DSM4137 (Rabe et al., 2017) (see also Chapter 3 of this thesis).

Diterpene synthases acting on GGPP are rarer in *Streptomyces*, but several have been discovered and characterised, including the synthase in the pathway to the antibiotics platensimycin and platencin in *Streptomyces platensis* (Smanski et al., 2012) and the cyclooctatin synthase (Figure 1.5) in *Streptomyces melanosporofaciens* (Lim et al., 2008). The latter pathway has also been found in several other *Streptomyces* including *Streptomyces* sp. LZ35 (Zhao et al., 2013), as well as in *S. malaysiensis* DSM4137 (see Chapter 3 of this thesis). The *cot* gene cluster consists of four adjacent, probably transcriptionally-linked, ORFs encoding respectively the GGPP synthase, the terpene cyclase, and two cytochrome P450 enzymes (Figure 1.5), but there is no adjacent regulatory gene.

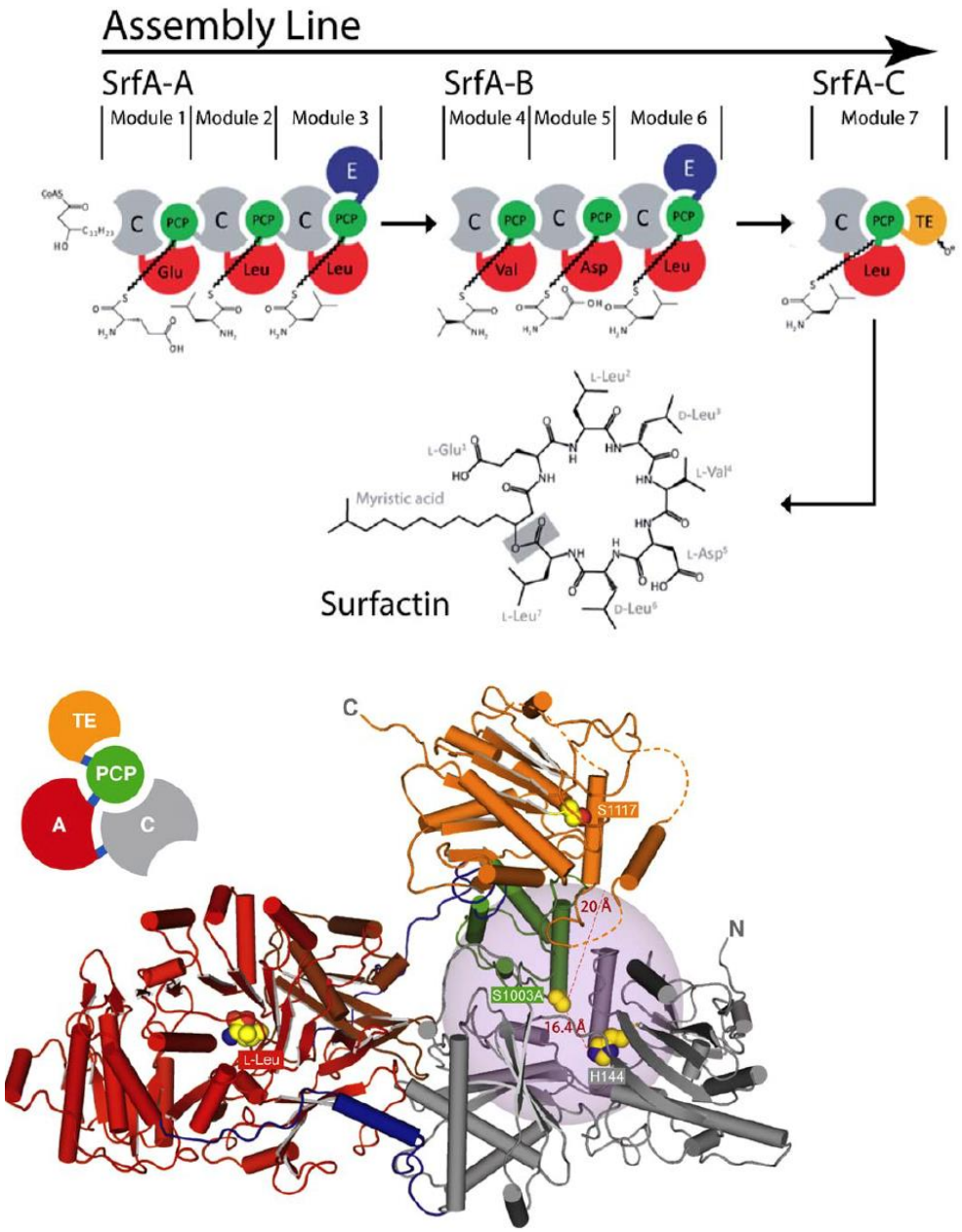


Figure 1.7. A. The nonribosomal peptide synthetase (NRPS) assembly-line for the lipoheptapeptide surfactin from *Bacillus subtilis.* (adapted from Marahiel, 2016) Three multienzymes Srf-A, B and C form the assembly-line. Condensation (C) domains are shown in grey; adenylation (A) domains in red; peptidyl carrier protein (PCP, or T) domains in green; epimerase (E) domains in blue; and the thioester/cyclase (TE) domain in orange. The starter unit is furnished by myristyl-CoA. B. The X-ray crystal structure of the intact final module of the surfactin NRPS (adapted from Strieker et al., 2010).

1.3.2 Peptides and peptide synthetases

Peptides are among the most diverse and valuable of bacterial specialised metabolites. It was recognised as early as 1954 by Lipmann that the production of some microbial peptides is unaffected by inhibitors of ribosomal protein synthesis, and this led Lipmann to his multienzyme thiotemplate model (reviewed by Kleinkauf and Koischwitz, 1980) in which intermediates are elongated on large multienzymes to which they remain covalently bound in thioester linkage. Fungal and bacterial nonribosomal peptide synthetases were successfully purified and used to study the biosynthesis of, for example, the tetrapeptide tyrocidine and the cyclic decapeptide gramicidin S from *Bacillus brevis*, the cyclohexadepsipeptide enniatin from *Fusarium oxysporum*, and the cyclic undecapaptide cyclosporin from *Tolypocladium inflatum*. In these systems, amino acids were shown to be activated as aminoacyl adenylates, before being transferred to thioester linkage on the enzyme. However, until 1990 it was believed that all intermediates were bound to a single "central thiol" (Kleinkauf and von Döhren, 1987).

Once the modular nature of complex polyketide synthases had been revealed (see below), rapid progress was made, via gene cloning and sequencing, in defining a similarly modular arrangement for nonribosomal peptide synthetases (Guttiérez et al., 1991; Tonin et al., 1991; Mootz and Marahiel, 2007). It is now established that in general, there is a different extension module for each amino acid introduced, arrayed within the multienzyme in theorder in which the modules are used. A canonical extension module consists minimally of a condensation (C) domain, an activating adenylation (A) domain, and a peptidyl carrier protein (PCP) also known as the thiolation (T) domain. The A domain activates the amino acid monomer as an aminoacyl adenylate and transfers the acyl group to the thiol of the 4'-phosphopantetheine prosthetic group on the PCP domain. The PCP domain delivers the acyl group to the C domain where it accepts the peptidyl group from the previous module to form the amide link. The PCP then delivers the elongated peptide to the next module. An epimerisation (E) domain is (in general) present if a D-amino acid is to be introduced, and a methylation (M) domain may be included to specify *N*-methylation of the peptide bond, and in such modules the PCP-bound peptide is tailored in those active sites before transfer to the next module. A thioesterase domain (TE) is typically present at the C-terminus of the final extension module, to catalyse product release via hydrolysis or cyclisation.

As an example, Figure 1.7 shows the organisation of the assembly-line multienzyme for the cyclic peptide surfactin from *Bacillus subtilis* (Marahiel, 2016). This model at once explains the selectivity of the synthetase, because each successive amino acid is selected

and activated by a specific A domain. A set of contact residues lining the specificity pocket of the conserved A domains are the primary determinant of the specificity (Conti et al., 1997), and it has proved possible, from knowledge of the A domain amino acid sequence, to predict the likely nature of the amino acid that is inserted (Stachelhaus et al., 1999; Challis et al., 2000). This property has proved widely useful in establishing the identity of orphan NRPS genes discovered during genome sequencing, and in encouraging attempts to engineer NRPS assembly-lines to produce novel products (Winn et al., 2016). However, correct prediction remains difficult in the case of those many A domains that recruit unusual, non-proteinogenic amino acids.

The three-dimensional structure of an intact extension module, the final module of the surfactin synthetase, has been determined by X-ray crystallography Figure 1.7) (Tanović et al., 2008). Interestingly, this structure, as well as other structures determined for individual domains and didomains, indicates that there must be significant conformational change during the elongation cycle for the PCP-bound species to reach the various active sites with which it must interact namely, the A and C domains within its own module, the downstream C domain, and tailoring domains if they are present (Weissman, 2015; Kittilä et al., 2017).

In contrast to the products of assembly-line nonribosomal peptide synthetases (NRPSs), other bacterial peptides are produced on ribosomes as pre-peptides with a characteristic N-terminal leader peptide, and undergo significant post-translational processing to form distinctive crosslinks and topological structures. They are now described collectively as *Ribosomally synthesized and Post-translationally modified Peptides* (RiPPs), and they display a wide variety of biological activities. The best known and understood are the so-called lanthipeptides, in which the characteristic lanthionine and methyllanthionine residues are formed by dehydration of serine and threonine residues to dehydroalanine (Dha) and dehydrobutyrine (Dhb) respectively, and the formation of thioether bridges to cysteine residues through 1,4 addition onto the dehydro amino acids.

The first of these compounds to be studied in depth was the antimicrobial lanthipeptide (lantibiotic) nisin from *Lactococcus lactis* (Sahl, 1994), but with the rapid increase of genomic sequence information over the last two decades hundreds of lanthipeptide biosynthetic gene clusters have since been identified in (particularly) actinobacterial genomes. They have been classified according to the details of the enzymology that is involved in the conversion of the pre-peptide into the crosslinked product (Van der Donk and Nair, 2014; Repka et al., 2017). Class I lanthipeptides utilise a dehydratase enzyme (LanB) and a coupled cyclase (LanC), and the dehydration is facilitated by intermediate ester formation to glutamate catalysed by a glutamyl-tRNA-dependent enzyme (Repka et al., 2017; Funk and van der Donk, 2017). The

leader peptide (20-30 amino acids) is important for recognition by LanB and LanC. Class II lanthipeptides like cinnamycin (O'Rourke et al., 2017) are formed by the action of a single dehydratase-cyclase enzyme LanM. The lanthipeptide peptide SapB involved in *Streptomyces* aerial mycelium formation and the novel cross-linked labyrinthopeptins (Müller et al., 2010) are formed by a different, so-called Class III or LanKC enzyme. Both class II and class III enzymes activate the hydroxy amino acid for the dehydration step by phosphorylation (Repka et al., 2017). Other RiPPs such as the linaridins (Claesen and Bibb, 2010) do not have lanthionine bridges, although they contain dehydro-amino acid residues.

A second, emerging type of RiPP that appears to be encoded in many actionomycete bacteria comprises so-called lasso peptides (Ortega and van der Donk, 2016). These are characterised by a macrolactam ring formed by the N-terminal amino group forming an amide with an internal Asp or Glu residue, and a C-terminal tail that is threaded through the ring and held in place by bulky sidechain residues to form a lariat structure. Their biosynthesis involves a pre-protein that is acted upon by an ATP-dependent proteinase that unmasks the N-amino group; and a second enzyme that activates the internal acidic sidechain and catalyses lactam formation. Bio-informatic tools have been crucial to the growing appreciation of the widespread occurrence of these gene clusters in actinomycete genomes (Maksimov and Link, 2014; Tietz et al., 2017).

1.3.3 Polyketides and polyketide synthases

The polyketides are one of the largest and most diverse classes of specialised metabolites, and this is reflected in the diverse bioactivities that they show. Numerous actinomycete-derived polyketides have proved useful leads in drug discovery, while others have themselves been developed into useful drugs and biocontrol compounds, such as the antibacterials tetracycline and erythromycin A; the immunosuppressant rapamycin; the antifungal compound amphotericin B; the insecticidal spinosyn A; the anthelminthic avermectin; and the anticancer compounds doxorubicin and mithramycin

The polyketide pathway strongly resembles the process of fatty acid biosynthesis, in which chain assembly is initiated with acetyl-CoA and successive extender units, supplied as malonyl-CoA, are added onto the growing chain in an enzyme-catalysed Claisen-like condensation reaction leading initially to a 3-ketoacyl thioester. In a typical fatty acid synthase (FAS) synthesising a long-chain saturated fatty acid, each condensation step is followed by reduction of the newly-introduced 3-keto group by an NADPH-dependent ketoreductase (KR),

dehydration by a dehydratase (DH) and further reduction by an NADPH-dependent enoylreductase (ER), producing a fully saturated acyl chain which enters the next extension cycle (Weissman, 2009, 2014). Once the desired chain length is reached, a thioesterase (TE) enzyme releases the product as either a free acid or as a fatty acyl thioester for further metabolism. Polyketide synthases (PKSs) utilise the same basic mechanism to construct the polyketide chain but in contrast the introduced ketide unit may be only partially reduced, or not at all, before the next cycle of chain extension. Also, a wider range of starter and extender units is used. Chain release can also occur by multiple different mechanisms, often including cyclisation. These factors create a far greater chemical and stereochemical diversity in polyketides than in fatty acids. For PKSs catalysing little or no reduction, the polyketide chain is usually stabilised by formation of aromatic groups (Hertweck 2009). As with FAS enzymes, the evolution of PKSs has led to a variety of different quaternary structures for these enzymes, which forms the basis of their classification into type III, type II and type I systems, which will be discussed here in that order.

Type III enzymes are simple, homodimeric ketosynthase enzymes that almost always utilise acyl-CoA or malonyl-CoA esters as starter and (usually) malonyl-CoA as extender units and carry out iterative chain extensions using a single active site. They produce a wide variety of phenolic compounds (chalcones, stilbenes, flavonoids, resorcylic acids) that perform vital roles in plants, including protection against UV, flower colour, and chemical defence (Yu et al., 2012; Stewart et al., 2013). Many of these compounds are now recognised as beneficial to human health, such as the phenylpropanoid cucurmin (Katsuyama et al., 2009) or the anti-oxidant resveratrol. Originally thought to be confined to plant secondary metabolism, these enzymes were later reported to be involved in the biosynthesis of bacterial compounds as well. These include 1,3,6,8-tetrahydroxynaphthalene (THN) from *Streptomyces griseus* and its spontaneous oxidation product flaviolin (Funa et al., 1999), also found in other actinomycetes (Cortés et al., 2002); 3,5-dihydroxyphenylacetyl-CoA, the source of the 3,5-dihydrophenylglycine building block for biosynthesis of glycopeptide antibiotics in *Amycolatopsis* spp. (Li et al., 2001); the unusual benzenoid starter unit for the polyketide kendomycin in *Streptomyces* spp. (Wenzel et al., 2008); and the 3,5-dihydroxybenzoic starter unit for the polyketide venemycin in *Streptomyces venezuelae* (Thanapipatsiri et al., 2016). These examples are useful guides in interpreting the potential role of genes for orphan type III enzymes found in actinomycete genome sequences.

Type II PKSs are defined as consisting of discrete proteins each bearing one, or at most two, of the individual enzyme activities needed to synthesise a polyketide chain. It is possible that they form a multienzyme complex within the cell. They are widespread in actinomycete bacteria, notably for the production of spore pigment but also responsible for a

great variety of aromatic polyketides arising from the folding, cyclization and tailoring of essentially unreduced polyketide chains based almost exclusively on malonate units (Hertweck et al., 2007; Zhang, 2009). Products of such type II PKSs include tetracyclines and the anticancer compound doxorubicin. In these enzymes chain initiation, elongation and termination takes place on a heterodimeric KSα:KSβ PKS, where KSα is the catalytically competent synthase and KSβ modulates initiation and chain length. The growing chains are tethered on a discrete acyl carrier protein (ACP) via thioester linkage to a 4'-phospantetheinyl prosthetic group, and loading of extender units onto the ACP is believed to require the FAS malonyl-CoA:ACP acyltransferase. Ketoreductases [KR], cyclases [CYC], and aromatases [ARO] direct chain folding and further remodelling and decoration of the polyphenolic chain occurs from oxygenases, glycosyltransferases and methyltransferases (Hertweck et al., 2007). Despite intensive research on their structure and function, it remains challenging to predict the nature of the product of orphan type II PKS genes (Zhang et al., 2017).

Type I PKSs, in contrast to type II PKSs, house their constituent enzyme activities within one or a few multifunctional protein subunits, with an integrated ACP domain and multiple enzyme domains. In actinomycete bacteria, many form a dimer of identical subunits, with the same order of active sites along the polypeptide chain (KS, AT, DH, ER, KR, ACP, TE) as in animal FAS, and sometimes including an integrated methyltransferase domain. The active sites are used iteratively until the appropriate number of chain extensions have occurred. The archetype of this type of PKS is the well-studied 6-methylsalicyclic acid synthase (MSAS) of the fungus *Penicillium patulum* (Spencer and Jordan, 1992). In *Streptomyces* and related bacteria, MSAS-like PKSs are responsible for the 6-MSA moiety of the antibiotic chlorothricin (Jia et al., 2006), for the orsellinic acid moiety in both the potent anticancer enediyne calicheamycin (Ahlert et al., 2002) and the antibiotic polyketomycin (Daum et al., 2009), and in formation of the naphthoate moiety of the antitumor compounds neocarzinostatin (Liu et al., 2005) and azinomycin (Zhao et al., 2008).

1.3.4 Modular polyketide synthases and the assembly-line paradigm

In 1990-1991, sequences were published, independently by researchers at Abbott Laboratories (Donadio et al., 1991) and here in Cambridge (Cortés et al., 1990), of the genes encoding the PKS from *Saccharopolyspora erythraea* for biosynthesis of 6-deoxyerythronolide (6-DEB), the aglycone of the macrolide antibiotic erythromycin A. Surprisingly, this revealed a wholly different arrangement to that previously established for other PKSs. Instead of a single

set of enzymes acting repeatedly to build the polyketide chain, a different set (or 'module') of enzymes acts in each cycle of chain extension. The modules are arrayed within Type I giant PKS multienzymes in the order in which they are used. This novel assembly-line paradigm for enzymatic catalysis immediately accounted for the observed selectivity of polyketide chain extension in reduced complex polyketides, and rationalised the biosynthetic relationships between hundreds of known polyketides.

The direct correspondence between the (order, number and content of) modules and the chemical nature of the product, proved to be widely applicable to a large number of PKS systems (Hertweck, 2009). This ‘colinearity rule’ allows confident predictions of metabolite structure given the sequence of the genes (Weissman and Müller, 2009; Weissman, 2009) giving a significant impulse to genome mining efforts. Co-linearity seems to have evolved as a consequence of continuous evolutionary pressure for novel polyketides, and of new polyketide pathways being formed by horizontal transfer and recombination of PKS-encoding DNA (Callahan et al., 2009).

The 6-deoxyerythronolide synthase (DEBS) assembly-line (Figure 1.10) comprises three homodimeric PKS subunits (DEBS1, DEBS2 and DEBS3) (Caffrey et al., 1992) each housing two extension modules, and containing in total 28 domains. The precursors consist of one propionyl-CoA starter unit and six (2*S*)-methylmalonyl-Coenzyme A (CoA) units (Marsden et al., 1994). A loading module at the N-terminus of DEBS1 consisting of a propionyl-CoA-selective acyltransferase domain and an ACP domain recruits the starter unit, and a thioesterase/cyclase domain at the C-terminus of DEBS3 offloads and cyclises the full-length heptaketide to give 6-DEB.

Further work on cloning and sequencing of the PKS for the immunosuppressant rapamycin from *Streptomyces hygroscopicus* (since renamed *Streptomyces rapamycinicus*) showed that after 14 cycles of polyketide chain extension, an amino acid (L-pipecolic acid or L-proline) is incorporated via a hybrid PKS-NRPS assembly-line, in which the final NRPS module (C-A-T-C domains) activates and incorporates the amino acid, and then the C-terminal C domain catalyses the macrocyclization and release of the product (Schwecke et al., 1995).

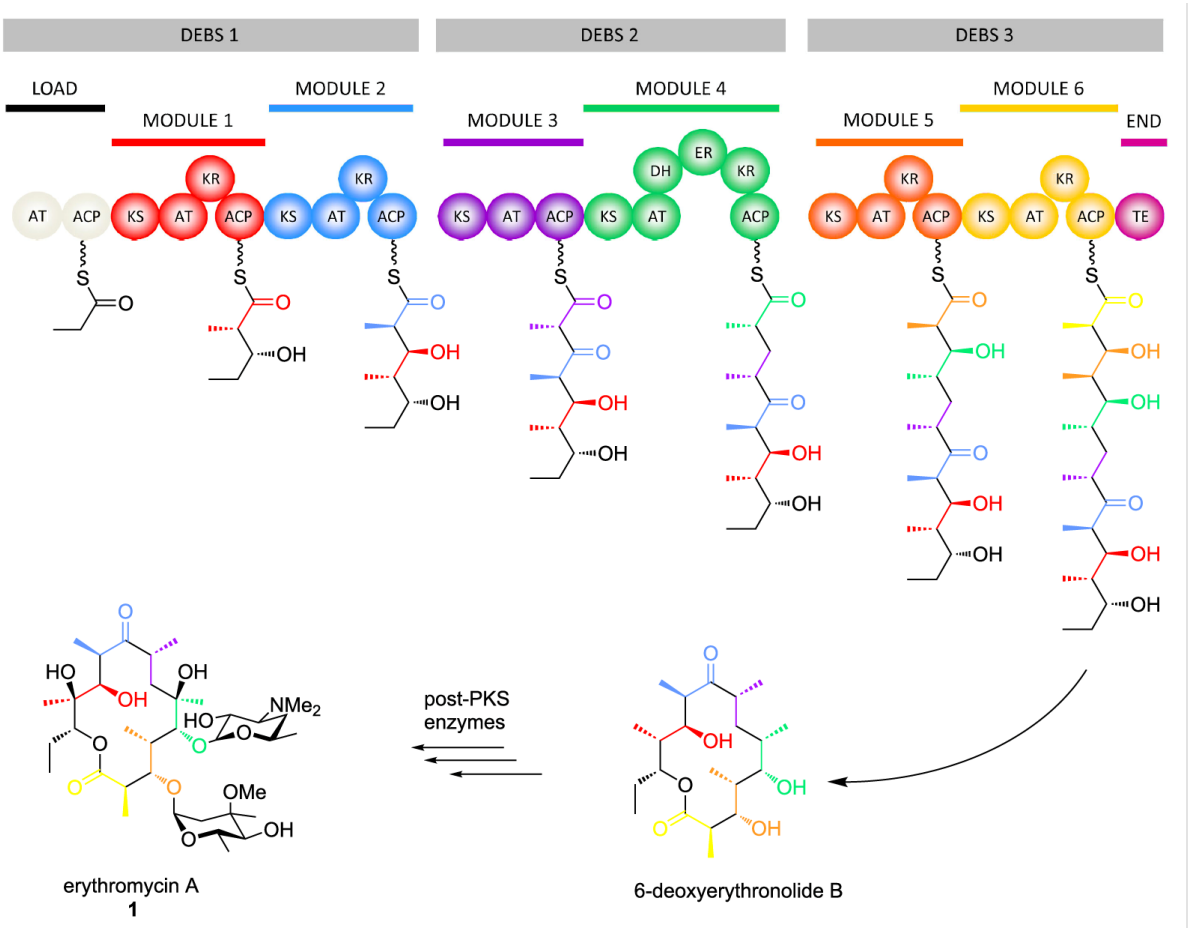


Figure 1.10. The erythromycin PKS assembly-line. (adapted from Weissman, 2017)

Not all assembly-line PKSs follow the canonical pattern of the erythromycin PKS. In particular, numerous so-called *trans*-AT or AT-less type I modular PKSs have extension modules lacking an AT domain, and instead all the extension units are provided *in trans* by a stand-alone AT domain which binds non-covalently at multiple sites on the PKS. The same AT is therefore used iteratively to acylate each module and each module, with rare exceptions (kirromycin reference) therefore recruits the same (malonate) unit. The first *trans*-AT PKS to be noted in the genome sequence of *Bacillus subtilis* 168 in 1993 and an almost identical sequence was linked later to production of the hybrid peptide-polyketide bacillaene (Chen et al., 2006). Meanwhile, *trans*-AT systems were discovered and their significance grasped in the pederin PKS from an unidentified bacterium in the gut of *Pederus* beetles (Piel, 2002) and in the PKS for the anticancer compound leinamycin from *Streptomyces atroolivaceus* (Cheng et al., 2003). Since then, large numbers of *trans*-AT PKS have been described (reviewed in Cheng et al., 2009; Helfrich and Piel, 2016). Although relatively uncommon in actinomycetes,

they have been characterised from a wide range of bacteria, especially *Burkholderia*, *Pseudomonas*, cyanobacteria, and bacterial symbionts of sponges (Helfrich and Piel, 2016).

These systems appear to have evolved separately from canonical *cis*-AT assembly-line PKSs (Nguyen et al., 2009; Lohman et al., 2015). They are also very frequently characterised by irregular placement of domains, the inclusion of apparently inactive domains, and incorporation of a wider range of enzyme types. Methyl branches, for example, arise not from use of methylmalonyl-CoA to supply extension units, but from *in situ* methylation of intermediates by integrated methyltransferase domains. A few PKS systems, such as those for etnangien (Menche et al., 2008), kirromycin (Musiol et al., 2013) and enacyloxin (Mahenthiralingam et al., 2011) have modules of both *cis*- and *trans*-AT types, or can incorporate other extender units other than malonate (Zhao et al., 2010; Musiol et al., 2013). Also, the "proof-reading" function in *trans*-ATs differs from that in *cis*-AT PKSs (Heathcote et al., 2001) and in NRPSs (Schwarzer et al., 2002), where discrete thioesterase enzymes (TEIIs) hydrolyse stalled intermediates from the multienzymes. Instead, as shown for the pederin PKS, a second stand-alone AT enzyme performs this role (Jensen et al., 2012). Figure 1.11 shows, as an example, the assembly-line for the *trans*-AT PKS for virginiamycin (Weissman, 2017).

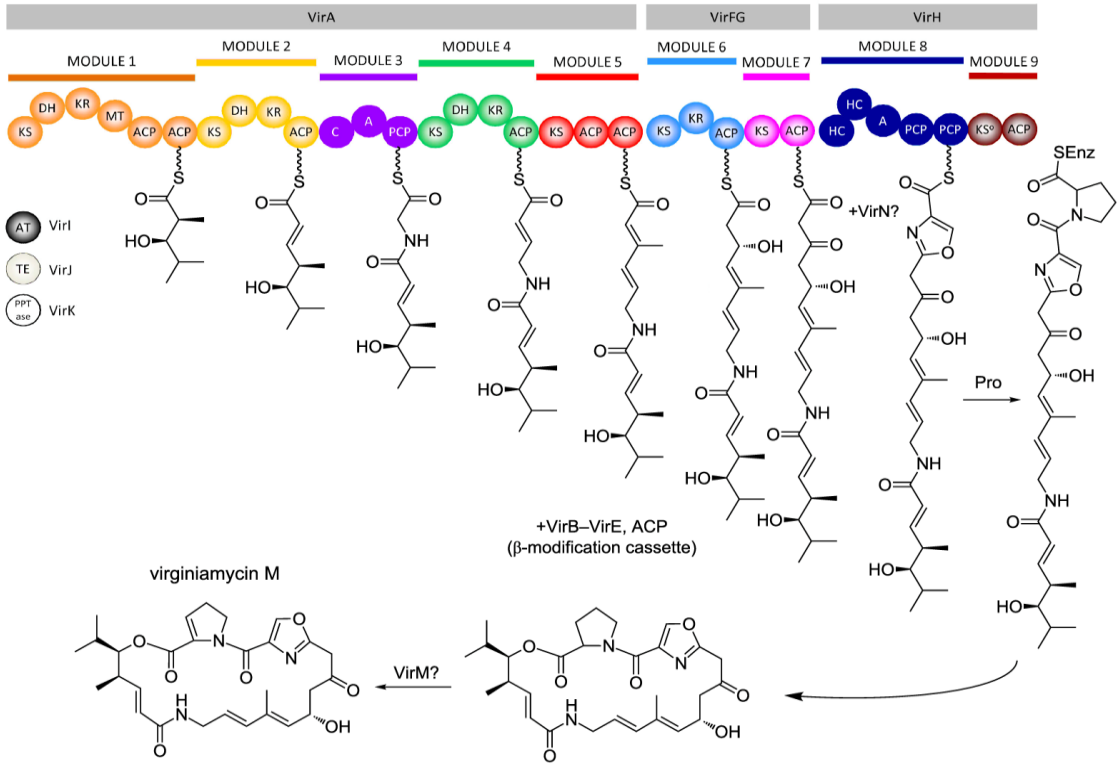


FIgure 1.11 The assembly-line of the virginiamycin *trans*-AT PKS. (adapted from Weissman, 2017).

1.3.5 Polyketide synthase (stereo)selectivity

The sequence of the DEBS PKS did not reveal the basis for extender unit choice, because all extension modules recruit (2*S*)-methylmalonyl-CoA (Marsden et al., 1994; Weissman, 2017). However, the sequencing of the rapamycin PKS allowed sequence motifs to be identified that correlated with the utilization by specific AT domains of either malonyl- or methylmalonyl-CoA (Haydock et al., 1995a). As the database of AT domains grew, it was possible to identify the particularly informative fingerprint residues HAF**H** (methylmalonyl) or YAS**H** (malonyl) around one of the essential active site His residues, about 100 residues C-terminal of the active site Ser residue (Reeves et al., 2001; Del Vecchio et al., 2003). Variants of this motif could later be correlated to the introduction of less frequently used extender units, for example from methoxymalonyl-CoA (Haydock et al., 2005b). The crystal structure of a KS-AT didomain from DEBS (Tang et al., 2006) has confirmed that the structural basis for this AT selectivity, and for the discrimination against the (2*R*)-isomer of methylmalonyl-CoA, lies in predicted steric clashes with the amino acid residues for the YASH/HAFH motif (Weissman, 2017).

Similarly, analysis of KR sequences led to the recognition (Caffrey, 2003; Reid et al., 2003) that certain residues could be correlated to the stereochemistry of β-ketoreduction in the growing polyketide chain. These observations have been confirmed and extended by X-ray crystal structure analysis of KR domains, and codified into a scheme for predicting the stereochemical outcome at both β- and (where appropriate) α-branched positions (Keatinge-Clay, 2007). The same is true of DH domains where an initial identification of active site residues (Bevitt et al., 1993) has also been confirmed by crystallographic analysis (Keatinge-Clay, 2008): a conserved His residue acts as a general base to deprotonate at C-2, and a conserved Asp residue acts as a general acid to stabilize the C-3 hydroxy leaving group. A conserved tyrosine in a YPG motif, which is proposed to form a hydrogen bond to the hydroxy group, has very recently been shown also to be essential for activity (Zhang et al., 2017). Finally, a correlation has been found in the ER active site of modular PKSs which correlates with the (2*R*)- or (2*S*)- configuration in C-2 branched, fully-reduced ketide units (Kwan et al., 2008), but as yet no crystal structure has been obtained for this domain. Taken together, these bio-informatic, structural and biochemical insights have, first, provided the essential basis for attempts at rational engineering of modular PKS assembly-lines to create novel products; and secondly, they have provided the essential information to train automated gene cluster prediction software for use in genome mining.

1.3.6. Nucleosides and their biosynthetic pathways

Nucleoside natural products are produced widely (several hundred are known) and almost exclusively by *Streptomyces* and related actinomycetes. As a class, they show a remarkably broad range of bioactivity, including antibacterial, antifungal, antitumour, insecticidal, antiviral and immunomodulatory compounds (Niu and Tan, 2015; Chen et al., 2017). Antibacterial nucleosides like the uridyl peptide antibiotic pacidamycin (Figure 1.12) and the uridyl lipopeptide tunicamycin (Figure 1.12) have been shown to inhibit a critical step in peptidoglycan synthesis, catalysed by phospho-*N*-acetylmuramyl-pentapeptide translocase (translocase I, MraY) (Winn et al., 2010), presumably because they mimic the normal substrate.

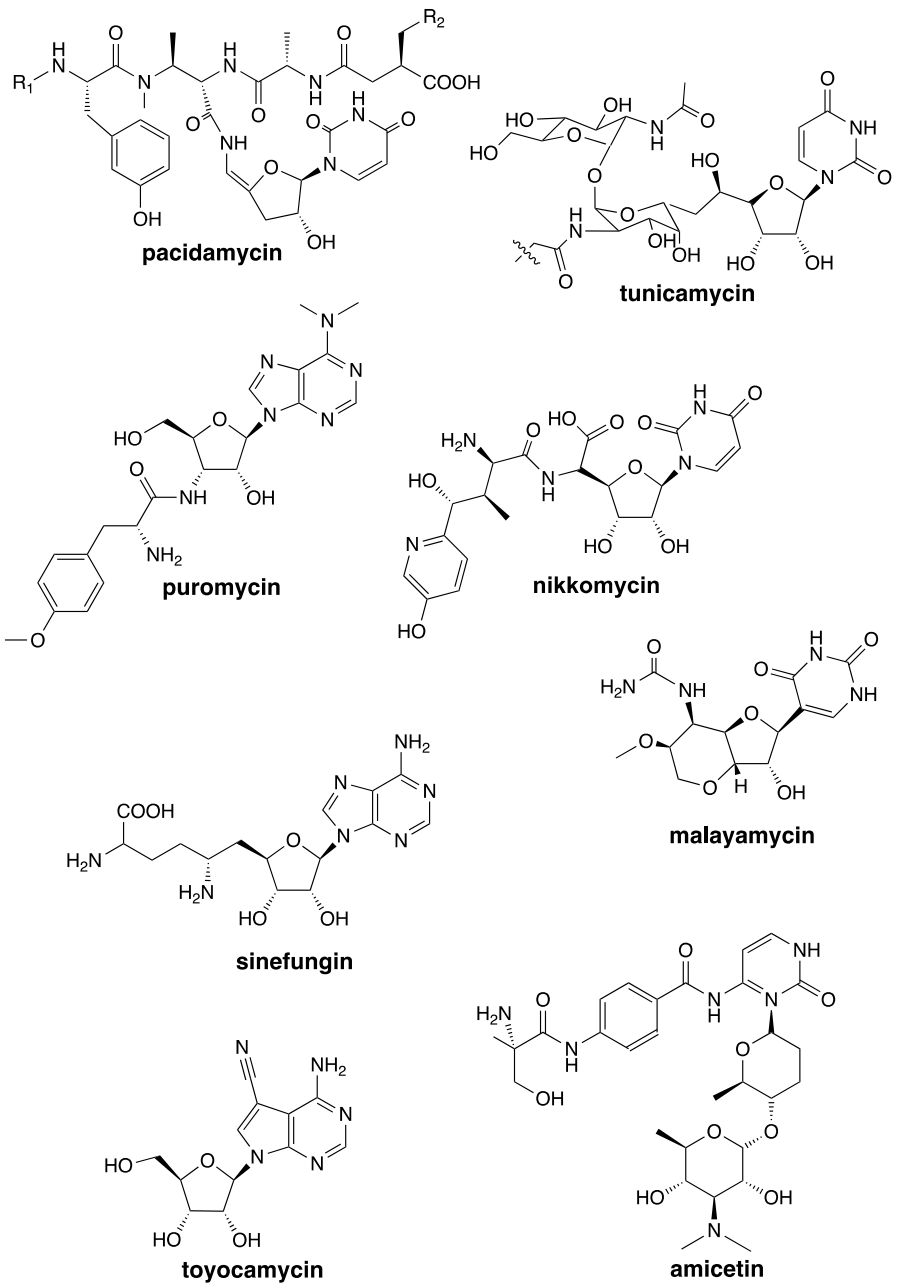


Figure 1.12. Nucleoside antibiotics from *Streptomyces* spp. Tunicamycin and pacidamycin are inhibitors of MraY in cell wall biosynthesis; puromycin and amicetin are protein synthesis inhibitors; nikkomycin and malayamycin are inhibitors of chitin synthase; sinefungin inhibits SAM-dependent methyltransferase; and toyocamycin inhibits RNA processing. In the pacidamycin structure, different natural pacidamycins 1-3 have R_1= Ala, and R_2 = indolyl, phenyl or 4-hydroxyphenyl respectively. Tetrapeptide pacidamycins have R_1= H.

The pacidamycin pathway has been independently studied by the research groups of Rebecca Goss and Chris Walsh. Its biosynthesis differs from that of a conventional ribosomal or NRPS product, and it therefore provides a good illustration of the genome-based strategy

(and the challenges) of uncovering and characterising novel enzymology in the pathways to *Streptomyces* specialised metabolites. The pacidamycins are a set of pentapeptides in which, for the major species, the N-terminal residue is L-Ala and the C-terminal residue is either Trp, Phe or the non-proteinogenic amino acid *meta*-tyrosine (Figure 1.13).

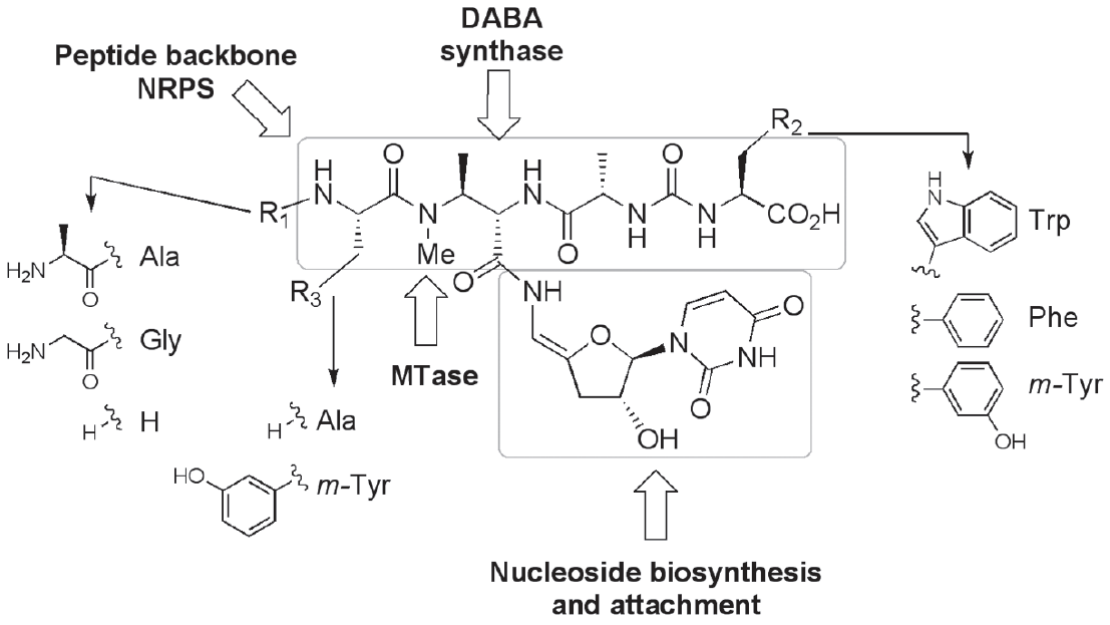


Figure 1.13. Retrobiosynthetic analysis of the pathway to pentapeptide pacidamycins. (adapted from Rackham et al., 2011a)

The presence of the non-proteinogenic amino acid 2,3-diaminobutanoic acid (DABA) in the peptide suggests a possible origin via an NRPS-catalysed pathway, but the structure also reveals an unusual ureido linkage in the peptide chain, as well as the presence of a highly unusual 3'-deoxyenamino-uridine nucleoside moiety. In attempting to clone the relevant gene cluster from *Streptomyces coeruleorubidus*, consideration of how the molecule might plausibly be assembled (Fig. 1.13) led Goss and colleagues initially to design hybridization probes based on known diaminopropionate and diaminobutyrate synthase genes (reviewed by Rackham et al., 2011). When this approach failed, they turned to *in silico* analysis of a shotgun genome sequence of *S. coeruleorubidus* and identified a strong candidate for the DABA synthase flanked by NRPS-related activities. This approach was fully vindicated when a cosmid clone containing this region was cloned into *S. lividans* and conferred on this

heterologous strain the ability to produce tetrapeptide pacidamycins. These authors were able to suggest plausible roles for many of the genes in the *pac* cluster, but because the individual peptide synthetase enzymes were present as tri-, di- and single domain proteins, rather than in a conventional NRPS assembly-line, more precise assignment was not possible. There are four proteins containing adenylation (A) domains, two of them free-standing. There are also four PCP proteins containing thiolation (T) domains, and three proteins with condensation (C) domains, although five amino acid units need to be connected.

Immediately after appearance of their publication (Rackham et al., 2010) Walsh and colleagues submitted an independent analysis of the same region also based on genome mining (Zhang et al., 2010). The arrangement within the cluster of the 22 genes Pac1-Pac22 (or PacA-PacV) was identical, but the Walsh group also expressed individual A domains and PCPs in *E. coli* and reconstituted parts of the *pac* peptide synthesis *in vitro*.

Further *in vitro* as well as gene deletion work from both groups (Zhang et al., 2011; Zhang and Walsh, 2011; Ragab et al., 2011) has now given a fairly complete picture of the timing and nature of the various steps involving a number of novel enzymes (Figure 1.14). Reactions take place between activated groups tethered to PCPs. The synthesis begins with the attachment of DABA (residue 3) to the PCP PacH (Pac8). Meanwhile, the Ala_4-ureido-Tyr_5 dipeptide unit is constructed by a condensation reaction during which CO_2 reacts with the amino group of mTyr5 to give an N-carboxyamino intermediate (a reaction precedented in syringolin biosynthesis (Imker et al., 2009)). Then m-Tyr (residue 2) is coupled to the β-amino group of DABA, and the ureido-linked dipeptide is coupled to the α-amino group of DABA. Uridine is meanwhile transformed by three enzymes to an aminonucleoside, and this is coupled to the tetrapeptide. Finally, attachment of Ala or another amino acid at the N-terminus of the tetrapeptide intermediate is catalysed by the unusual Ala-tRNA-dependent enzyme PacB (Pac2) (Zhang et al., 2011). At least some of the biosynthetic enzymes have relaxed substrate specificity, which is promising for future efforts to engineer this pathway to create novel nucleoside antibiotics. Also, having characterised this cluster, the lessons learned can be directly applied to an understanding of biosynthetic pathways to other known nucleoside antibiotics, as well as in the uncovering of pathways to novel compounds in this class.

PacLPacL

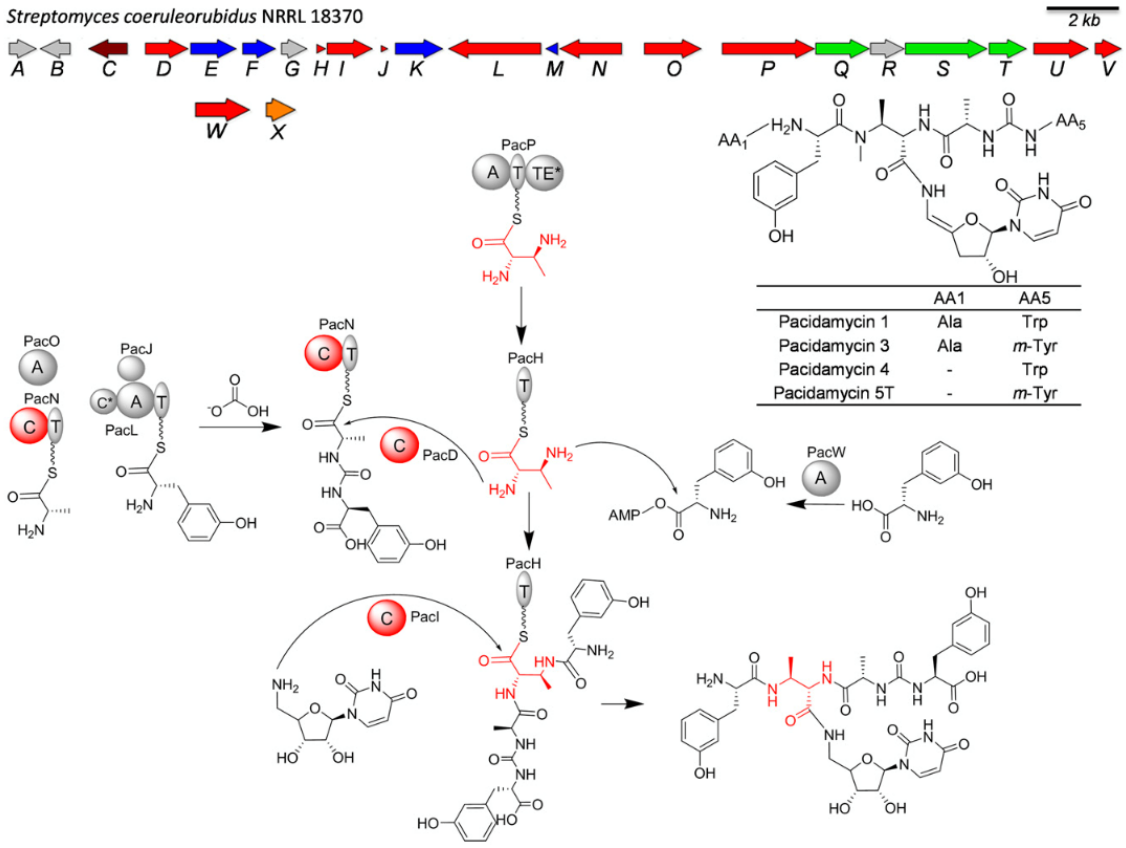


Figure 1.14. The biosynthetic pathway to pacidamycin (adapted from Zhang and Walsh, 2011; and Ragab et al., 2011).

Aims of the project

This project aimed to examine three different aspects of the challenge of linking an actinomycete gene cluster to the specialised metabolite whose biosynthesis it encodes.

First, the prolific antibiotic producing strain *Streptomyces malaysiensis* DSM 4137 has been analysed using RNAseq, to evaluate the use of transcriptomics in informing on those gene clusters (out of a total of at least 54 annotated in the genome) that are transcribed and likely to produce a compound under the conditions used.

Secondly, it was aimed to mobilise the gene cluster for the environmentally-friendly insecticidal spinosyns from the genetically intractable strain *Saccharopolyspora spinosa* and to attempt the refactoring of the cluster in a heterologous strain, to provide a platform for future engineering of the cluster to give new derivatives.

Finally, a genome mining approach has been taken to the discovery of the biosynthetic gene cluster for the remarkable insecticidal nucleoside antibiotic albucidin from *Streptomyces albus* subsp. *chlorinus*, produced in negligible amounts and by an unknown pathway.

Chapter 2 Materials and Methods

2.1 Materials

2.1.1 Chemicals

All chemicals were purchased from Sigma-Aldrich and unless otherwise stated, were analytical grade. Solvents were purchased from Fischer Scientific and were HPLC grade. Distilled water was purified using the Millipore Milli-Q water purification system (Millipore, France).

2.1.2 Antibiotics

Antibiotics were purchased from Sigma-Aldrich. Stock solutions were prepared as in Table 2.1, sterilised through a 0.22 µm filter (Millipore), stored at -20 °C and used at the appropriate concentrations.

Table 2.1 List of antibiotics and the concentrations used.

Antibiotic	Solvent	Stock concentration, mg mL^{-1}	Working concentration µg mL^{-1}
Ampicillin (Amp)	H_20	100	100
Apramycin (Apr)	H_20	50	50
Carbenicillin (Carb)	H_20	100	100
Chloramphenicol (Cam)	C_2H_5OH	25	25
Kanamycin (Kan)	H_20	50	50
Nalidixic acid (Nal)	0.15 M NaOH (aqueous)	25	25
Thiostrepton (Thio)	DMSO	50	50

2.1.3 Enzymes

Fast digest restriction endonucleases were purchased from ThermoFisher Scientific. Calf intestinal alkaline phosphatase, Phusion® high-fidelity PCR master mix with GC buffer, Q5 High-Fidelity DNA Polymerase, T4 DNA ligase, Exonuclease I, Lambda Exonuclease and Gibson assembly® master mix were purchased from New England Biolabs (NEB). Biomix™Red and Biomix™ were purchased from Bioline. PrimeSTAR® GXL polymerase was purchased from Takara. Single stranded (ss)-DNA from salmon testes was purchased from Sigma-Aldrich. Proteinase K powder was manufactured by Melford Laboratories Ltd. All enzymes were stored at -20 °C and used with the buffers provided.

2.14 Culture Media

Table 2.2 Media used in this study

Medium	Composition
Solid	
2TY	1.6 % tryptone; 0.5 % yeast extract; 0.5 % NaCl; 2 % agar
SFM	2 % soya flour; 2 % mannitol; 2 % agar; with tap water
Oatmeal	6 % oatmeal; 12.5 % agar; pH 6.0
ABB13	0.5 % soytone; 0.5 % soluble starch; 0.3 % $CaCO_3$; 2.1 % MOPS buffer, 2 % agar
Liquid	
2TY	1.6 % tryptone; 0.5 % yeast extract; 0.5 % NaCl; pH 7.5
TSB	3% tryptic soya broth (Difco)
SOC	2% tryptone; 0.5% yeast extract; 10 mM NaCl; 2.5 mM KCl; 10 mM $MgCl_2$; 20 mM glucose
SFM	2 % soya flour; 2 % mannitol; with tap water
TSBY	3 % tryptic soya broth (Difco); 103 % sucrose; 5 % yeast extract

Tap Water Medium	1 % glucose; 0.5 % tryptone; 1 % sucrose; 0.25 % yeast extract; 0.0036 % EDTA; pH7.1
Spinosyn production medium	6.8 % glucose; 2.2 % cotton seed flour; 2.5 % peptonised milk nutrient; 1.45 % corn steep liquor; 4.0 % (w/v) methyl oleate; 0.5 % $CaCO_3$;
Oxetanocin production medium	2.0 % soluble starch; 1.5 % soy bean meal; 0.3 % KH_2PO_4; 0.2 % Na_2HPO_4; 0.05 % $MgSO_4 \cdot 7H_2O$; 0.0002 % $CoCl_2 \cdot 6H_2O$; 0.0002 % $FeSO_4 \cdot 7H_2O$; pH 6.0
Nikkomicin production medium SP	3 % mannitol, 1 % soluble starch, 0.75 % yeast extract; 0.5 % soy peptone; pH 6.0

2.15 Buffers

Table 2.3 Buffers used in this work

Chemically competent cell buffers	
Buffer A	10 mM CH_3COOK; 50 mM $CaCl_2$; pH 6.0
Buffer B	10 mM CH_3COOK; 50 mM $CaCl_2$; 20% (v/v) glycerol; pH 6.2

Buffer for genomic DNA purification	
Lysozyme solution	2 mg mL^{-1} lysozyme; 0.3 M Sucrose (10.3%), 25 mM Tris HCl (pH 8.0), 25 mM EDTA (pH 8.0)

Buffer for plasmid alkaline lysis	
Buffer P1	25 mM Tris; 50 mM glucose; 10 mM EDTA; 60 mg mL^{-1} RNAase (added just before use); pH 8.0
Buffer P2	1 % (m/v) SDS; 0.2 M NaOH
Buffer P3	5 M CH_3COOK; adjusted up to pH 4.8 with CH_3COOH

2.16 Oligonucleotides

See appendix 3. Oligonucleotides were purchased from Invitrogen, Eurofins and Sigma-Aldrich.

2.17 Vectors

Table 2.4 Vectors used in this work

Plasmid	Characteristic	Reference
pIB139	aac(3)IV, oriT, attP (ΦC31), int, PermE*, oripUC integration/complementation of mutant strains in trans	(Wilkinson *et al.*, 2002)
pESAC13	neo, tsr, bla, parA, parB, sacB, oriT, attP (ΦC31), int, oriRP1, oriLP1	(Sosio *et al.*, 2000)
pGP9	aac(3)IV, oriT, attP (ΦBT1), int, $_{PactII\text{-}ORF4}$, ΦBT1 site integrative vector with *actII-ORF4/PactI* promoter system, Apr^R	(Gregory *et al.*, 2003)
pUZ8002	*F- dam- 13::Tn9 dcm- 6 hsdM hsdR recF143 zjj-202::Tn10 galK2 galT22 ara14 pacY1 xyl-5 leuB6 thi-1*.	(Paget *et al.*, 1999)
pUB307	kan, RP4	(Flett *et al.*, 1997)
(MacNeil *et al.*, 1992)pAM51△rad	pCRISPomyces2 based plasmid designed for in frame deletion of the 4Fe-4S cluster in *acdB*	This study
F1R1:Alb	pIB139 derived, contains the genes necessary for albucidin production	This study
F1R2:Alb	pIB139 derived, contains the genes necessary for albucidin production	This study
F2R1:Alb	pIB139 derived, contains the genes necessary for oxetanocin production	This study
F2R2:Alb	pIB139 derived, contains the genes necessary for oxetanocin production	This study
pFD54:Alb	pSET152 derived, contains entire AntiSMASH predictive region which contains the albucidin cluster	This study
pIB139:AcdB	pIB139 derived, contains *acdA*	This study
pIB139:AcdA	pIB139 derived, contains *acdA*	This study
pFD030:Nig	pESAC13 derived, contains nigericin gene cluster	This study
pFD0040	pESAC13 derived, contains spinosyn gene cluster	This study
pFD0047	pGP9 derived, contains the rhamnose precursor genes	This study

2.18 Strains

Table 2.5 Strains used in this work

***Escherichia coli* strains**	**Genotype/Characteristics**	**Reference**
DH10B	F- *mcrA* Δ(*mrr- hsd* RMS- *mcrBC*) Φ80*lac*Z ΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ (*ara leu*)7697 *gal*U *gal*K *rps*L *nup*G λ-. Host for general DNA manipulation.	Invitrogen
ET12567/ pUZ8002	F- *dam*- 13::Tn9 *dcm*- 6 *hsdM hsdR* recF143 zjj-202::Tn10 *galK2 galT22 ara14 pacY1 xyl-5 leuB6 thi-1*. Donor strain for conjugation between *E. coli* and *Streptomyces*.	(MacNeil *et al.*, 1992)
TOPO10/pR9604	F- *mcrA* Δ(*mrr- hsd* RMS- *mcrBC*) Φ80*lac*Z ΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ (*ara leu*)7697 7 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G Carbr. Contains mobilisation vector for *E. coli* to *Streptomyces* transfer.	Novagen

***Streptomyces* and *saccharopolyspora* strains**	**Genotype/Characteristics**	**Reference**
S. malaysiensis sp. DSM 4137	See chapter 3	DSMZ
Saccharopolyspora spinosa A83543	Spinosyn-producing wild type strain	(Waldron *et al.*, 2000)
S. coelicolor sp. M1154	Heterologous host strain	(Gomez-Escribano and Bibb, 2011)
S. lividans sp. TK24	Heterologous host strain	(Hopwood *et al.*, 1983)
S. albus sp. J1074	Heterologous host strain	(Zaburannyi *et al.*, 2014)
S. erythraea BIOT2085	Heterologous host strain	PFL
M1154:Nig	M1154 with the nigericin cluster heterologously transferred	This work
TK24:Nig	TK24 with the nigericin cluster heterologously transferred	This work
M1154:Spin1	M1154 with the spinosyn cluster heterologously transferred	This work
TK24:Spin1-3	TK24 with the spinosyn cluster heterologously transferred	This work
BIOT2085:Spin	BIOT2085 with the spinosyn cluster heterologously transferred	This work
J1074:F1R1	J1074 with the F1R1 cluster heterologously transferred	This work

J1074:F1R2	J1074 with the albucidin cluster heterologously transferred	This work
J1074:F2R1	J1074 with the albucidin cluster heterologously transferred	This work
J1074:F2R2	J1074 with the albucidin cluster heterologously transferred	This work
J1074:Alb	J1074 with the albucidin cluster heterologously transferred	This work
J1074:AcdA	J1074 with the albucidin cluster heterologously transferred	This work
J1074:AcdB	J1074 with the albucidin cluster heterologously transferred	This work

2.3 Microbiological Methods

2.31 Growth and maintenance of *E. coli* strains

E. coli strains were grown on solid or liquid LB or 2TY media supplemented with the appropriate antibiotics. Unless otherwise stated, all cultures were grown at 37 ˚C. Liquid cultures were grown in canonical flasks or Falcon tubes and shaken at 220 rpm during incubation.

2.32 Culture conditions for RNA analysis

DSM4137 strains were grown in 10 mL TSBY medium for 36 hours as described. A 30 mL culture was inoculated with 10 % seed medium. Samples were taken at 12, 18 and 30 h in duplicates for RNA extraction.

2.33 Preparation and transformation of chemical competent cells

A fresh plate of *E. coli* cells was streaked on 2TY solid agar to obtain single colonies and grown overnight at 37 °C. 5 mL 2TY was inoculated with a single colony of *E. coli* and incubated overnight at 37 °C, 220 rpm before transferring to 50 mL 2TY. The subsequent culture was again incubated at 37 °C, 220 rpm until an absorbance (A600) of 0.3 - 0.5 was reached. The cells were harvested by centrifugation at 2,200 *x g*, 5 min, 4 °C, and suspended

in 30 mL ice-cold buffer A. After 1 h incubation on ice with occasional agitation, the cells were spun down at 2,200 *x g*, 5 min, 4 °C, and suspended in 3 mL ice-cold buffer B. Aliquots of 100 µL were transferred into pre-cooled 1.5 mL Eppendorf tubes, flash-frozen with dry ice, and stored at -80 °C.

For transformation with plasmid DNA, chemically competent DNA was thawed on ice followed by the addition of 1 µL (10 – 200 ng) purified plasmid DNA. The cells were then left to rest on ice for 30 min before been subjected to heat shock at 42 ˚C for 45 sec. The cells were immediately placed back on ice for 2 min before 500 µL of SOC medium was added and 1 h incubation at 37 ˚C, 220 rpm. The culture was spread onto agar plates containing the appropriate antibiotic(s) and left to grow overnight in a 37 °C incubator.

2.34 Preparation and transformation of electrocompetent cells

For the preparation of electrocompetent *E. coli* cells, overnight inoculations were transferred into 50 mL 2TY as above. The culture was transferred onto ice for 10 min followed by centrifugation for 5 min at 2,200 *x g*, 4°C, to harvest the cells. The cell pellet was carefully resuspended in 20 mL of glycerol (10%, 4°C). To remove salts, a wash step was performed four times, consisting of: (1) centrifugation for 5 min at 2,200 *x g* at 4°C; (2) removal of supernatant; (3) resuspension of pellet in 20 mL of glycerol (15%, 4°C). After the final wash, supernatant was discarded and the pellet resuspended in 1 mL of glycerol (10%, 4°C). Aliquots of 50 µL were transferred into pre-cooled 1.5 mL Eppendorf tubes, flash-frozen with dry ice, and stored at -80 °C.

During routine transformation, an aliquot of electrocompetent cells was thawed on ice and mixed with 0.5 – 2 µL (20 – 200 ng) DNA. The mixture was immediately transferred into an ice-cold 1 mm electroporation cuvette and pulsed at 2.5 kV/mm (25 mF, 200 W) for approximately 5.0 ms – 7.0 ms using a Bio-Rad Gene Pulser II. Electroporation was followed by immediate addition of 750 µL of SOC medium, and the cells were incubated for 1 h at 37 °C, 250 rpm. The culture was streaked onto LB agar plates containing appropriate antibiotic(s) and incubated overnight at 37 °C.

2.35 Electroporation of high molecular weight plasmid DNA

The electroporation efficiency of high molecular weight DNA is low. High efficiency DH10B electrocompetent *E. coli* cells were purchased from NEB. Large plasmid DNA or isothermal assembly mixtures were purified by drop dialysis. Up to 50 µL of plasmid DNA or isothermal assembly mix was dropped onto a 25 mm Type-VS Millipore membrane (shiny side up) in a Petri dish of Milli-Q water. After 1 to 2 h, the drop was carefully collected and the membrane rinsed with 5 µL Milli-Q water. The total amount (up to 30 µL) of DNA or isothermal mixture was then added to a 25 µL aliquot of NEB DH10B electrocompetent cells and the mixture was immediately transferred into an ice-cold 1 mm electroporation cuvette. Electroporation was carried out as described by Sheng, Mancino, and Birren (1995), the mixture was pulsed at 0.9 – 1.2 kV/mm (25 mF, 100 W) for approximately 5.0 ms – 7.0 ms using a Bio-Rad Gene Pulser II. Immediately following electroporation, 500 µL of ice cold SOC medium was added, and the cells were incubated for 1 h at 37°C, 220 rpm. The culture was streaked onto LB plates containing appropriate antibiotic(s) and incubated overnight at 37 °C.

2.36 Growth and maintenance of *Saccharopolyspora* and *Streptomyces* strains

Streptomyces and *Saccharopolyspora* strains were typically maintained in TSBY liquid medium and on SFM plates. For growth in liquid medium, strains were grown from 30 to 72 h in 50 mL flasks (with 7 mL culture) or 250 mL flasks (with 30 mL culture) containing metal springs at 30 ˚C at 220 rpm. For growth on solid media, strains were incubated at 30 ˚C for 5 to 14 days. For spore isolation, strains were grown on SFM, except for *S. albus* subsp. *chlorinus*, *S. erythraea* and *S. spinosa* which were grown on Oatmeal, ABB13 and BHI agar, respectively.

2.37 Growth of *Streptomyces* cultures for metabolite production

S. spinosa and BIOT2085:Spin were grown in 7 mL CSM seed medium for 48 h at 30 ˚C, 220 rpm in 50 mL flasks without springs. A 10 % (v/v) inoculum from seed medium was added to Spinosyn production medium or TSBY. Cultures were grown for 8 days at 30 ˚C, 220 rpm in 250 mL flasks without springs.

TK24:Nig, M1154:Nig and DSM4137 were grown in 7 mL TSBY seed medium for 48 h at 30 ˚C, 220 rpm in 50 mL flasks with metal springs. A 10 % (v/v) inoculum from seed medium was added to 50 ml TSBY or liquid SFM. Cultures were grown for 8 days at 30 ˚C, 220 rpm in 250 ml flasks with metal springs.

S. albus subsp. *chlorinus* was grown in 7 mL TSBY seed medium for 48 h at 30 ˚C, 220 rpm in 50 mL flasks with metal springs. A 10 % (v/v) inoculum from seed medium was added to SFM. Cultures were grown for 8 days at 30 ˚C, 220 rpm in 250 mL flasks with metal springs.

For metabolite production on solid media, strains were grown for 30 – 48 h in 5 mL liquid media and then 200 μL was spread on 30 mL solid media in a Petri dish. Strains were grown for 5 to 14 days until ready for extraction.

2.38 Transformation of *Streptomyces* and *Saccharopolyspora* by conjugation

Transformation of *Streptomyces* mycelia with integrative plasmids was carried out by intergeneric conjugation with *E. coli* following the protocols described in the Laboratory Manual "Practical Streptomyces Genetics". A culture of *Streptomyces* or *Saccharopolyspora* was grown in TSBY medium for 2 to 3 days. 5 mL of the appropriate *E. coli* strain was grown overnight in LB medium containing the appropriate antibiotics at 37 ˚C, 220 rpm. A 5 mL culture of fresh LB medium containing the appropriate antibiotics was then inoculated with 100 µL of the overnight culture and grown at 37 ˚C, 220 rpm until (A600) reached 0.4 – 0.6 . Cells were harvested at 3,000 *x g*, 1 minute, 4 ˚C and the supernatant discarded. Cells were washed three times in LB medium to remove antibiotics. The actinomycetes strain was centrifuged at 3,500 *g* for 5 min and washed twice in LB medium. The actinomycetes strain was resuspended in 500 µL of LB medium. 200 µL was spread on SFM agar as a control plate and 300 µL was mixed with the relevant *E. coli* strains, then spread on SFM or ABB13 agar. Plates were overlaid with 1.5 mL H_20 containing nalidixic acid (25 µg mL^{-1}) and thiostrepton (50 µg mL^{-1}) and half the control plate with thiostrepton (BIOT2085 strains were overlaid with a 10 µg mL^{-1} concentration).

Mycelial or spore conjugation of *Streptomyces* spp. using the direct mating approach was carried out as described above with the exception that the target plasmid was retained in DH10B. The mobilisation plasmid pUB307, was also transformed into DH10B and the two *E. coli* strains were incubated until absorbance (A600) reached 0.3 – 0.5. A three-part mixture

containing 100 µL from each of the DH10B suspensions and 100 µL from the spore or mycelial *Streptomyces* suspension was generated. 200 µL of the mixture was spread on SFM plates and incubated at 30 °C. After 16 – 20 h, mycelial or spore conjugation plates were overlaid with the appropriate antibiotics then incubated at 30 °C until exconjugant colonies appeared.

For *S. spinosa,* conjugations were performed following the method by Xue *et al*. (2013). *S. spinosa* was grown in 30 mL TSBY for 2 to 3 days. The day before conjugation, 10 mL 2TY containing the appropriate antibiotics was inoculated with ET12567/pUZ8002 containing the plasmid of interest and grown at 30 °C, 220 rpm for 12 h. A 50 mL Falcon tube containing 10 mL 2TY was inoculated with 10 % (v/v) overnight culture until A600 reached 0.4 – 0.7. *E. coli* was centrifuged at 2,500 *x g* for 5 min at 4 °C before the supernatant was discarded and then resuspended in 10 mL ice-cold 2TY medium. This was repeated twice more before resuspending in 200 µL H_20. At the same time, myelia were washed twice in 2TY medium and resuspended in 5 mL 2TY medium. 200 uL of mycelia was added to 200 uL of *E.* coli and spread on SFM or R6 medium. After incubation at 37 °C for 24 h, plates were overlaid with the appropriate antibiotics.

2.39 Transformation into *Saccharopolyspora*

Electroporation was performed by the method of Wang et al., 2008. Before spore transformation, plasmids were incubated in cell free extract with the addition of *S*-adenosyl-methionine, as per Donahue *et al.*, 2000, before electroporation. Cell free extracts were prepared from *Saccharopolyspora* sp. in log phase. Cell pellets were resuspended in extraction buffer [20 mM Tris-acetate (pH 7.9), 50 mM potassium acetate, 5 mM Na2EDTA, 1 mM dithiothreitol (DTT), protease inhibitor cocktail (Complete, Mini; Boehringer Mannheim)], sonicated on ice and then centrifuged at 5,*000 x g*, 15 min, 4 ˚C. Extracted plasmid was prepared and added to CFE extract containing 400 mg protein in reaction buffer (20 mM Tris-acetate (pH 7.9), 50 mM potassium acetate, 5 mM Na_2EDTA, 1 mM DTT and 200 mM SAM) to a total volume of 50 µL.

2.4 Molecular biology methods

2.41 Isolation of plasmid DNA from *E. coli*

Isolation of plasmid DNA from an overnight culture was performed using the E.Z.N.A. HP Plasmid Mini Kit I (Omega Bio-Tek), according to the manufacturer's protocol.

2.42 Isolation of PAC DNA

High-molecular weight PAC DNA was isolated from *E. coli* overnight culture by alkaline lysis according to Osoegawa *et al*. (2000). As the alkaline lysis method for PAC purification resulted in fairly impure DNA, the ZR BAC DNA mini-prep kit (Zymo Research) was used in instances when extremely pure PAC DNA was required. Further purification using lambda and exonuclease I resulted in pure PAC DNA.

2.43 Isolation of genomic DNA from *Streptomyces*

For genomic DNA extraction, *Streptomyces* mycelia were spun down and the supernatant discarded. 30 µL of pellet was mixed with 200 uL lysozyme solution [2 mg mL^{-1} lysozyme, 0.3 M Sucrose (10.3%), 25 mM Tris Cl (pH 8.0), 25 mM EDTA (pH 8.0)] followed by incubation at 37 ˚C for 1 to 3 hours with gentle agitation. 200 uL SDS (2%) was added and incubated again for another 50 min. 100 µL phenol/chloroform mixture was then added and inverted gently for 30 min until the liquid looked evenly white. After centrifuging at 4,000 *x g* for 15 min at 4 ˚C, the supernatant was removed with a cut blue pipette tip and 100 µL phenol/chloroform mixture was then added again. This was repeated three times. After the final centrifugation step, the supernatant was transferred with a cut tip to another tube supplemented with 800 µL isopropanol and 80 uL NaAC (3 M). The tube was inverted until genomic DNA appeared. Genomic DNA was spooled on a glass rod and transferred to a clean tube containing 70 % ethanol. After the sample was air dried (10 min at 37 ˚C), 200 – 500 µL water was added to dissolve the genomic DNA.

2.44 RNA extraction

For RNA extraction, 1 mL of sample was mixed with twice the amount of RNAprotect Bacteria Reagent (QIAGEN, 76506), vortexed for 5 sec, incubated for 5 min at room temperature and then centrifuged at 10 min at 5000 *x g*. The supernatant was carefully decanted and dried by dabbing on a paper towel. The resultant pellet was flash-frozen in liquid nitrogen and then stored at -80 ˚C until needed. Preferably, RNA was extracted from samples immediately. With DSM4137, this showed a marked increase in RNA quality.

RNA extraction was prepared according to the RNeasy Bacterial RNA extraction protocol with a minor but important difference. DSM4137 pellets were incubated for 20 min with the appropriate quantity of RLT buffer and lysozyme. Proteinase K was added as per the protocol.

2.45 Polymerase chain reaction

PCR amplifications were carried out using a GeneAmp® PCR system 9700 (PE Applied Biosystems). Phusion® high-fidelity PCR master mix with GC buffer or PrimeSTAR® GXL polymerase was used for cloning, and BioMixTM Red master mix was used for screening purposes. A typical PCR reaction mixture for cloning using Phusion® contained 25 µL of polymerase master mix, 1.5 µL of each primer stock solution, 1.5 µL of DMSO, 1 µL of DNA template, and water up to a final volume of 50 µL. A typical PCR reaction mixture for cloning using PrimeSTAR® GXL contained 1 µL of polymerase, 10 µL of PrimeSTAR® GXL buffer, 0.5 µL of each primer stock solution, 4 µL of dNTP mixture, 10 ng of DNA template, and water up to a final volume of 50 µL. A typical PCR reaction mixture for screening using BioMixTM Red contained 10 µL of polymerase master mix, 1 µL of each primer stock solution, 1 µL of DMSO, 1 µL of DNA template, and water up to a final volume of 20 µL. The following PCR programme was typically used with Phusion®: initial denaturation at 98 ℃ for 10 sec; then 30 thermal cycles which included, 10 sec at 98 ˚C, and then 30 s/kb of DNA at 72 ˚C; final extension at 72 ˚C for 5 min terminated the PCR reaction. For BioMixTM Red: initial denaturation at 98 ℃ for 20 sec; then 30 thermal cycles which included, 10 sec at 98 ˚C, and then 10 sec at 68 ˚C; final extension at 72 ˚C for 5 min terminated the PCR reaction.

2.46 Colony PCR for *E. coli*

For rapid screening of *E. coli* colonies, single colonies were typically restreaked on LB medium (supplemented with the appropriate antibiotic) and grown overnight to confirm the resistance phenotype. The following day a toothpick was used to touch the colony and then was dipped in a 20 µL Biomix reaction and run as above.

2.47 Agarose gel electrophoresis

Agarose gel electrophoresis was carried out using Flowgen horizontal gel tanks (BioRad). Generally, 0.7% (w/v) agarose gels were made using TAE buffer added with SYBR Safe™ stain (3 µL per 100 mL of buffer). DNA samples were mixed with 6X loading dye (Fermentas) before loading on a gel. As a molecular DNA size marker, Generuler 1kb Plus DNA Ladder (Fermentas) was used. DNA was visualised under a UV transilluminator (254 nm) and photographed using a UVP camera.

2.48 Recovery of DNA from agarose gel

Purification of DNA fragments from agarose gels was performed using the KeyPrep Gel DNA Clean Up Kit (Anachem) according to the manufacturer's instructions.

2.49 Recovery of DNA from low melting point agarose gel

Extraction from low melting point gel was undertaken using ß-agarase (NEB) and performed as per protocol. ß-agarase digestions were desalted using drop dialysis as per electroporation protocol.

2.45 Restriction enzyme digestion of DNA

Analytical digestion was carried out in 10 µL final reaction volume (0.2 µL of enzyme(s), 100 ng of DNA, 1 µL of 10X restriction buffer, and water up to the final volume). All reaction

volume was loaded on agarose gel for analysis.

Preparative digestion was performed in 50 µL final reaction volume (2 µL of enzyme(s), 1.5 µg of DNA, 5 µL of 10X restriction buffer, and water up to the final volume). The reaction was incubated in a water bath at 37 °C for 0.5 – 3 h.

2.46 Dephosphorylation of linear DNA

Calf intestinal alkaline phosphatase (CIP) was used to prevent self-ligation of a linearised vector. Immediately following digestion, 1 µL of CIP was added into the reaction. The sample was incubated in water bath at 37 °C for 30 min, followed by gel purification.

2.47 Ligation of DNA

The ligation reaction was typically carried out in 10 µL final volume (1 µL of T4 DNA ligase, 20 – 50 ng of linearised vector, 5X molar excess of DNA insert, 1 µL of 10X ligation buffer, and water up to the final volume). The ligation reaction was incubated at 16 °C for 14 h, after which all reaction volume was used for transformation of *E. coli* DH10B.

2.48 Isothermal assembly and Hot Fusion of DNA

Gibson assembly of DNA was performed according to the manufacturer's guideline. Generally 50 – 70 ng of linearised vector were mixed with at least 3X molar excess of DNA insert(s) and water to give a final volume 10 µL, which was mixed with 10 µL of 2X Gibson assembly® master mix. The reaction was kept at 50 °C for 1 h, after which 10 µL of the reaction was used for transformation of chemically competent *E. coli* DH10B.

For three and four-part isothermal assembly for capturing a gene cluster of interest, 1 µL of purified cut genomic DNA was added to 50 ng of PCR-amplified or cut vector. *In situ* fragments were diluted to supply a 1:1 molar ratio with the plasmid. Hot Fusion was performed as per Gibson assembly but the Hot Fusion mix was assembled as per Fu et al., 2014.

2.49 DNA sequencing

All cloning manipulations were confirmed by sequencing. DNA sequencing was carried out by the DNA Sequencing Facility in the Department of Biochemistry, University of Cambridge.

2.5 Analytical and Chemical Methods

2.51 Analysis of molecules

For analytical purposes, 1 mL samples of culture broth were extracted with 1 mL of ethyl acetate. After the solvent dried out, the residue was redissolved in 500 µL of methanol, centrifuged at 20,000 *x g* for 5 min, and used for HPLC-MS analysis.

For extraction from solid cultures, an agar plate, after 7 – 10 days growth, was cut into small pieces and extracted with equal volume of ethyl acetate. The organic phase was evaporated, the residue was dissolved in 1 mL of methanol and centrifuged at 20,000 *x g* for 5 min, and analysed by HPLC-MS.

For analysis of samples for albucidin and oxetanocin detection, 1 mL samples of culture broth were centrifuged at 20,000 *x g* for 5 min and then transferred to a fresh tube. After 10 min at 4 ˚C, the samples were again centrifuged at 20,000 *x g* for 5 min, and analysed by HPLC-MS.

2.52 Liquid chromatography and mass spectrometry

HPLC-MS analysis was performed using an HPLC (Hewlett Packard, Agilent Technologies 1200 series) coupled to a Finnigan MAT LTQ mass spectrometer (Thermo Finnigan) fitted with an electrospray ionisation (ESI) source. The mass spectrometer was run in positive ionisation mode, scanning from *m/z* 150 to 1800 with fragmentation at 12 % normalised collision energy.

2.53 HPLC analysis

The HPLC was fitted with a Prodigy C18 (250 mm × 4.6 mm, 5µm, Phenomenex) column. A solvent system of acetonitrile and water both containing 0.1 % (v/v) formic acid (FA) was used. Nigericin samples were eluted with a linear gradient of 80 to 100 % of acetonitrile over 20 min, then 100% acetonitrile over 13 min at a flow rate 0.7 mL/min (method A). Spinosyn samples were eluted with a linear gradient of 40 to 75% of acetonitrile over 15 min, then 75 to 95% acetonitrile over 5 min at a flow rate 0.7 mL min^{-1} (method B). Albucidin and oxetanocin samples were eluted with a linear gradient of 2 to 6% of 40 to 75% over 10 min, then 6% 40 to 75% over 11 min at a flow rate 1 mL min^{-1} (method C).

2.54 Measurement of DNA and protein concentration

Concentrations of DNA and protein samples were measured using a Nanodrop spectrophotometer (ND-1000 v3.8.1).

2.55 High resolution mass spectrometry

ESI high resolution MS (HRMS) was carried out by Dr. H. Hong on a ThermoFisher Orbitrap in the Department of Chemistry, University of Cambridge. Spectra were recorded in positive ionisation mode, scanning from *m/z* 100 to 1000, with 60 K resolution and high-resolution fragmentation at 15 % normalised collision energy.

2.56 cDNA sequencing

Total RNA quality and quantity was first checked on an agarose gel and with a nanodrop. Appropriate samples were then checked by using the Agilent RNA 6000 Pico Chip Kit to determine the RNA Integrity Number (RIN) (Agilent Technologies, Böblingen, Germany) before rRNA depletion using the Ribo-Zero rRNA Removal Kit (Bacteria) (Illumina, San Diego, CA, USA). cDNA libraries were created with TruSeq Stranded mRNA Library Prep Kit (Illumina, San Diego, CA, USA), using 200bp RNA fragmentation. Sequencing was undertaken on a Illumina MiSeq system (San Diego, CA, USA) in 76x2 paired read mode.

2.57 Read Mapping

Reads were mapped to the DSM4137 reference sequence using BWA with standard settings. Artemis was used for visualising raw transcripts and for visualising expression graphs. Geneious with Tophat2 was used to calculate expression levels and transcripts per million reads (TPM) of every CDS in the genome. DEseq with R was used for differential expression analysis. Heat maps and graphs were created using GraphPad Prism 7. DEseq with R was used for differential expression analysis. Heat maps and graphs were performed using GraphPad Prism 7 (GraphPad Prism version 7.00 for Mac, GraphPad Software, La Jolla California USA, www.graphpad.com).

2.58 Bioinformatics and software

Genomic sequence DNA was analysed using Artemis Release (Rutherford *et al.*, 2000).

Searches for DNA and protein homologues in public databases were performed using BLAST (Altschul et al., 1990). Multiple protein and DNA sequence alignments were compiled by Geneious version R10 (Kearse et al., 2012). The AntiSMASH 4.0 online software was used as a genome annotation and visualisation tool (Blin et al., 2017). Restriction enzyme cutting sites were checked by Geneious. Geneious was used for molecular biology design, visualisation and analysis. ChemDraw Professional 16.0 was used for chemical drawing and calculation of exact molecular weight of molecules (PerkinElmer, CambridgeSoft). HPLC-MSn, HRMSn data were processed and deconvoluted using Xcalibur v. 2.1 (Thermo Finnigan).

Chapter 3 Transcriptome analysis as a guide for genome mining of the fully-sequenced strain *Streptomyces malaysiensis* DSM4137

3.1 Introduction

Because of the remarkable advances that have been made in DNA sequencing technology, as outlined in Chapter 1, there are now over 80 *Streptomyces* genomes that have been sequenced to completion (defined as the assembly of a high-quality single contig for the entire chromosome) and published (www.ncbi.nlm.nih.gov/genome/browse). However, this number is still dwarfed by the number of highly fragmented draft sequences in the public databases. There is therefore real value in undertaking detailed analysis of the biosynthetic potential of such fully-sequenced strains, not only in their own right but also because the results can draw together and help systematise data from other, related strains.

In this laboratory over the last decade, a number of biosynthetic clusters encoding assembly-line biosynthetic systems have been individually sequenced and characterised in the strain originally deposited as *Streptomyces* sp. DSM 4137 by researchers at Aventis Pharma (Grabley *et al*., 1990). The exact origin of the strain is uncertain, as it first appeared as a contaminant on a plate in the laboratory (J. Wink, personal communication *via* P. F. Leadlay). However, it was quickly recognised as belonging to the large and productive *Streptomyces violaceusniger* clade (Kumar *et al*., 2007), which includes the producers of several clinically valuable compounds. More recently, multilocus sequence *analysis* (*MLSA*) has confirmed this (Samborskyy *et al*., ms in preparation) and allowed the strain to be identified as belonging to the species *Streptomyces malaysiensis*. Next-generation sequencing and an in-house assembly pipeline have recently delivered a complete sequence of the DSM4137 chromosome, comprising 10,694,678 bp of high G+C% DNA (Samborskyy *et al*., ms in preparation). This is only the second species from the *S. violaceusniger* clade to be fully sequenced (the other is *S. violaceusniger* Tü4113, GenBank CP002994.1).

The work described in this chapter concerns the re-annotation of the DSM4137 genome in the light of the availability of the complete sequence; and an evaluation of the use of RNAseq to guide the genome mining of clusters from this strain. It is now a familiar feature

of actinomycete biology that a large proportion of annotated biosynthetic gene clusters are known, or suspected, not to be detectably expressed under laboratory conditions, and information about transcription might shed valuable light on the reasons for this. DSM4137 is a useful and interesting model system to explore this, because it is already known to produce an uncommonly large number of specialised metabolites that have been characterised and that are the products of modular type I PKS, NRPS, or mixed PKS-NRPS systems. Monitoring the behaviour of these known clusters should help both to calibrate the usefulness of RNAseq data and to judge its reliability. Pointers might also be gained to the (better) expression of clusters for unknown specialised metabolites, or of known metabolites not previously connected with this strain.

3.1.1 The use of RNAseq data for *Streptomyces* genome mining

Transcriptomic analysis, the genome-wide analysis of gene transcription, was first made possible using custom-built microarrays. This approach has been frequently used in the model system *S. coelicolor*: for example, mutation analysis of genes that controlled the regulation of actinorhodin and undecylprodigiosin helped uncover expression patterns associated with antibiotic regulation and developmental stages (Huang *et al.*, 2001). Microarray technology has been used to probe the multifaceted regulatory networks of natural product biosynthesis (Kang *et al.*, 2007; Niraula *et al.*, 2010; Noh *et al.*, 2010); to identify substrates for RNase III (Gatewood *et al*., 2012); and to monitor cell wall homeostasis (Hong and Hesketh, 2016). Microarray analysis has also been extended to other *streptomycetes*. For example, global analysis of *S. avermitilis* comparing a high-level and a low-level producer of avermectin led to the discovery that a TetR-family transcriptional regulator was a global upregulator for antibiotic production in *Streptomyces* species (Duong *et al.*, 2009); and microarray analysis confirmed the absence of large *S. coelicolor* genomic islands in *S. lividans* (Jayapal et al., 2007).

With the introduction of next-generation sequencing, RNAseq has become the method of choice for analysing transcriptomes. Instead of relying on hybridisation of custom-made oligonucleotides to a particular region of interest, RNAseq matches transcripts to reference genomes instead (Croucher and Thomson, 2010)(Mantione *et al.*, 2014). Transcription is studied in an unbiased way, sequencing depth is greatly increased (Croucher and Thomson, 2010);(Mantione *et al.*, 2014), genetic features such as operons and promoters can be

confirmed, and the higher resolution allows repetitive regions to be better mapped. RNAseq offers output expressed in transcripts per million (TPM) or reads per kilobase CDS length per million (RPKM) rather than in terms of fluorescence intensity where saturation issues can occur.

Even for strains where a full transcriptomic profile has been undertaken using microarrays as for *S. erythraea* (Peano *et al.*, 2007), subsequent RNAseq analysis has allowed complete re-annotation of the genome, the identification of novel transcripts (Marcellin *et al.*, 2013a), and the detailed monitoring of a genetic switch in metabolism and morphology (Marcellin *et al.*, 2013b). As discussed in Chapter 1, both metabolomic and proteomic approaches have been enlisted to increase the efficiency and scope of *Streptomyces* genome mining. Transcriptomic analysis has been less explored, although it is worth noting the early success in using PCR analysis of cDNA templates to gain insight into the identity of expressed biosynthetic clusters in *Streptomyces flaveolus* (Qu *et al*., 2011). In this chapter, an overview is given of the biosynthetic potential of DSM 4137; and the transcriptomic analysis of those biosynthetic gene clusters will be discussed where the product has both been identified and linked to its corresponding cluster. In the following chapter, Chapter 4, the analysis is extended to examples of clusters which are 'silent', that is, clusters where a reasonably confident prediction can be made of the nature of the product, but that product has not been previously identified from DSM4137.

3.2 Results and Discussion

3.2.1 Updating the annotation of the *S. malaysiensis* DSM4137 genome

The completion of the genome sequence of DSM4137 has revealed a linear chromosome of 10,694,678 nucleotides (Figure 3.1). A small circular plasmid was also sequenced, but it bears no genes for specialised metabolite biosynthesis, and is therefore not considered further here. A number of biosynthetic gene clusters had already been manually annotated, using standard BLAST searches (Altschul *et al*., 1990), including those for the macrocyclic diolide elaiophylin (Haydock et al., 2004), the neuroprotectant meridamycin (Sun *et al*., 2006), the polyether nigericin (Harvey et al., 2007) the antifungal macrolactone azalomycin (Hong et al., 2016) and the giant linear antifungal polyene clethramycin (Hong et al., 2017) (Figure 3.2). The genome sequence had also been scanned with an earlier version

of the automated annotation tool for locating and analysing such clusters, AntiSMASH 3.0 (Blin et al., 2013). For re-annotation, AntiSMASH 4.0 (Blin *et al.*, 2017) was used.

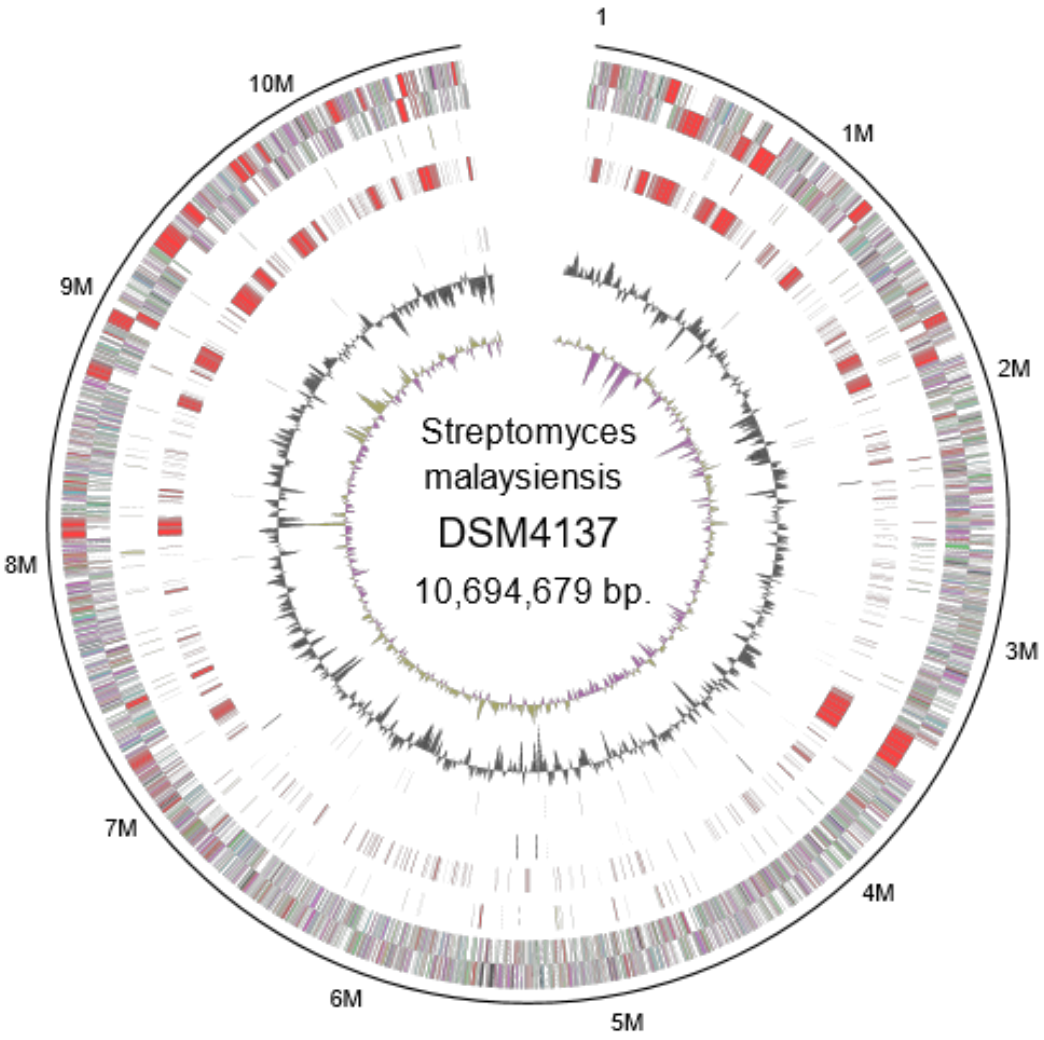


Figure 3.1. The complete genome of *Streptomyces malaysiensis* DSM4137. (Samborskyy *et al*., *ms* in preparation). A circular representation is shown, but the genome is in fact a linear one of 10.7 Mbp. The two outermost rings record the genes on the opposing strands of DNA, colour-coded according to their functional classification (those involved in specialised metabolite biosynthesis third from outside in red). The next ring inwards specifically identifies the location of biosynthetic gene clusters, showing that, as in other sequenced streptomycetes, such clusters are largely concentrated near the ends of the linear chromosome.

Table 3.1 Manually corrected* AntiSMASH 4.0-predicted biosynthetic gene clusters encoded in the genome of *S. malaysiensis* DSM4137.

AntiSMASH Cluster No.	AntiSMASH descriptor	From (nt)	To (nt)	ID range	Annotation
1a	lantibiotic	72117	75738 (75980)*	0071-0072	class II lantibiotic
1b	NRPS	76032	110236	0073-0090	padanamide
1c	terpene	126206	127114	0103	phytoene
2	terpene	253053	260001	0219-0224	diterpene
4a	T1 PKS	369832	412946	0286-0310	conserved
4b	T1 PKS	413021	504880	0311-0339	linear polyene
5	T1 PKS	631600	706630	0447-0481	hygrocin
6	T1 PKS	730929	825626	0506-0524	**meridamycin**
7	terpene	1039981	1044348	0715-0716	**iso-africanol**
not detected	terpene	1073265	1081794	0741-0747	cyclooctatin
8	T1 PKS	1188962	1249297	0854-0870	*geldanamycin*
9a	terpene	1580727	1582810	1169-1170	*methylisoborneol*
9b	other PKS	1586163	1597932	1173-1179	galbonolide
10	T1 PKS	1699304	1766513	1267-1296	*cuevaene*
11	T1 PKS	1812800	1878335	1341-1369	neoansamycin
12	terpene	2180619	2188854	1617-1623	
13	siderophore	2242994	2255645	1667-1676	non-NRPS
14	ectoine	2422479	2435280	1820-1831	
15	terpene	2925401	2929657	2258-2261	
16	lasso peptide	3003641	3025329	2335-2354	
17a	lantibiotic	3444406	3451991	2710-2715	
17b	T1 PKS	3426316	3609691	2696-2709 2716-2730	**clethramycin**
18	ladderane	3789218	3810718	2893-2916	

19	NRPS	3911324	3918285	3024-3026	
20	prenylindole	4150977	4159650	3208-3214	
21	terpene	4779409	4781805	3748	geosmin
22	siderophore	5802498	5810895	4653-4659	desferrioxamines
23	other	6432725	6451073	5228, 5237-5246	echosides
24	siderophore	7024338	7036512	5712-5719	non-NRPS
25	skyllamycin-like depsipeptide	7134319	7197169	5799-5841	"haoxinamide"
26	bacteriocin	7275199	7287706	5901-5909	
27	T1 PKS	7366999	7402288	5985-5998	pentaketide
28	T2 PKS	7559423	7570537	6136-6147	spore pigment
29	terpene	7797128	7812103	6340-6351	hopene
30	NRPS	8078932	8168308	6597-6636	
31	T1 PKS	8268212	8307236	6722-6744	
32	butyrolactone	855035	8505970	6930	
33	homoserine lactone	8520184	8543321	6943-6959	
34	T1 PKS	8710348	8777605	7105-7130	**elaiophylin**
35a	T1 PKS	8878865	8910585	7221-7240	
35b	T1 PKS	8925858	9030545	7255-7282	**nigericin**
36	T1 PKS	9281479	9395505	7490-7512	**azalomycin**
37	NRPS	9440479	9502509	7558-7579	laspartomycin
38	butyrolactone	9550842	9551837	7619	
39	T3 PKS-T1 PKS-NRPS	9684034	9787652	7729-7779	
40	NRPS	9823993	9848268	7817-7828	coelichelin
41	NRPS	10130369	10181285	8091-8120	putative depsipeptide
42	T1 PKS	10260321	10289353	8202-8219	

43a	T1 PKS-NRPS	10399348	10460794	8312-8320	"catramycin"
43b	NRPS	10461123	10482686	8320-8335	prajinamide?
43c	NRPS	10483086	10494458	8336-8348	
44	NRPS	10646693	10662431	8503-8507	

*Compounds shown are either confirmed or are strong predictions for DSM4137. Those in **bold** were detected from DSM4137 in advance of transcriptomic analysis, those in *italics* were subsequently searched for and identified (see section 4.2). Compounds in blue were first detected (after either deletion of genes for major metabolites, or after manipulation of regulatory genes) in near-identical clusters in *Streptomyces* sp. LZ35 (Shandong), which from partial 16S rRNA comparisons may also be a *S. malaysiensis* strain. "Haoxinamide" detection has been reported by the same group in a database entry but its structure has not been published. The limits of the clusters shown were determined by manual analysis. AntiSMASH 4.0 numbering was kept (ID range), but some clusters were judged to be composite and were split up into adjacent distinct clusters. One terpene cluster (cyclooctatin) was not identified by AntiSMASH 4.0. The total number of clusters annotated is therefore 53, approximately double the number in the model organism *S. coelicolor* A3(2).

The analysis revealed a remarkable 44 predicted biosynthetic gene clusters including nine T1 PKS, one T2 PKS, two T1 PKS-NRPS hybrid systems, one T1-T3-PKS_NRPS hybrid and six NRPS, as well as a clutch of terpenes. Examination of the sequence was greatly facilitated by displaying features in the versatile programme Artemis (Carver *et al*., 2012), linked to detailed BLAST results for each identified ORF. In general, AntiSMASH 4.0 found the same complement of assembly-line systems as had an earlier version, but there was improved prediction of RiPPs, and a prenylated indole previously found manually was now detected automatically. The limits chosen by the programme for each cluster are relatively "greedy" and there is a tendency to amalgamate adjacent clusters, but this remains the most useful automated tool available for initial examination of a bacterial genome sequence for its biosynthetic potential.

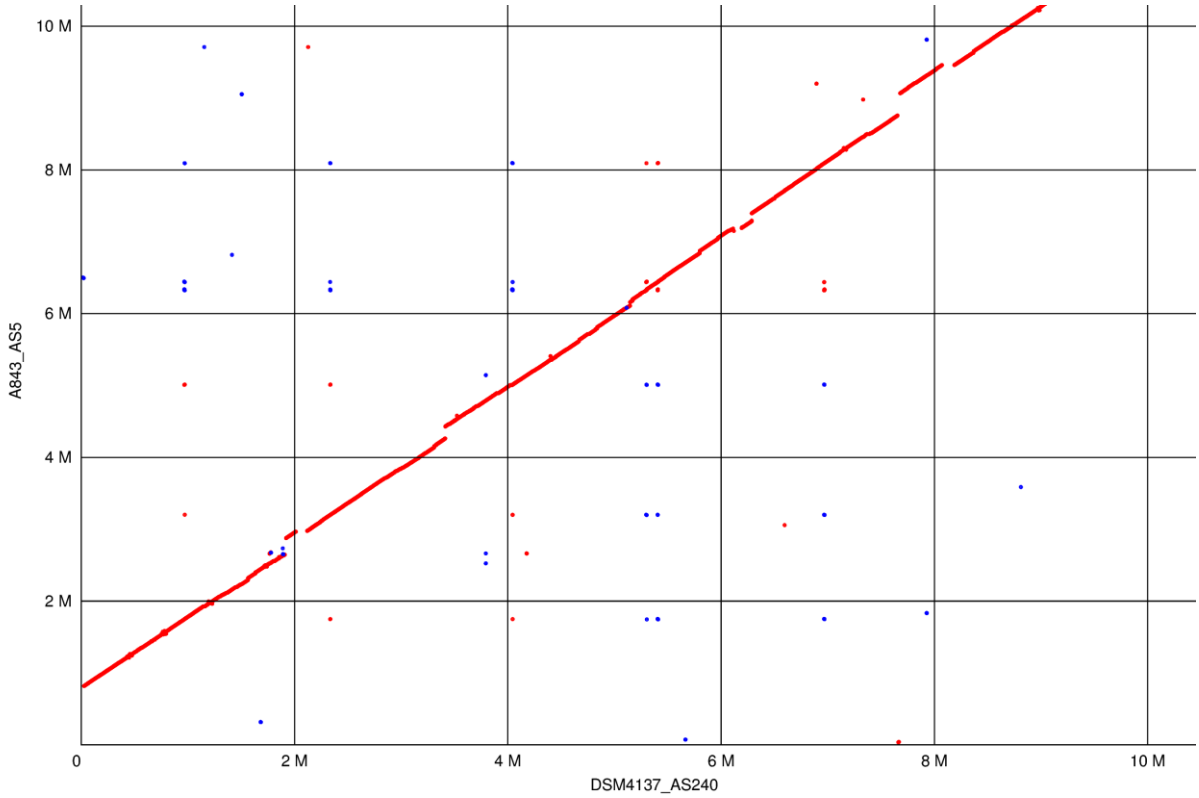


Figure 3.2. Genome sequence alignment of DSM4137 and DSM14702 (A843).

The comparison, made using the program Mummer 3 (Kurtz et al., 2004) shows the expected high degree of conservation of gene order (synteny) between the two strains of *S. malaysiensis* over much of their length. The ends of the chromosome, though, differ markedly: DSM4137 lacks around 1 Mbp present in A843 (this region containing biosynthetic gene clusters 1*-6* in DSM14702) and A843 is missing around 1 Mbp present in DSM4137 at the other end (this region contains biosynthetic gene clusters 37-44 in DSM4137).

3.2.2 Annotation of the genome of *Streptomyces malaysiensis* DSM14702, producer of the antifungal nucleoside malayamycin

Recent work in this laboratory by Dr Hui Hong has revealed the biosynthetic gene cluster for the potent antifungal compound malayamycin (Li et al., 2008), and uncovered the roles of key biosynthetic enzymes (Hong *et al*., ms in preparation). The cluster was uncovered by whole-genome sequencing, and this gave the opportunity in the present work to compare a second strain of *S. malaysiensis* with DSM4137. After re-assembly of the genome by Dr Markiyan Samborskyy into two contigs (a total of 10,635,977 bp), the two genome sequences were compared using the program *Mummer 3* (Kurtz *et al*., 2004). The comparison (Figure 3.2) showed extensive overlap in their complement of biosynthetic gene clusters, which greatly assisted the annotation of DSM14702 (listed in this laboratory's strain collection as A843) using AntiSMASH 4.0 and the previously annotated DSM4137. The location, extent and predicted identity of the clusters uncovered, corrected manually by comparison with DSM4137, are summarised in Table 3.2.

3.2.3 Production of known major metabolites of DSM4137 under the conditions to be used for transcriptomic analysis

DSM4137 was grown in SFM liquid medium and samples from extracts of both culture supernatant and cell pellets were analysed by HPLC-MS after two, four and six days, by Dr J. Xu in this laboratory. The results can be briefly summarised as follows: the polyether nigericin was present at all times tested in both supernatant and pellet; the macrocyclic polyene azalomycin and the linear polyene clethramycin were both present at all times tested, in the pellet only; and the macrocyclic neuroprotectant meridamycin was only present at the two days time point, in both supernatant and pellet. The macrodiolide polyketide elaiophylin was not detected during growth on this medium under these conditions, although it was clearly detectable on other media. The structures of these compounds are shown in **Figure 3.3**. Nigericin, azalomycin, clethramycin and elaiophylin are all the products of canonical modular type I PKS systems, and meridamycin is the product of a hybrid modular type I PKS-NRPS. Two of these well-characterised clusters (nigericin and clethramycin) will be used in this chapter to exemplify the ability of transcriptomics to monitor the timing and level of expression. The full RNAseq data for clusters not described in detail here can be found in Appendix 1.

Table 3.2 Manually corrected* AntiSMASH 4.0-predicted biosynthetic gene clusters encoded in the genome of *S. malaysiensis* DSM14702

AntiSMASH Cluster No.	AntiSMASH descriptor	From (nt)	To (nt)	ID range	Annotation
1*	TIII PKS	138847	140031	0148	
2*	T1 PKS	201435	227040	0202-0203	hexaketide
3*	T1 PKS	227173	318209	0204-0215	bafilomycin
4*	T1 PKS	318310	338185	0216-0222	triketide
5*	nucleoside	338796	362530	0224-0246	**malayamycin**
6*	NRPS	547069	557636	0423-0429	NRPS
1a	lantibiotic	868248	868418	0722-0723	class II lantibiotic
1b	NRPS	868712	902934	0724-0742	padanamide
1c	terpene	918904	919812	0756	phytoene
2	terpene	1042644	1049592	0874-0879	brasilicardin-like
3	NRPS	1095324	1141523	0922-0931	conserved
4a	T1 PKS	1157279	1199331	0944-0968	conserved
4b	T1 PKS	1199406	1291299	0969-0995	linear polyene
5	T1 PKS	1412645	1487308	1105-1137	hygrocin
6	T1 PKS	1510278	1605020	1162-1177	meridamycin
7	terpene	1820087	1824457	1370-1371	iso-africanol
not detected	terpene	1854734	1863443	1397-1403	cyclooctatin
8	T1 PKS	1957608	2017982	1499-1515	geldanamycin
8a*	NRPS	2288140	2326836	1769-1793	
9a	terpene	2336112	2339326	1806-1808	methylisoborneol
9b	other PKS	2341445	2353463	1810-1816	galbonolide
10	T1 PKS	2461989	2523604	1908-1936	cuevaene
11	T1 PKS	2557453	2621725	1973-1999	neoansamycin
12	terpene	3038631	3054830	2385-2395	indole terpene
13	siderophore	3096684	3114444	2432-2446	non-NRPS
14	ectoine	3285716	3300260	2595-2605	
15	terpene	3779820	3792887	3041-3052	
16	lasso peptide	3003641	3025329	2335-2354	

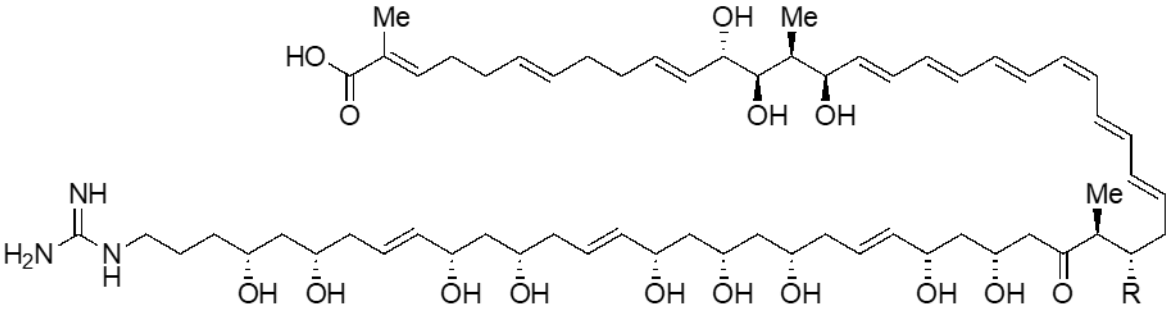
17a	lantibiotic		*Not present*		
17b	T1 PKS	4443032	4617399	3660-3697	clethramycin
18	ladderane	4784884	4806312	3845-3868	
19	NRPS	4893810	4898802	3964-3965	
20	prenylindole	5115553	5124246	4132-4139	
21	terpene	5733083	5735425	4677	geosmin
22	siderophore	6874952	6875611	5691-5697	desferrioxamines
23	other	7540491	7559932	6341, 6352-6361	echosides
24	siderophore	8151678	8162449	6851-6858	non-NRPS
25a*	terpene	8248000	8259963	6931-6941	
25	skyllamycin-like depsipeptide	8266137	8328674	6946-6985	"haoxinamide"
26	bacteriocin	8405793	8417963	7044-7051	
27	*T1 PKS*		*Not present*		*pentaketide*
27a*	RiPP	8473568	8478368	7105-7108	putative linaridin
28	T2 PKS	8658873	8669987	7274-7285	spore pigment
********	*********	*******	*******	********	end of scaffold 1
29	terpene	281671	297588	0811-0805	hopene
30	NRPS		*Not present*		
31	T1 PKS	629197	639567	0811-0805	PKS
32	butyrolactone	876292	879780	0584-0581	
33	homoserine lactone	893966	918840	6943-6959	
34	T1 PKS	1083473	1143635	0407-0386	elaiophylin
35a	T1 PKS	1241845	1265791	0294-0279	PKS
35b	T1 PKS	1288709	1394384	0259-0230	nigericin
35c*	NRPS	2288140	2326836	1769-1793	
35d*	NRPS-PKS	1428401	1506232	0200-0171	putative tetramate
36	T1 PKS	1708406	end	0004-end	*incomplete azalomycin cluster*
37	NRPS		*Not present*		laspartomycin
38	butyrolactone		*Not present*		
39	T3 PKS-T1 PKS-NRPS		*Not present*		

40	NRPS	*Not present*	coelichelin
41	NRPS	*Not present*	putative depsipeptide
42	T1 PKS	*Not present*	
43a	T1 PKS-NRPS	*Not present*	"catramycin"
43b	NRPS	*Not present*	
43c	NRPS	*Not present*	
44	NRPS	*Not present*	

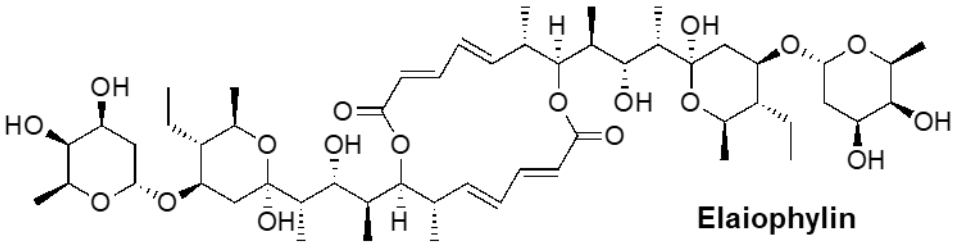
*Clusters shown with an asterisk are present in DSM14702 only. Clusters shown in italics are present in DSM4137 only. Apart from malayamycin, the secondary metabolome of DSM14702 has not yet been investigated.

Nigericin was the first polyether to be described (Harned *et al*., 1951), and study of its biosynthesis led to a significant revision in our model of the post-PKS oxidative cyclisation steps in such pathways (Harvey *et al*., 2007). Clethramycin was first isolated from *Streptomyces hygroscopicus* TA-A0623 (Furumai *et al*., 2003) and found to inhibit plant pollen tube emergence, as well as showing antifungal activity against *Candida albicans* and *Candida glabrata* (Igarashi *et al*., 2003). Since then it has been detected in several other species within the *S. violaceusniger* clade, including DSM4137 where it occurs along with the co-metabolite desulfoclethramycin, *Streptomyces mediocidicus* (Hong *et al*., 2017, in the press) and the rapamycin-producing *Streptomyces rapamycinicus* (K. Usachova, unpublished). The clethramycin PKS is one of the largest known modular PKS, housing 27 extension modules.

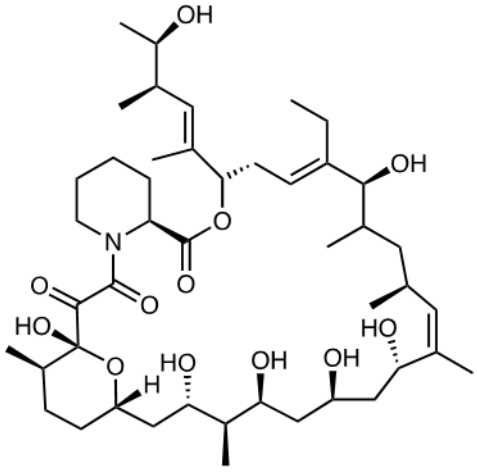
It can be seen from Table 3.1 that DSM4137, in common with very many other streptomycetes, harbours several clusters that have been extensively characterised in the model organism *S. coelicolor*. Among these are the iron-scavenging compounds desferrioxamine E (cluster 22) and coelichelin (cluster 40), where the gene arrangement within each cluster is identical to the counterpart cluster in *S. coelicolor*; signalling molecules butyrolactone (clusters 32 and 38) and homoserine lactone (cluster 33); ectoine (cluster 14) and bacteriocin (cluster 26); and hopene/aminotrihydroxybacteriohopane (ATBH) (cluster 29) (Challis, 2014). These compounds are also shown for reference in Figure 3.3. No attempt was made to search DSM4137 cultures for the presence of these "generalist" metabolites.



Clethramycin



Elaiophylin



Meridamycin

Azalomycin

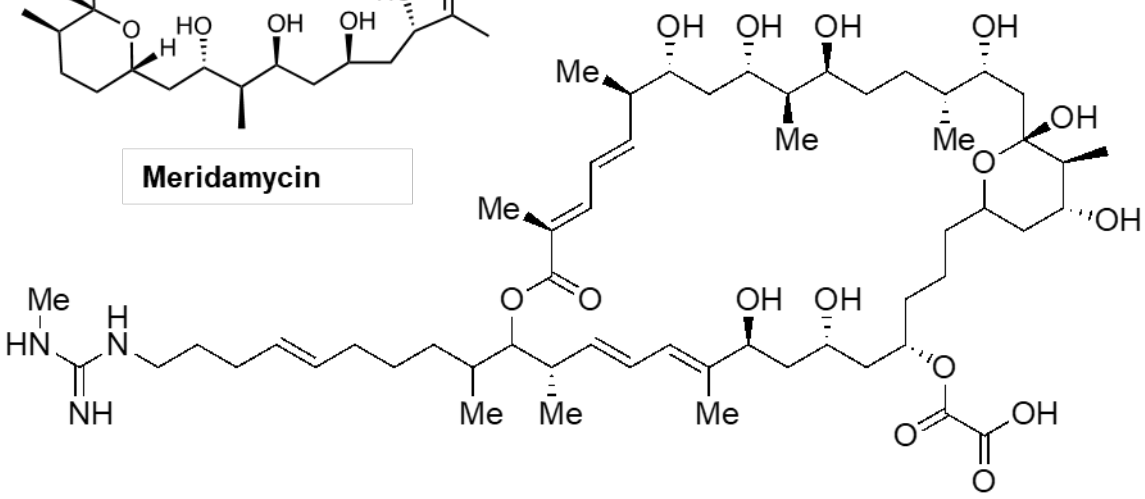


Figure 3.3 The major metabolites of DSM4137

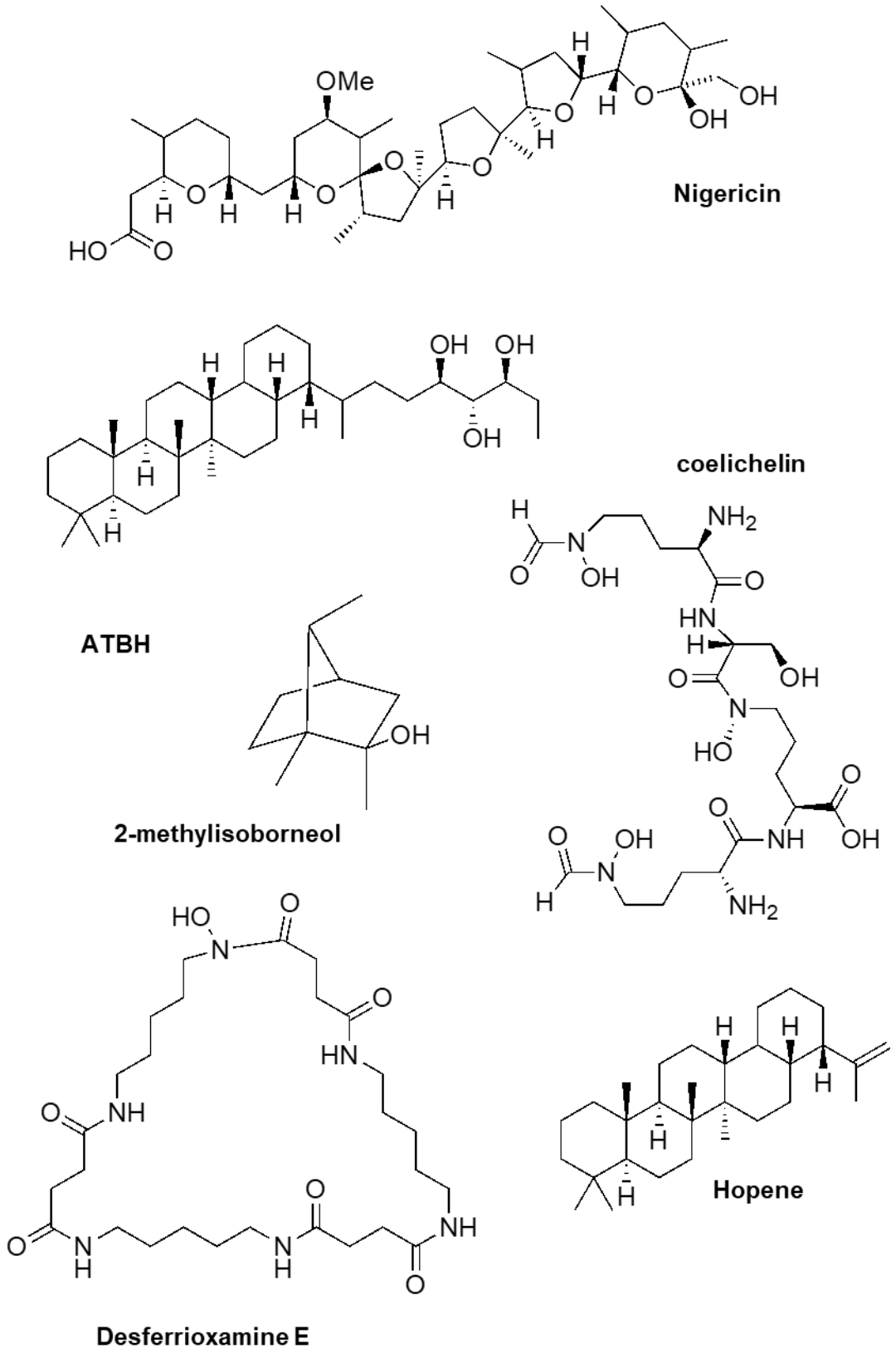


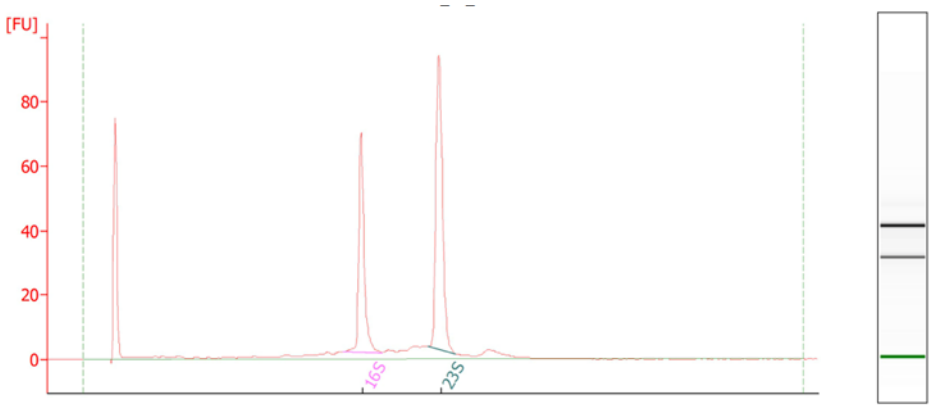
Figure 3.3 b. Common metabolites from *Streptomyces*

3.2.4 Isolation of RNA from *S. malaysiensis* DSM4137 and preparation of libraries for for RNAseq analysis

Significant experimentation was needed to optimise RNA extraction and purification to obtain RNA of good quality from DSM4137. Despite repeated attempts, RNA isolated from later stages of growth (>50-60 hours) was particularly vulnerable to degradation, perhaps because of a greatly increased level of degradative enzymes at this stage of growth. Nevertheless, satisfactory RNA preparations were made from samples grown for 12, 24, 30 and 48 hours respectively. Clearly, it would be beneficial eventually to extend the analysis to later time points, but there is some comfort in evidence from transcriptomic analysis of multiple *Salinispora* spp. that in those actinomycetes, a surprising majority of biosynthetic gene clusters that are expressed are already transcribed during an early phase of growth (Letzel *et al.*, 2017; and P. R. Jensen, personal communication).

Extraction of RNA was done using the RNeasy kit (Qiagen). 50 mL of SFM and TSBY liquid media were inoculated with 5 mL of a culture of DSM4137 grown in TSBY for 3 days. After 12, 18, 30 and 48 hours respectively, 1 mL of culture was added to 2 mL "RNA protect", vortexed and left at room temperature for 5 minutes. After centrifugation, the pellet was flash frozen in liquid nitrogen. For DSM4137, this protocol tended to give poor quality RNA. Better quality RNA was achieved from RNA that was extracted immediately. Several modifications were tried: for example, DSM4137 cell pellets were pre-incubated with lysozyme for 15-20 minutes instead of the recommended 10 minutes. However, this also increased RNA degradation. After RNA extraction with the RNeasy kit, RNA was analysed using gel electrophoresis to check for degradation and then analysed on a Bioanalyzer (Agilent) to check for the RNA Integrity Number (RIN) (Figure 3.4). An RIN of 8 or more is high enough to warrant preparing an RNA library (Imbeaud *et al*., 2005).

A detailed analysis of RNA library preparation, rRNA depletion and sequencing can be found in materials and methods section 2.56 to 2.58. Once RNAseq data had been aligned and assembled, numerous options were explored to display the data in the most convenient and informative way including displaying RNA transcriptits above the genome annotations with the Artemis genome browser (Rutherford *et al.*, 2000). The heatmap was deided to be the best tool to use, using the rainbow heatmap it is possible to observe subtle changes in RNA levels and compare the transcript levels of different clusters.



Overall Results for sample 1 : DSM4137_d2_1

RNA Area:	294.2	rRNA Ratio [23s / 16s]:	1.5
RNA Concentration:	946 pg/µl	RNA Integrity Number (RIN):	9.8 (B.02.08)

Fragment table for sample 1 : DSM4137_d2_1

Name	Start Size [nt]	End Size [nt]	Area	% of total Area
16S	1,297	1,789	74.4	25.3
23S	2,551	3,076	108.0	36.7

Figure 3.4. Estimation of the RNA Integrity Number for DSM4137 RNA using the RNA Bioanalyser. An RIN of over 8 represents RNA of sufficient quality for RNA-seq analysis.

3.2.5 Transcription of the clethramycin biosynthetic gene cluster

The annotated gene cluster for clethramycin lies within a region of 43 contiguous ORFs (SMALA_2696 to SMALA_2738, although not all genes in this region are assigned to clethramycin biosynthesis. It is an interesting example of "a cluster within a cluster" because a number of adjacent genes within this region are predicted to be involved in biosynthesis of an unknown lantibiotic (section 3.2.6). The detailed annotation of the clethramycin cluster is given in Table 3.2, alongside the RNAseq data. On the following page is shown Figure 3.5, the heat map representation of the data of Table 3.2.

From the heat map it is clear that clethramycin genes are transcribed at the earliest stages of DSM4137 fermentation and this is fully consistent with the LC-MS data. The PKS genes, SMALA_2722 to SMALA_2730 are all turned on early and together. It is noticeable that expression of the PKS genes remains low, and transcription levels then drop (though the significance of the low level at 30 hours needs to be confirmed by further experiment). The major clethramycin metabolite observed at all stages of the DSM4137 fermentation is 32-*O*-desulfoclethramycin, the immediate precursor of clethramycin (Figure 3.3). The molar ratio of desulfo/clethramycin remains constant during the fermentation, and the transcriptomic analysis shows that the sulfotransferase (*stf*) gene catalysing this last step is being transcribed

at all time points studied. Notably, all the putative regulatory genes (highlighted in yellow) are transcribed.

Table 3.3. The results of transcriptional analysis of the clethramycin biosynthetic gene cluster in *S. malaysiensis* DSM4137.

locus_tag	gene	Annotated name	Length	S1_12hr_ TPM	S2_18hr_ TPM	S1_30hr_ TPM	S1_48hr_ TPM
SMALA_2696		LysR regulator	936	2.95	2.55	2.95	7.4
SMALA_2697	stf	stf	1044	35.75	41.9	44.45	33.9
SMALA_2698	gdmRII	gdmRII - LuxR family transcriptional regulator	2649	10.6	12.65	10.8	7.9
SMALA_2699		arginine oxidase	1659	43.5	48	45.35	34.7
SMALA_2700		CoL	1410	31.55	36.3	28.6	11.4
SMALA_2701		oleoyl-_acyl-carrier-protein_ hydrolase	624	103.8	144.05	109.35	35.8
SMALA_2702		daunorubicin resistance ABC transporter ATPase subunit	981	185.1	214.25	185.3	45.9
SMALA_2703	ABC transpo	ABC transporter permease	1434	103.7	114.5	110.35	42.9
SMALA_2704		_acyl-carrier-protein_S-malonyltransferase-like protein	939	128.55	126.15	119.6	40
SMALA_2705		transcriptional regulatory protein	600	4.6	6.05	7.2	16
SMALA_2706		hydrolase	915	5.15	6.55	4.75	12.6
SMALA_2707		TetR family transcriptional regulator	600	30.1	22.35	24.8	37.4
SMALA_2708		short-chain dehydrogenase_reductase SDR	747	9.9	8.55	11.45	10.2
SMALA_2709		glycosyl transferase	435	115	132.45	116.65	54.9
SMALA_2710		protein-L-isoaspartate_D-aspartate_O-methyltransferase	504	8.8	7.45	6.6	35.4
SMALA_2711		lanthionine synthetase C-like	945	0.85	0.75	0.45	11.8
SMALA_2712		lantibiotic dehydratase-like	3039	1.35	0.9	1.2	11.8
SMALA_2713		hypothetical protein	165	6.3	3.35	6.1	30.4
SMALA_2714		protein-L-isoaspartate_D-aspartate_ O-methyltransferase	1203	6.55	5.35	7.1	18.8
SMALA_2715		taurine catabolism dioxygenase TauD_TfdA	864	3.95	2.75	3.6	17.8
SMALA_2716		NUDIX hydrolase	468	124	55.9	97.3	64.5
SMALA_2717		Xre transcriptional regulator	1170	17.85	13.55	20	32.5
SMALA_2718		regulatory protein	210	57	57.15	56	131.7
SMALA_2719		putative DNA-binding protein	660	10.4	11.15	12.5	122.6
SMALA_2720		hypothetical protein	321	8.85	8.45	7.35	18.7
SMALA_2721		hypothetical protein	216	16.45	19.5	16.05	16.9
SMALA_2722	cleA9	cleA9 - modular polyketide synthase	11709	13.4	12.45	15.1	5.9
SMALA_2723	cleA8	cleA8 - type I modular polyketide synthase	34698	8.7	7.5	8.2	2.4
SMALA_2724	cleA7	cleA7 - type I modular polyketide synthase	9621	22.55	21.15	22.3	7.6
SMALA_2725	cleA6	cleA6 - modular polyketide synthase	22038	19.45	22.95	20.5	10
SMALA_2726	cleA5	cleA5 - type I modular polyketide synthase	15753	23.7	30.55	22.95	7.3
SMALA_2727	cleA4	cleA4 - modular polyketide synthase	4995	27.45	23.05	29.9	12.1
SMALA_2728	cleA3	cleA3 - type I modular polyketide synthase	25800	15.15	13.8	15.35	5.5
SMALA_2729	cleA2	cleA2 - modular polyketide synthase	10335	13.6	14.75	12.9	5.5
SMALA_2730	cleA1	cleA1	24624	9.95	12.6	10.35	5.1
SMALA_2731		membrane-flanked domain-containing protein	1545	61.75	62.35	61.3	126.7
SMALA_2732		membrane-flanked domain protein	501	140.5	140.7	148.1	293
SMALA_2733		integral membrane protein	558	590.6	684.25	523.25	2119.3
SMALA_2734		sensor kinase, two-component system	1245	60.7	46	56.75	45.4
SMALA_2735		two component transcriptional regulator, LuxR family	561	37.3	24.25	33	19.5
SMALA_2736		gdmRII - LuxR family transcriptional regulator	2556	12.3	12.3	11.2	12.4
SMALA_2737		oleoyl-_acyl-carrier-protein_ hydrolase	762	164.1	222.85	175.8	102
SMALA_2738		putative integral membrane protein	2115	10.8	5.9	10.35	7.5

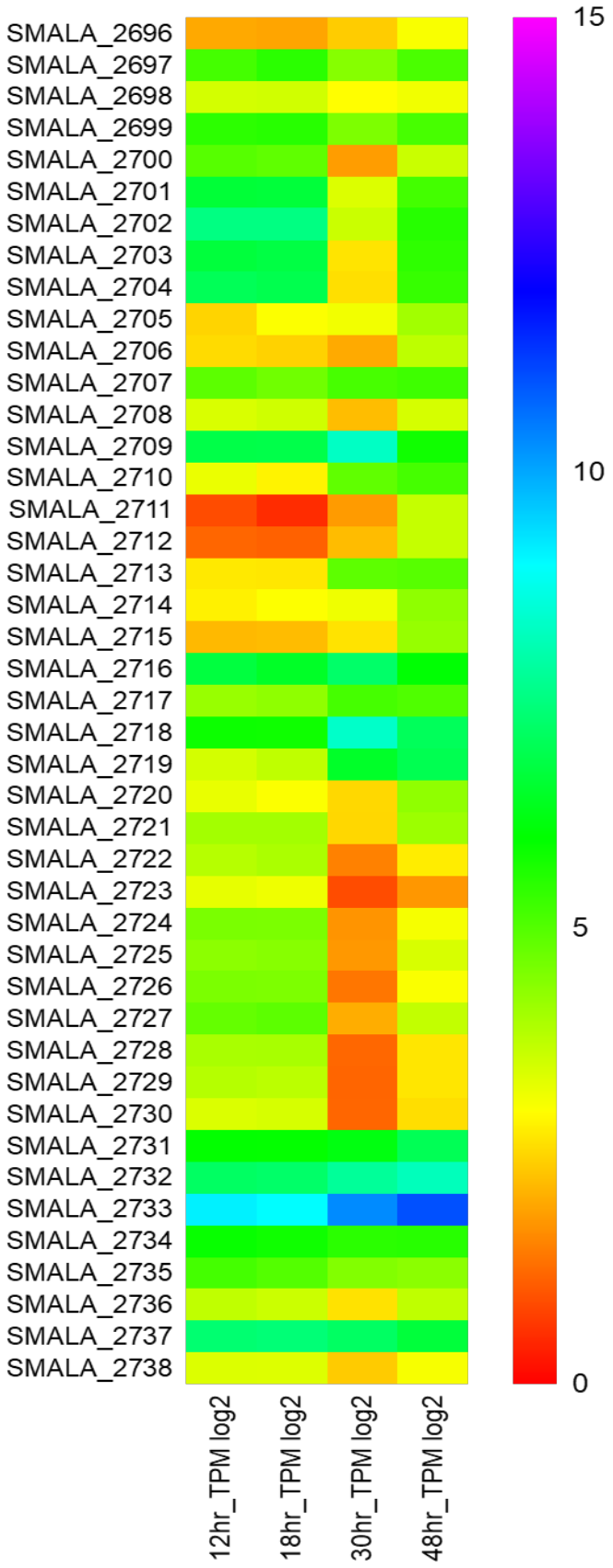


Figure 3.5. Heat map of RNAseq data for the clethramycin biosynthetic gene cluster *S. malaysiensis* DSM4137*

3.2.5 Transcription of a putative lantibiotic biosynthetic gene cluster in DSM4137

SMALA_2713 encodes the preprotein, and SMAL_2711 and SMALA_2712 encode key enzymes that act upon it to generate the mature antibiotic. Flanking genes may provide additional tailoring. It is striking that SMALA_2711 and SMALA_2712 behave differently to the surrounding clethramycin genes. These two genes are apparently silent until 48 hours, when all the genes of the putative cluster are being transcribed. It is likely that this is an active cluster in DSM4137.

Table 3.4. The results of transcriptional analysis of a putative lantibiotic biosynthetic gene cluster in *S. malaysiensis* DSM4137.

locus_tag	gene	Annotated name	Length	S1S2_12hr _TPM	S1S2_18hr _TPM	S1S2_30hr _TPM	S1S2_48hr _TPM
SMALA_2710		protein-L-isoaspartate_D-aspartate_O-methyltransferase	504	8.8	7.45	6.6	35.4
SMALA_2711		lanthionine synthetase C-like	945	0.85	0.75	0.45	11.8
SMALA_2712		lantibiotic dehydratase-like	3039	1.35	0.9	1.2	11.8
SMALA_2713		hypothetical protein	165	6.3	3.35	6.1	30.4
SMALA_2714		protein-L-isoaspartate_D-aspartate_ O-methyltransferase	1203	6.55	5.35	7.1	18.8
SMALA_2715		taurine catabolism dioxygenase TauD_TfdA	864	3.95	2.75	3.6	17.8

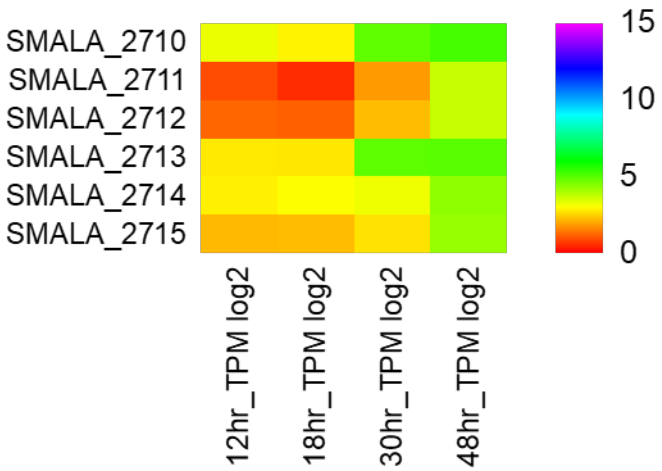


Figure 3.5. Heatmap representation of RNAseq data for nigericin biosynthetic gene cluster.

3.2.5 Transcription of the nigericin biosynthetic gene cluster in DSM4137

The annotated nigericin cluster shown in **Table 3.5** comprises 59 contiguous ORFs but the original annotation (Harvey *et al*., 2007) extended the cluster only from SMALA_2763, encoding the cluster-situated regulator (CSR) NigR, a member of the SARP family of DNA-binding proteins (Wietzorrek and Bibb, 1997), to SMALA_7282 which probably encodes a member of a novel family of spirocyclase enzymes (O. Bilyk, personal communication) that could play a role in nigericin biosynthesis. SMALA_7284 encodes a transposase, so this flank of the cluster is reasonably well-defined. For the purposes of transcriptomic analysis, a number of adjacent genes on the other flank are analysed here, up to SMALA_7255, to see how their transcription differs from those of the cluster itself.

The heat map for this region (Figure 3.6) shows a different pattern to that of clethramycin biosynthetic gene cluster. Nigericin is produced later than clethramycin: it is undetectable on day one, can first be detected on day two, and transcription increases thereafter, plateauing. after day seven, again later than clethramycin. The gene for the CSR NigR, SMALA_7263, is transcribed throughout and its transcription increases in step with increasing production of nigericin. This is consistent with its previous identification as an activator (Harvey *et al*., 2007).

3.3 Concluding remarks

The comparison of the genomes of two strains of *Streptomyces malaysiensis*, DSM4137 and DSM14702, has provided a striking example of how the availability of fully-sequenced genomes has allowed a deep analysis of the relationship between them. It seems highly probable that further strains currently simply described as "*Streptomyces* sp. XYZ from soil" may turn out also to be closely related to *S. malaysiensis* and will benefit from the current analysis, as well as focussing attention on endemic, or strain-specific, clusters that are, perhaps, more likely to be novel.

The examples presented here, of the transcriptomic analysis of the biosynthetic gene clusters for clethramycin and nigericin, together show that the RNAseq analysis of DSM4137, even though limited to a single growth medium and to relatively early time points in the fermentation, nevertheless has delivered a wealth of information about the timing and coordination of transcription of cluster genes. To the extent that the data can be correlated

with data from LC-MS analysis of the production of the specialised metabolite, there is also a reasonable match. Indeed, Jensen and colleagues report (Letzel *et al.*, 2017).

Table 3.5. Transcriptional analysis of the nigericin biosynthetic cluster in DSM4137

locus_tag	gene	Annotated name	Length	S1S2_12 hr_TPM	S1S2_18 hr_TPM	S1S2_30 hr_TPM	S1S2_48 hr_TPM
SMALA_7255		RNA polymerase, sigma-24 subunit, ECF subfamily	873	0.25	0.45	0.6	5.9
SMALA_7256	prpB6	prpB6 - magnesium or manganese-dependent protein phosphatase	1110	0.55	0.4	1.1	10.4
SMALA_7257	clpA	clpA - Clp protease ATP binding subunit	2610	0.75	0.35	0.6	9
SMALA_7258		capsule synthesis protein, CapA	1170	0.35	0.35	0.25	10.3
SMALA_7259		hypothetical protein	162	0.45	0	0	4.3
SMALA_7260		methyltransferase	852	0.5	0.45	0.65	2.3
SMALA_7261		x-prolyl-dipeptidyl aminopeptidase	1911	0.8	0.3	0.5	12.9
SMALA_7262		faal	231	3.3	3.3	2.4	67.3
SMALA_7263	nigR	SARP family pathway specific transcriptional activator	783	4.65	3.95	3.65	63.9
SMALA_7264	nigE	methyltransferase type 11	813	6.4	2.85	4.85	38.8
SMALA_7265	nigA1	nigA1 - modular polyketide synthase	7248	0.15	0.15	0.15	3.3
SMALA_7266	nigA1I	nigA1I - type I modular polyketide synthase	6702	0.25	0.1	0.05	8.4
SMALA_7267	nigAIII	nigAIII - type I modular polyketide synthase	12252	0.05	0	0	2.2
SMALA_7268	nigAIV	nigAIV - type I modular polyketide synthase	12162	0.2	0.15	0.2	9.1
SMALA_7269	nigAV	nigV - type I modular polyketide synthase	12144	0.2	0.1	0.1	10.1
SMALA_7270	nigAVI	nigAVI - modular polyketide synthase	5106	0.15	0.05	0.15	3.5
SMALA_7271		nigCII	867	2.5	2.15	3.3	21.4
SMALA_7272	nigD	cytochrome P450 CYP124E1	1260	0.6	0.25	0.5	25.3
SMALA_7273	nigAXI	NigAXI - modular polyketide synthase	339	1.3	1.1	1.2	17.2
SMALA_7274	nigAX	nigAX - modular polyketide synthase	3900	0.25	0.15	0.15	3.2
SMALA_7275		nigCI	1431	0.45	0.05	0.15	8
SMALA_7276		nigBII	468	1.6	0.55	0.35	45.3
SMALA_7277		nigBI	438	0.75	0.8	1.3	24.2
SMALA_7278	nigAIX	beta-ketoacyl synthase	5829	0.2	0.3	0.3	5.8
SMALA_7279		erythronolide synthase, modules 3 and 4	6624	0.2	0.15	0.2	4.6
SMALA_7281	nigAVII	nigAVII - modular polyketide synthase	14361	0.5	0.25	0.35	13.6
SMALA_7282		hypothetical protein	648	1.6	0.95	1.25	35.7
SMALA_7283		hypothetical protein	1398	0.65	1.25	1.55	11.5

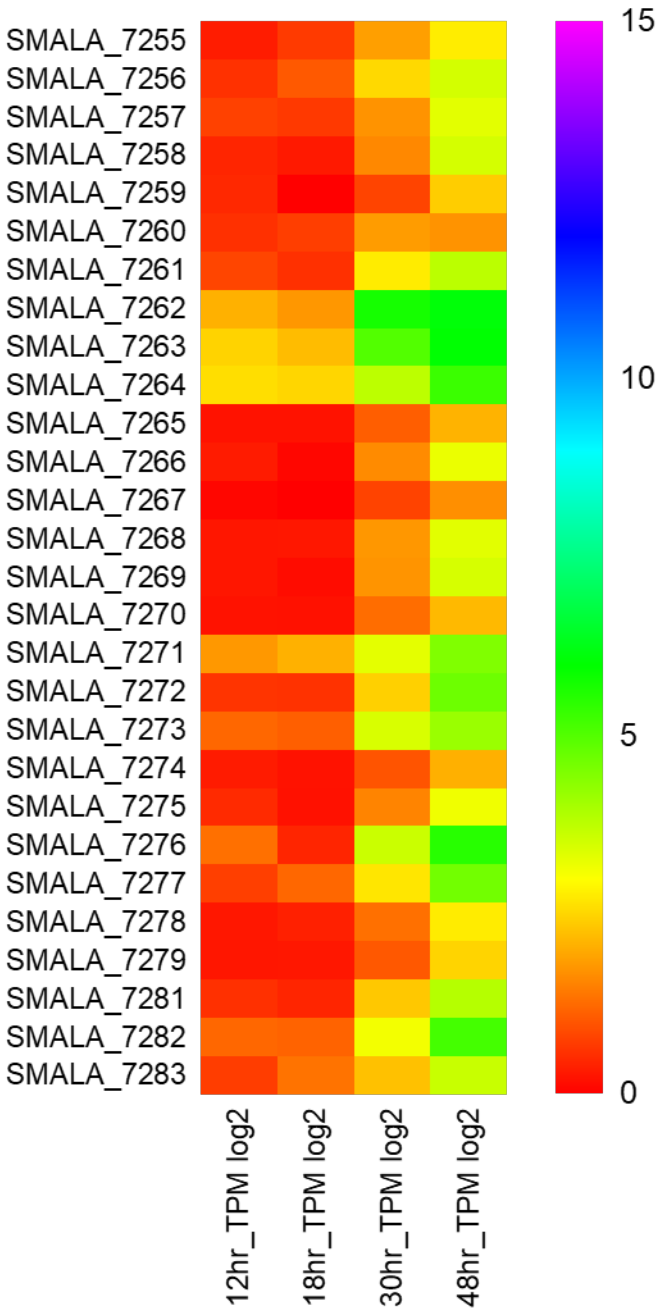


Figure 3.5. Heat map of RNAseq data for the nigericin biosynthetic gene cluster *S. malaysiensis* DSM4137*

Jensen, personal communication) that for *Salinispora* spp., there is an encouragingly good match for the great majority of specialised metabolites they tracked, between transcription level and the production of the natural product. Although transcription levels of these actinomycete biosynthetic clusters are not high relative to many genes of primary metabolism, it has proved possible to make a confident distinction between genes that are active and those that are not.

The regulation of most of the biosynthetic gene clusters in DSM4137 has never been studied, beyond noting the presence of various CSRs whose functional properties have been studied in other strains. In the next chapter, the transcriptomic analysis will be extended to further biosynthetic gene clusters in DSM4137, where the predicted products have never been detected in the fermentation of this strain. The compounds are either known from the analysis of closely similar clusters in other actinomycete strains, but not previously suspected of being produced by DSM4137, or they are predicted to be novel, and although their general properties can be predicted, their precise structure is unknown. For this reason, they may even be produced at low level but escaped detection. Alternatively, the relevant gene cluster may be entirely silent. For such clusters, transcriptomic analysis has the potential to distinguish between these possibilities. There will be even more emphasis on analysis of the presence, identity and behaviour of genes encoding CSRs, since pointers may be gained from the transcriptomic analysis of which regulatory genes it would be worthwhile to target in order to activate a 'silent' cluster.

Chapter 4. Transcriptome analysis of 'silent' biosynthetic gene clusters in *Streptomyces malaysiensis* DSM4137

4.1 Introduction

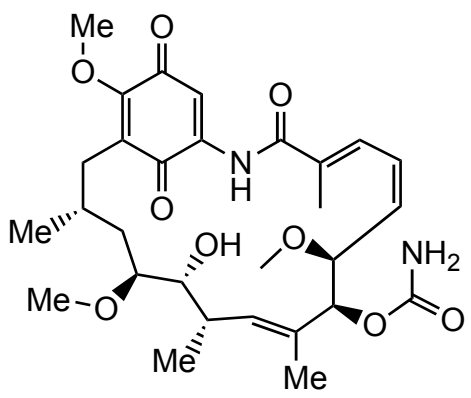
DSM4137 is an unusually prolific actinomycete in that it simultaneously produces a number of assembly-line polyketide natural products, but the analysis of its genome has confirmed that, as for other actinomycete strains, the great majority of the 53 biosynthetic clusters that it houses do not appear to produce an identifiable product when fermented under laboratory conditions. The work in this chapter seeks to use transcriptome analysis to discover whether at least some of these apparently silent clusters are transcribed under the conditions used; and in particular to monitor the expression level of putative CSRs in these clusters for clues as to how manipulation of such regulators might promote the activation of the cluster. This analysis is made easier by the fact that in some cases essentially identical clusters in other actinomycetes have already been shown to produce a specific product; or have been successfully engineered for (higher) expression.

4.2 Results and Discussion

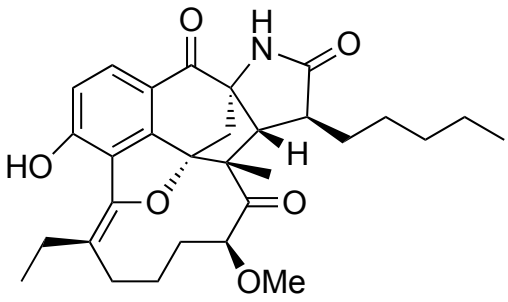
Figure 4.1 shows the structures of the diverse specialised metabolites that are the predicted products of some of the 'silent' clusters of DSM4137. They are drawn from the major groups of natural products discussed in Chapter 1.

4.2.1 Transcription of terpene biosynthetic gene clusters in DSM4137

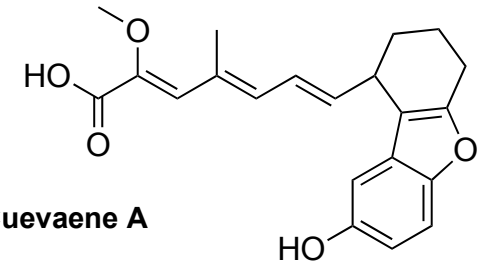
The DSM4137 genome contains several gene clusters predicted to encode the biosynthesis of terpenes. These include the widely-distributed clusters for biosynthesis of geosmin (cluster 21) and 2-methylisoborneol (2-MIB) (cluster 9a), which provide the characteristic odour of *Streptomyces* bacteria. These clusters and their products have been well-characterised, for example in *S. coelicolor* (Challis, 2014). *S. coelicolor* produces a mixture of hopene and aminotrihydroxybacterio-hopane (ATBH) (Poralla *et al*., 2000) and the cluster proposed to govern this pathway (Challis, 2014) is present in DSM4137 (cluster 29). The terpene synthase encoded in DSM4137 (cluster 7)



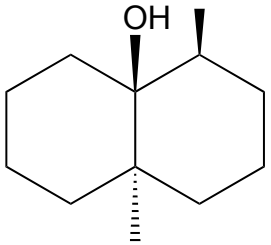
Geldanamycin



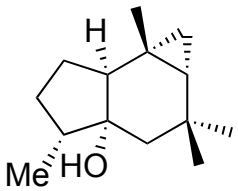
Neoansamycin D



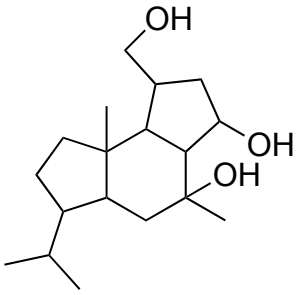
Cuevaene A



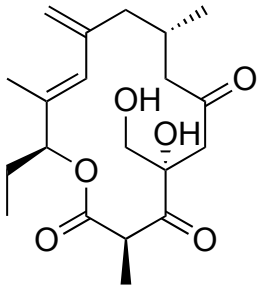
Geosmin



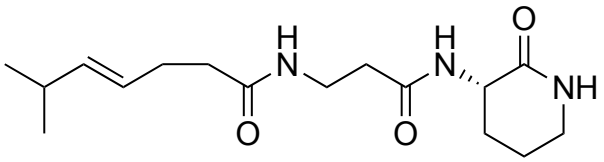
Isoafricanol



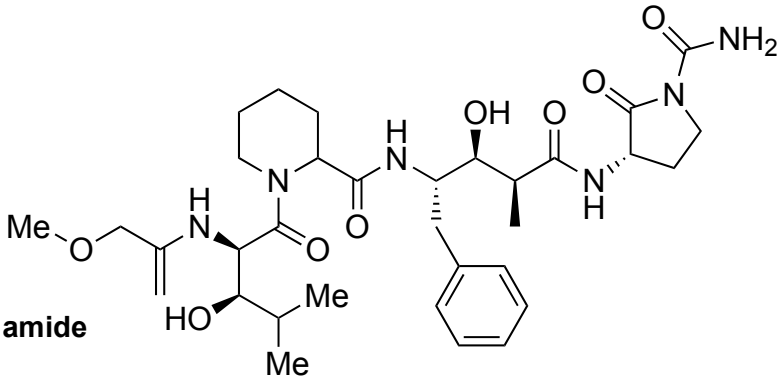
Cyclooctatin



Galbonolide E



Prajinamide



Padanamide

Figure 4.1 Predicted products of some DSM4137 'silent' clusters

closely resembles authentic (+)-isoafricanol synthases from other *Streptomyces* spp. including *S. violaceusniger* (Riclea *et al*., 2014). The volatile sesquiterpene alcohol isoafricanol has now been detected in headspace extracts of an agar-based culture of DSM4137, and its identity has been confirmed by extensive *in vitro* analysis (Rabe *et al*., 2017).

The heat maps for geosmin and 2-MIB clusters are shown in Figure 4.2. Clearly, the geosmin synthase is transcribed throughout the period analysed, while in contrast the isoafricanol synthase is transcribed later, accompanied by a rise in the transcription of a neighbouring CSR.

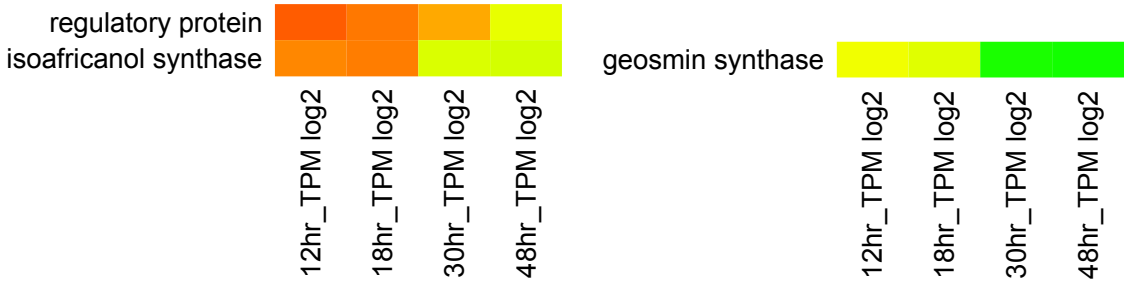


Figure 4.2. Heat map of RNAseq data for the iso-africanol and geosmin biosynthetic gene cluster *S. malaysiensis* DSM4137*

The unusual diterpene cyclooctatin **(Figure 4.1)** was first discovered as a product of *Streptomyces melanosporofaciens* (Aoyagi *et al*., 1992) and its biosynthetic gene cluster was then characterised in *Streptomyces* sp. LZ35 (Kim *et al*., 2009). Based on 16S rRNA sequence comparisons, LZ35 may be a *Streptomyces malaysiensis* strain (M. Samborskyy, unpublished). Cyclooctatin has been detected in extracts of a *Streptomyces* sp. LZ35 strain mutated to delete clusters governing four abundant specialised metabolites (Zhao *et al*., 2013). The cyclooctatin (*cot*) gene cluster of DSM4137, which is essentially identical to that of LZ35, is given with the transcription data in Table 4.1, and the corresponding heat map representation of the RNAseq data for the *cot* cluster is shown **b** From the appearance of the heat map it seems clear that the genes are all expressed from the earliest time point and throughout the period examined. These genes include a CSR (SMALA_0746) belonging to the TetR family. The TetR regulators are widely associated with antibiotic resistance and negative regulation which is relieved by ligand binding (Bibb, 2005), although in some cases

TetR regulators can show contrasting effects on different clusters (Jiang et al., 2017). Whatever the details of the regulation, there would seem a good chance that DSM4137 produces cyclooctatin on this medium.

Table 4.1. The results of transcriptional analysis of the cyclooctatin biosynthetic gene cluster in *S. malaysiensis* DSM4137

locus_tag	gene	Annotated name	Length	S1S2_12 hr_TPM	S1S2_18 hr_TPM	S1S2_30 hr_TPM	S1S2_48 hr_TPM
SMALA_0741	cotB1	geranylgeranyl pyrophosphate synthase	900	43.2	7.8	50	200.8
SMALA_0742	cotB2	terpene cyclase	924	159	52.55	167.3	666.6
SMALA_0743	cotB3	aveE - cytochrome P450 hydroxylase	1362	44.35	8.35	52.45	219.3
SMALA_0744	cotB4	ptlI - cytochrome P450	1362	23.4	7.2	26	69.8
SMALA_0745	cypA	cypA - cytochrome P450-like enzyme	1101	31.6	31.6	33.65	58.4
SMALA_0746		transcriptional regulator, TetR family	681	31.8	24.15	36.9	64.6
SMALA_0747		integral membrane transport protein	1428	15.15	8.95	11.6	63.4

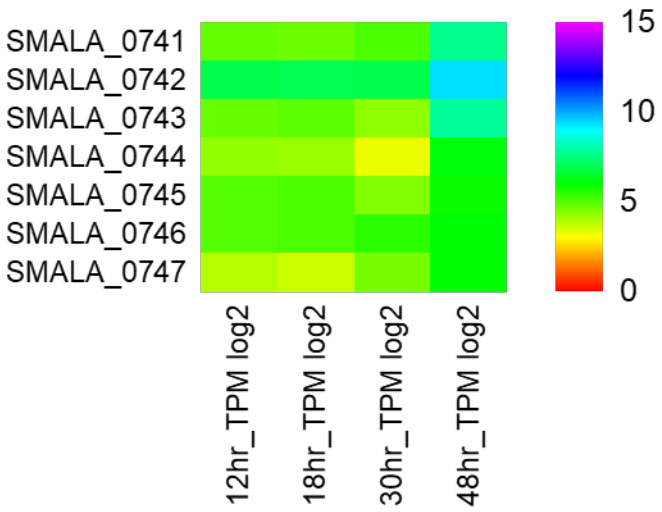


Figure 4.3. Heat map of RNAseq data for the cyclooctatin biosynthetic gene cluster *S. malaysiensis* DSM4137*

4.2.2 Transcription of ansamycin polyketide biosynthetic gene clusters in DSM4137

The genome of DSM4137 houses no fewer than four clusters potentially governing the biosynthesis of so-called "ansa" (bridged) polyketides: geldanamycin, cuevaene and neoansamycin (Figure 4.1), as well as hygrocin. The first three will be considered here. These are each the products of modular type I PKS systems in which the starter unit is derived from the shikimate pathway.

Geldanamycin is a 1,4-benzoquinone ansamycin, first discovered as a product of *Streptomyces hygroscopicus* (DeBoer *et al.*,1970). The geldanamycin biosynthetic gene cluster in DSM4137 (cluster 8) contains 20 genes, including three putative regulatory genes: two LuxR regulators (*gdmRI* and *gdmRIII)* and a TetR regulator (*gdmRIII*). Deletion of *gdmRIII* in the very similar cluster in the geldanamycin-producing *Streptomyces autolyticus* CGMCC0516 ablated production (Jiang *et al*., 2017).

Deletion of a PKS gene in the geldanamycin gene cluster of LZ35 proved a successful tactic to allow the detection of the triene cuevaene A (Jiang *et al*., 2013). There is an essentially identical cluster in DSM4137 (cluster 10) to the LZ35 cuevaene cluster, allowing the DSM4137 cluster to be identified.

A further putative ansamycin cluster in the *S. sp*. LZ35 genome (Li et al. 2015) was identified by engineering its upregulation: biosynthetic analysis of the cluster revealed a LuxR regulator, a regulator that is frequently a positive regulator for polyketide biosynthesis (Laureti et al., 2011). Overexpression of this LuxR regulator under the control of the constitutive permE* promoter activated the cluster and allowed the isolation of three novel ansamycins; neoansamysins A, B and C (Li et al. 2015). Again, there is an essentially identical cluster in DSM4137 (cluster 11) to the LZ35 neoansamycin cluster, allowing this DSM4137 cluster to be identified.

The detailed annotation of the geldanamycin cluster is given in Table 4.2, alongside the RNAseq data. On the following page is shown Figure 4.4, the heat map representation of the data of Table 4.2. The data suggest very strongly that the pathway genes of the geldanamycin cluster are well expressed at later time points, with two of the three regulatory genes also transcribed at early time points and the product of SMALA_0866 (GdmRI) in contrast only expressed at the same time as the pathway genes. These results certainly encourage a more focussed search for this compound in DSM4137 extracts.

Table 4.2 The results of transcriptional analysis of the geldanamycin biosynthetic gene cluster in *S. malaysiensis* DSM4137*

locus_tag	gene	Annotated name	Length	S1S2_12 hr_TPM	S1S2_18 hr_TPM	S1S2_30 hr_TPM	S1S2_48 hr_TPM
SMALA_0854	gdmA1	gdmA1 - type I modular polyketide synthase	20571	0.25	0.15	0.15	5.8
SMALA_0855	gdmA2	pteA1 - modular polyketide synthase	10248	0.25	0.3	0.4	7.6
SMALA_0856	gdmA3	mlsA1 - type I modular polyketide synthase	11610	0.15	0.05	0.05	4.5
SMALA_0857	gdmF	N-acetyltransferase	774	1.05	0.4	0.05	14.7
SMALA_0858	gdmM	mhpA - 3-_3-hydroxyphenyl_propionate hydroxylase	1623	0.4	0.2	0.25	4.4
SMALA_0859		O-linked N-acetylglucosamine transferase	297	0	0.25	0.25	0.6
SMALA_0860	gdmN	carbamoyltransferase	2058	1.05	0.6	1.4	15.2
SMALA_0861	gdmH	FkbH like protein	1095	0.7	0.05	0.6	6.3
SMALA_0862	gdmI	acyl-CoA dehydrogenase domain	1113	0.2	0.1	0.25	2.7
SMALA_0863		gdmJ	276	1.85	1.25	1.7	27.8
SMALA_0864	gdmK	3-hydroxyacyl-CoA dehydrogenase NAD-binding	867	0.7	0.7	1.7	13.9
SMALA_0865	gdmG	caffeoyl-CoA O-methyltransferase	657	1	0.8	1.55	15.7
SMALA_0866	gdmRI	gdmRII - LuxR family transcriptional regulator	2619	0.25	0.25	0.2	2
SMALA_0867	gdmO	shnQ - 3-dehydroquinate	1065	0.25	0.3	0.7	5.9
SMALA_0868	gdmfdx	fdxB - ferredoxin	195	0.5	0.7	1	9
SMALA_0869	gdmP	cyp7 - cytochrome P450 hydroxylase	1203	1.6	0.5	1.05	14.5
SMALA_0870	gdmRII	gdmRII - LuxR family transcriptional regulator	2592	26.1	4.8	26.75	130.6
SMALA_0871		hypothetical protein	468	6.15	3.5	2.45	29.5
SMALA_0872		hydrolase	714	7.85	3.7	6	12.5
SMALA_0873		TetR family transcriptional regulator	627	6.3	4.35	6.35	7.7

*The PKS genes are shaded in blue and the three putative regulatory genes in yellow.

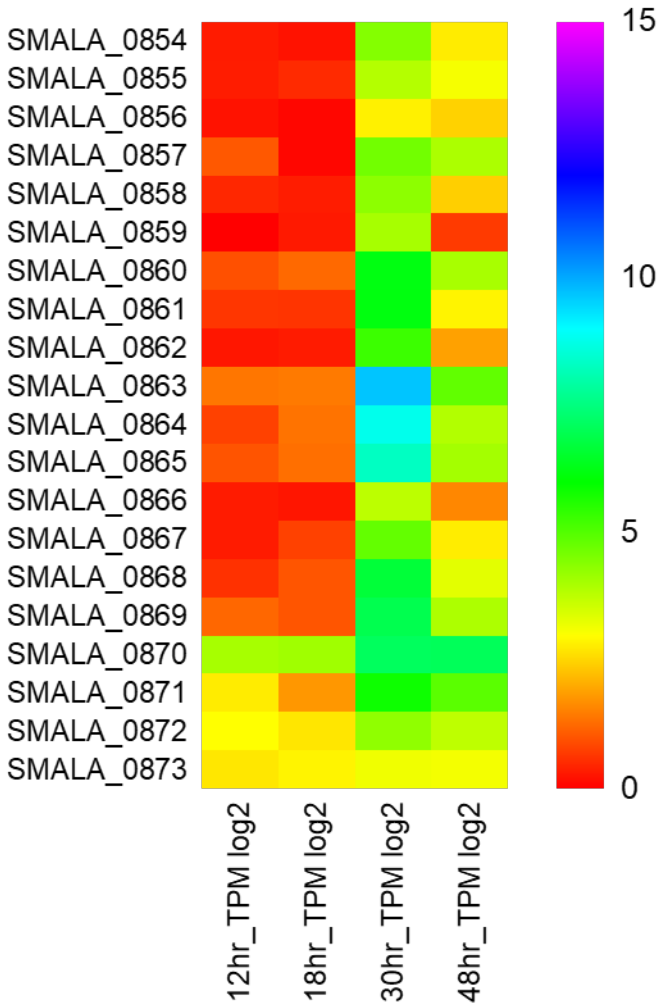


Figure 4.4 Heat map of RNAseq data for the geldanamycin biosynthetic gene cluster *S. malaysiensis* DSM4137*

The detailed annotation of the DSM4137 cuevaene cluster is given in Table 4.2, alongside the RNAseq data. On the following page is shown Figure 4.5, the heat map representation of the data of Table 4.2. For the fermentation period under consideration, it is clear that certain key PKS genes are poorly transcribed. However, more striking still is the status of the two likely positive CSRs and of the two likely negatively regulating CSRs. First, the likely activating LuxR-encoding gene (SMALA_1281) is not transcribed; secondly, the likely activating SARP-encoding PadR gene (SMALA_94) is mostly not transcribed; and (perhaps decisively) both of the likely negatively-acting TetR-encoding genes (SMALA_1287 and SMALA_1296) are both being transcribed.

Table 4.3 The results of transcriptional analysis of the cuevaene biosynthetic gene cluster in *S. malaysiensis* DSM4137*

locus_tag	gene	Annotated name	Length	S1S2_12 hr_TPM	S1S2_18 hr_TPM	S1S2_30 hr_TPM	S1S2_48 hr_TPM
SMALA_1266		dihydroorotate oxidase B, electron transfer subunit	1050	1.65	2.85	2.4	5.8
SMALA_1267	cuv1	pqrB - putative paraquat_multidrug resistance efflux pump	1536	3.05	2.15	2.75	8.3
SMALA_1268	cuv2	TetR family transcriptional regulator	555	0.95	1.4	2.1	2.2
SMALA_1269	cuv3	glyceryl transferase	1092	1.45	1.25	2	15.4
SMALA_1270	cuv4	acyl-CoA dehydrogenase	1098	0.2	0.35	0.35	5.7
SMALA_1271	cuv5	acyl-CoA dehydrogenase	264	0.3	0.55	0.4	17.4
SMALA_1272	cuv6	3-hydroxyacyl_Coa dehydrogenase	843	0.2	0.1	0.1	5.8
SMALA_1273	cuv7	SAM-dependent methyltransferase	684	0.4	0.6	0.3	3.5
SMALA_1274	cuv8	drug resistance transporter	1398	4.1	5.15	5.2	65.4
SMALA_1275	cuv9	domain of unknown function DUF1857	480	0.7	0.15	0.55	25.4
SMALA_1276	cuv10	FkbO	957	1.6	1.6	1.9	22.6
SMALA_1277	cuv11	loading module and extension modules 1 and 2	14145	0.4	0.35	0.4	10.2
SMALA_1278		PKS	6282	0.65	0.6	0.85	15.2
SMALA_1279	cuv13	mlsB - type I modular polyketide synthase	10275	0.8	0.5	0.6	9
SMALA_1280	cuv14	pteA1 - modular polyketide synthase	5835	0.35	0.35	0.55	7.4
SMALA_1281	cuv15	gdmRII - LuxR family transcriptional regulator	2550	0.15	0.1	0.1	1
SMALA_1282	cuv16	aroH - phospho-2-dehydro-3-deoxyheptonate aldolase	1179	0.15	0.15	0.15	2.6
SMALA_1283	cuv17	nucleotide-diphosphate-sugar epimerase	843	0.35	0.05	0.25	1.8
SMALA_1284	cuv18	flavin-dependent monooxygenase	1563	1.1	0.65	1.25	5.3
SMALA_1285	cuv19	ABC transporter related	402	14.9	10.55	17.15	44.9
SMALA_1286	cuv20	cuv20	2643	3.55	2.8	3.95	10.8
SMALA_1287	cuv21	TetR family transcriptional regulator	141	8	7.6	9.35	24
SMALA_1288	cuv22	aveA4 - type I polyketide synthase AVES 4	1212	0.05	0	0.2	0.7
SMALA_1289	cuv23	putative membrane-associated phosphoesterase	1146	4.1	2.15	3.75	13.1
SMALA_1290	cuv24	zinc-binding alcohol dehydrogenase family protein	996	4.95	4.2	6.3	21.1
SMALA_1291	cuv25	hypothetical protein	603	2.9	4.05	2.65	16.6
SMALA_1292	cuv26	hypothetical protein	612	58.75	61.75	66.7	137.4
SMALA_1293	cuv27	retinal pigment epithelial membrane protein	1248	11.25	10.2	9.4	30.4
SMALA_1294	cuv28	PadR-like family transcriptional regulator	519	1.85	1.65	3.05	14.9
SMALA_1295	cuv29	P450	1248	1.95	0.85	2.6	6.3
SMALA_1296	cuv30	TetR family transcriptional regulator	684	16.65	12.55	17.15	24.7

*The PKS genes are shaded in blue and the four putative regulatory genes in yellow

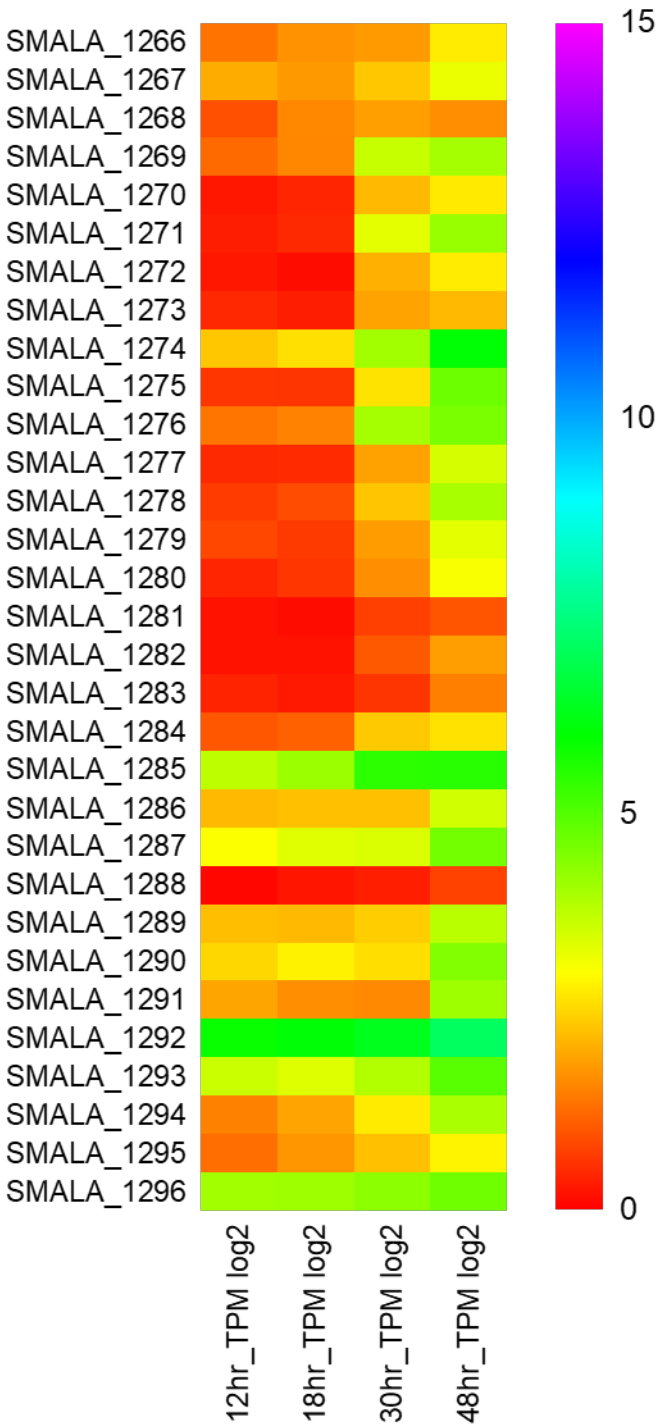


Figure 4.5 Heat map of RNAseq data for the cuevaene biosynthetic gene cluster *S. malaysiensis* DSM4137*

The detailed annotation of the DSM4137 neoansamycin cluster is given in Table 4.3, alongside the RNAseq data. On the following page is shown Figure 4.6, the heat map representation of the data of Table 4.3. In contrast to the data for cuevaene A, these data encourage the view that neoansamycin is produced in DSM4137. Notably, the activating CSR (SARP, SMALA_1369) is strongly transcribed later on.

Table 4.4 The results of transcriptional analysis of the neoansamycin biosynthetic gene cluster in *S. malaysiensis* DSM4137*

locus_tag	gene	Annotated name	Length	S1S2_12 hr_TPM	S1S2_18 hr_TPM	S1S2_30 hr_TPM	S1S2_48 hr_TPM
SMALA_1341		LysR family trancsriptional regulator	912	12.3	13.9	14.25	52.9
SMALA_1342		NADP oxidoreductase coenzyme F420-dependent	630	1.65	1.6	2.1	16.5
SMALA_1343		NAD-dependent epimerase_dehydratase	933	1.2	1.1	1.1	10.8
SMALA_1344		NADP oxidoreductase coenzyme F420-dependent	669	1.35	2.7	3	18.6
SMALA_1345		namR2	642	13.8	8.2	13.2	40.9
SMALA_1346		namR1	417	50.3	23.7	45.6	85.2
SMALA_1347		FabH nam10	996	0.2	0.3	0.6	2.6
SMALA_1348		nam9	1767	1.4	0.3	1.4	8.7
SMALA_1349	nam8	ccr nam8	1350	1.9	0.3	1.6	17
SMALA_1350	namJ	aroD - 3-dehydroquinate dehydratase	444	1.55	1.4	3.1	12.9
SMALA_1351	nam7	nam7 - 3-_3-hydroxyphenyl_propionate hydroxylase	1653	0.4	0.55	0.7	7.4
SMALA_1352	nam6	nam6 - cyclohexanecarboxylate-CoA ligase	1698	1.3	0.95	1.75	6.1
SMALA_1353		AHBA kinase	1008	0.2	0.05	0.3	1.6
SMALA_1354	namM	namM - phosphatase-like protein	585	1.95	1.5	2.15	13.3
SMALA_1355		oxidoreductase namL	1101	1.65	0.9	1.1	8.1
SMALA_1356		namK AHBA synthase	1284	2.85	2.15	2.65	15.7
SMALA_1357		hypothetical protein	93	2.55	3.3	6.65	24.6
SMALA_1358	namG	NamG 3-dehydroquinate synthase	1008	1.4	1.7	0.95	7.3
SMALA_1359		NamF AHBA synthase	834	14.05	15.3	14.5	57.1
SMALA_1360		namE	5289	2.5	2.7	3	15
SMALA_1361		namD	5490	2.5	2.15	2.25	16.6
SMALA_1362		namC	6495	2.85	2.4	2.75	22.4
SMALA_1363		namB	6498	2.35	1.95	2.4	22.2
SMALA_1364	namA	namA - type I modular polyketide synthase	16977	1.2	0.85	1.5	11.5
SMALA_1365		hypothetical protein	642	0.45	0	0.3	8.4
SMALA_1366		acyl-CoA dehydrogenase domain protein	117	1.2	0	0	3.8
SMALA_1367		putative methoxymalonate biosynthesis protein	201	2.6	0.85	0.8	20.3
SMALA_1368	ccrA2	ccrA2 - crotonyl-CoA reductase	1365	6.45	4.5	8.6	30.2
SMALA_1369		transcriptional regulator, LuxR family	1086	2.8	1	2.4	15.3

*The PKS genes are shaded blue and the four putative regulatory genes yellow.

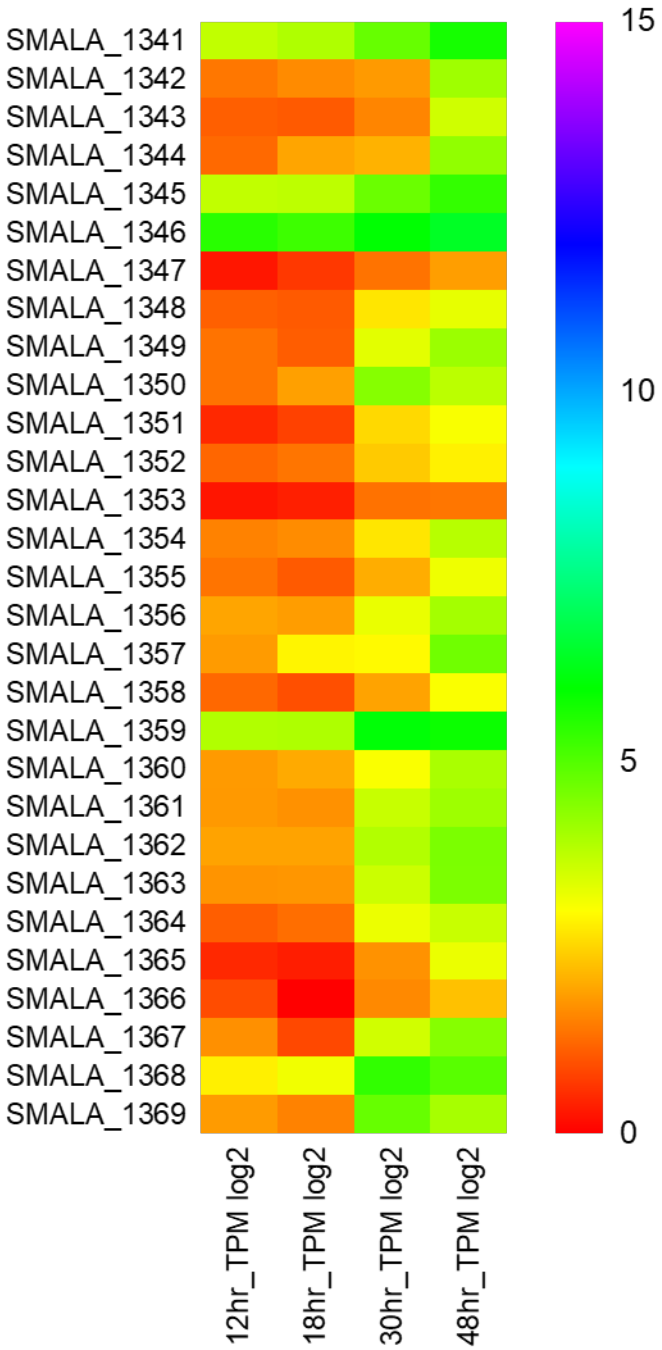


Figure 4.6 Heat map of RNAseq data for the neoansamycin biosynthetic gene cluster *S. malaysiensis* DSM4137*

4.2.3 Transcription of a hybrid fatty acid synthase-polyketide (galbonolide) biosynthetic gene cluster in DSM4137

Galbonolides (Figure 4.1) are non-glycosylated, rather unstable 14-membered macrocyclic polyketides, the first of which was isolated from *Micromonospora narashinoensis* as rustmicin, a potent antifungal against wheat stem rust fungus (Takatsu *et al*., 1985) and independently in *Streptomyces galbus* subsp. *eurythermus* (Achenbach *et al*., 1988). It is effective against the opportunistic fungal pathogen *Cryptococcus neoformans* inositol phosphoceramide synthase (IC_{50} 70 pM) (Mandala *et al*., 1998), but showed disappointing *in vivo* activity. The assembly of the macrocycle is catalysed by a wholly novel combination of fatty acid synthase and PKS activities working iteratively (Kim *et al*., 2014). The structural difference between galbonolides A, B and E is that the fourth extender unit to be added is either from methoxymalonyl-CoA (A), methylmalonyl-CoA (B) or malonyl-CoA (E). Galbonolides B and E have also been isolated from the DSM4137-related strain *Streptomyces* sp. LZ35 (Wang *et al*., 2013) and the gene cluster from this strain has been characterised (Liu *et al*., 2015a; Liu *et al*., 2015b). The galbonolide cluster in DSM4137 is identical to that in LZ35, and the detailed annotation is given in Table 4.5, alongside the RNAseq data. Figure 4.7 shows the heat map representation of the data of Table 4.5.

Table 4.5 The results of transcriptional analysis of the galbonolide biosynthetic gene cluster in *S. malaysiensis* DSM4137

locus_tag	gene	Annotated name	Length	S1S2_12 hr_TPM	S1S2_18 hr_TPM	S1S2_30 hr_TPM	S1S2_48 hr_TPM
SMALA_1173	gbnA	acyl transferase	1221	0.3	0.2	0.2	5.1
SMALA_1174	gbnB	galbonolide PKS KRI-KSI-KSII-KRII-DH	6084	0.2	0.2	0.3	3.7
SMALA_1175	gbnC	pks9-2 - 3-oxoacyl-ACP synthase III	1026	0.05	0.4	0.4	4.3
SMALA_1176	gbnD	cytochrome P450	1455	0.05	0	0.1	5.8
SMALA_1177	gbnE	dioxygenase	1110	0.15	0.15	0.05	3.8
SMALA_1178		3-hydroxyacyl-CoA dehydrogenase NAD-binding	396	0.2	0.2	0	1.6
SMALA_1179		putative methoxymalonate biosynthesis protein	291	1.8	0.85	0.85	9.7

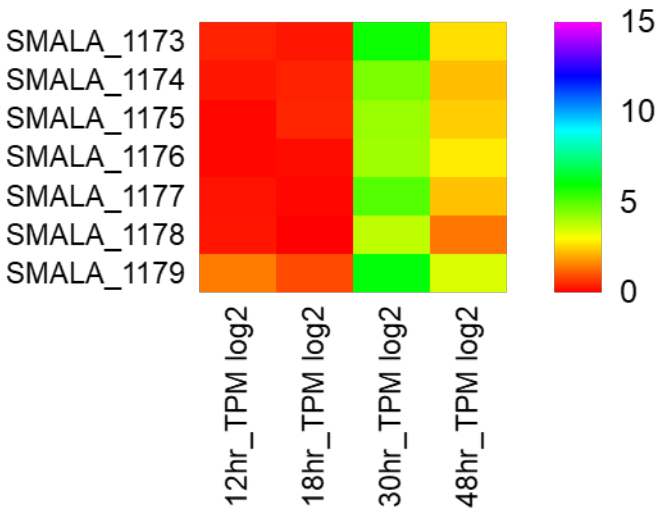


Figure 4.7. Heat map of RNAseq data for the galbonolide biosynthetic gene cluster *S. malaysiensis* DSM4137*

The data of Figure 4.7 are unequivocal in predicting that galbonolide is indeed produced by DSM4137, at least after 30 hours of fermentation.

4.2.3 Transcription of a novel hybrid PKS-NRPS (catramycin) biosynthetic gene cluster in DSM4137

Catramycin is a hypothetical member of a rare and valuable chemotype, the macrocyclic inhibitors of signal transduction mediated by FKBP12 and related proteins. Specialised metabolites in this class include the immunosuppressants FK506 and rapamycin. In this laboratory, attempts have been made to detect the predicted product, including by affinity chromatography on FKBP-Sepharose, without success (Yurkovich, 2016). The cluster was cloned from a PAC library and heterologously expressed; and the gene in the cluster (SMALA_8319) encoding a potential LuxR activator was placed under a constitutive promoter in both wild type and heterologous host strains, again without success (Yurkovich, 2016). the detailed annotation is given in Table 4.6, alongside the RNAseq data. Figure 4.8 shows the heat map representation of the data of Table 4.6.

Intriguingly, although the majority of the annotated cluster genes are not transcribed early, all of these genes are transcribed, including the putative SARP regulator, at later time points. There is still a chance that the predicted product is made, under the appropriate conditions.

Clearly, without a specific structure or activity to search for, genome mining is not a straightforward activity. There remains, too, the possibility that the 'catramycin' cluster is a remnant from a larger cluster that has recently undergone a recombination event that removed essential genes, and despite the positive results from the transcriptional analysis, the cluster is non-functional.

Table 4.6 The results of transcriptional analysis of the 'catramycin' biosynthetic gene cluster in *S. malaysiensis* DSM4137*

locus_tag	gene	Annotated name	Length	S1S2_12 hr_TPM	S1S2_18 hr_TPM	S1S2_30 hr_TPM	S1S2_48 hr_TPM
SMALA_8312		cytochrome P450	1182	1.4	1.05	1.4	19
SMALA_8313		cytochrome P450	1167	1.15	0.35	0.8	17.8
SMALA_8314		fkbO-like	1056	0.35	0.25	0.4	7.3
SMALA_8315	ocd	arcB - ornithine cyclodeaminase	1038	0.3	0.2	0.2	5.2
SMALA_8316	mlsA1	mlsA1 - type I modular polyketide synthase	24405	0.55	0.4	0.4	9.3
SMALA_8317		non-ribosomal peptide synthetase	4530	0.5	0.2	0.2	7.5
SMALA_8318	mlsB	mlsB - type I modular polyketide synthase	23769	0.4	0.35	0.45	9.9
SMALA_8319	gdmRII	gdmRII - LuxR family transcriptional regulator	2721	5.6	2.15	4.5	23
SMALA_8320		gdmRII - LuxR family transcriptional regulator	447	7.75	7.75	8.45	29.9

*The PKS genes are shaded blue and the four putative regulatory genes yellow.

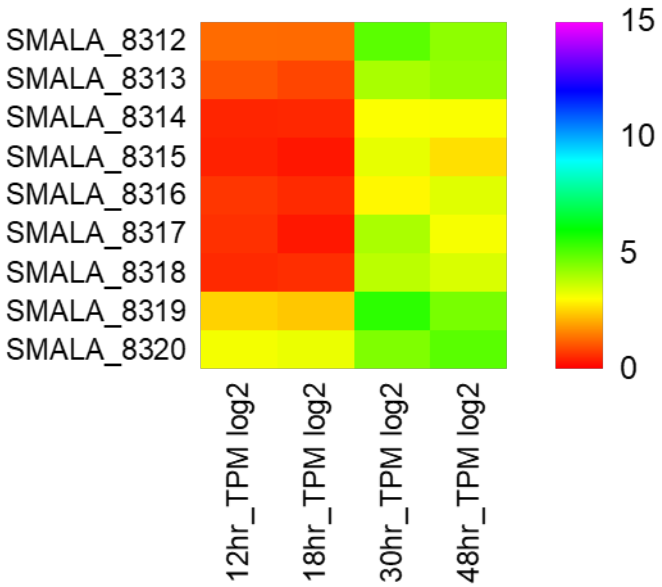


Figure 4.8. RNA-seq heatmap representation of RNAseq data for the 'catramycin' cluster of DSM4137.

4.2.4 Transcription of biosynthetic gene clusters for unusual peptides in DSM4137

Padanamides such as padanamide A (Figure 4.1) are modified linear peptides constructed from non-proteinogenic amino acids and a ketide unit (Du *et al*., 2013), isolated from *Streptomyces sp.* RJA2928, a marine isolate from the Pacific Ocean (Williams *et al.*, 2011). The biosynthetic gene cluster found in this marine isolate (Du *et al*., 2013) is also found, in near-identical form, in both DSM4137 (Table 3.1) and DSM14702 (Table 3.2, cluster 1b). Neither strain has been examined for padanamide production. Table 4.7 shows the annotation of the cluster alongside the transcriptional data. There are two potential CSRs: SMALA_0086 encodes a PaiB homologue, a negative transcriptional regulator of *Bacillus subtilis* sporulation (PadO), while SMALA_0090 encodes a SARP regulator (PadR) (Du *et al*. 2013). In fact, in a ∆PadO mutant padanamide biosynthesis was abolished (Du *et al*., 2013) so it may be an activator.

Figure 4.9 shows the heat map representation of the data of Table 4.7. The SARP is transcribed, and increasingly so as time progresses. SMALA_0086 also is transcribed at later time points, but the level of transcription of the NRPS genes remains low. It is an intriguing possibility that padanamide might be produced in DSM4137 (perhaps helped by overexpression of one or both activators) but there are caveats: first, not all the genes needed for padanamide biosynthesis are found in the cluster of the authentic padanamide producer. For the supply of the unusual building block L-2,4-diaminobutyrate (L-DAB), it is suggested (Du *et al*., 2013) that EctB, an enzyme in the pathway to ectoine (a molecule that serves to counteract high osmolarity in cells (Louis and Galinski, 1997) is co-opted to convert convert L-aspartic acid semialdehyde into L-DAB. Further, not all the enzymes required for synthesis of methoxymalonyl-ACP as the proposed starter unit for assembly of padanamide are in the cluster in RJA2928 (Du *et al*., 2013). Such uncertainties increase the risks in attempting heterologous expression, but are arguably of less importance for attempts to detect padanamide production in DSM4137.

Table 4.7 The results of transcriptional analysis of the padanamide biosynthetic gene cluster in *S. malaysiensis* DSM4137*

locus_tag	gene	Annotated name	Length	S1S2_12 hr_TPM	S1S2_18 hr_TPM	S1S2_30 hr_TPM	S1S2_48 hr_TPM
SMALA_0074	padH	dehydrogenase, E1 component, alpha	951	0.35	0.25	0.15	10.4
SMALA_0075	padI	dehydrogenase, E1 component, beta	1086	0	0.05	0.05	11.7
SMALA_0076	padJ	dehydrogenase complex E2 component	1230	0	0	0	3.6
SMALA_0077	padK	FkbH like protein	1704	0	0	0	2
SMALA_0078	padC	NRPS: C-A-T	3324	0.05	0	0	1.2
SMALA_0079	padD	PKS: KS-AT-KR-T	4617	0.05	0	0.05	3.1
SMALA_0080	padA1	NRPS: T-C-A	3351	0	0	0	2.3
SMALA_0081	padB	NRPS: C-A-T	3411	0.05	0.05	0.1	3.3
SMALA_0082	padA2	T-E didomain protein	1821	0	0	0	1.6
SMALA_0083	padM	O-methyltransferase	936	0.2	0.2	0.35	9
SMALA_0084	padL	MbtH domain protein	222	0	0	1.2	17.7
SMALA_0085	padN	ornithine N-hydroxylase	1347	0.7	0.3	0.75	32.1
SMALA_0086	padO	regulator	672	0.1	0.1	0.1	17.1
SMALA_0087	padP	hydroxylase	837	0.45	0	0	17.3
SMALA_0088	padE	NRPS: C-A-T-TE	4062	0.1	0	0.1	7.4
SMALA_0089	padQ	Asp/OTC	957	1.65	0.7	1.05	30.6
SMALA_0090	padR	SARP family regulator	1794	6	2.65	6.85	39.1

Transcriptomic analysis has also been used to examine a set of genes for a second unusual peptide, prajinamide (Figure 4.1). The locus is not specifically highlighted by AntiSMASH 4.0 but it clearly contains NRPS genes, flanking the putative catramycin PKS-NRPS cluster in the region 10460347-10482686 of the DSM4137 chromosome (SMALA_8320-8335).

The recently-published draft genome sequence of *Streptomyces* sp. SPMA113 (Komaki *et al*., 2016) is reported to encode several clusters that govern synthesis of *inter alia* elaiophylin, nigericin, galbonolide, laspartomycin, padanamide, geldanamycin, meridamycin and bafilomycin. Since all these clusters are present in DSM4137 and/or DSM14702, it was therefore of interest that SPMA113, collected from soil in Prajinburi Province, Thailand, is a known producer of the novel modified peptide prajinamide (Igarashi *et al*., 2012). The authors determined at that time, based on 16S rRNA comparisons, that SPMA113 is most closely related to the type strain of *S. malaysiensis* (ATB-11^{T}), but the biosynthetic gene cluster for prajinamide has never been reported.

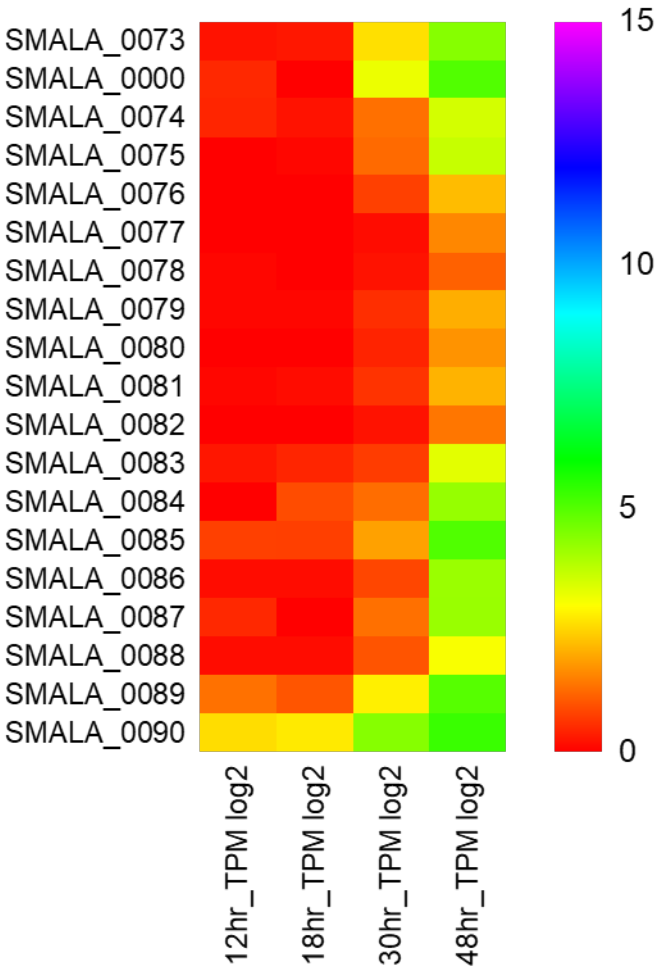


Figure 4.9. RNA-seq heatmap representation of RNAseq data for the padanamide cluster of DSM4137.

The prajinamide structure suggests that its biosynthesis might proceed by an NRPS assembly pathway, and utilise three simple building blocks: an unsaturated fatty acid (6-methyl (2Z,4E)-hepta-2,4-dienoic acid); β-alanine; and L-ornithine lactam. Scrutiny of the DSM4137 genome for an orphan NRPS cluster that might eventually be linked to this metabolite, only one cluster was small enough to be considered - SMALA_8320-8335. BLAST analysis showed that all these ORFs have close homologues (98-100% identity) in the prajinamide-producing strain, as well as in *Streptomyces* MnatMP-M27, a soil isolate sequenced at the USA DoE; and in *Streptomyces antioxydans*. In contrast, these ORFs are not present in DSM14702.

The putative cluster contains an NRPS gene encoding an enzyme with the domain arrangement C-A-T (SMALA_8326), preceded by a small gene for a discrete ACP or PCP (SMALA_8325) while two neighbouring genes encode stand-alone NRPS-related activities: SMALA_8328 an A domain, and SMALA_8329 a C domain. The adenylation specificity of the A-domains in SMALA_8326 is not clearly predicted by the algorithms used by AntiSMASH 4.0. However, the 10-residue active site motif (Stachelhaus *et al*., 1999; Challis *et al*., 2000) for

this A-domain bears a modest resemblance to the motif for the authentic bacterial A-domains known to be specific for β-alanine in, respectively NRPSs for tallysomicin Tao et al., 2007), bleomycin (Shen et al., 2002), and zorbamycin (Galm et al., 2009) (Table 4.8)

Table 4.8 The specificity-conferring code of the SMALA_8326 A domain compared to those for authentic β-alanine-activating A domains

Domain	235	236	239	278	299	301	322	330	331	517
Blm NRPS-2	V	D	W	V	V	**S**	**L**	**A**	D	**K**
Tlm NRPS-2	V	D	W	V	I	**S**	**L**	**A**	D	**K**
Zbm NRPS-2	V	D	A	L	V	**S**	**L**	**A**	D	**K**
DSM4137	D	V	Q	F	C	**S**	**L**	**A**	A	**K**
SMALA_8326										

The flanking genes SMALA_8321-8323 encode three enzymes of the pathway from L-glutamic acid to L-ornithine, and this pathway might furnish the precursor for the L-ornitholactam moiety of prajinamide. There are few clues as to the possible origin of the unusual unsaturated fatty acid moiety, but the A domain of SMALA_8328 together with the discrete PCP A_8325 might be involved in activating and loading this as a starter unit onto the NRPS. The stand-alone C domain of SMALA_8329 might act to release the peptide from the NRPS and the amide synthetase annotated as an asparagine synthetase (SMALA_8331) might act to close the ornitholactam ring. Clearly, functional studies will be required in order to test these speculations. Meanwhile it was of interest to examine the transcription of these genes. Table 4.11 shows the annotation of the putative cluster along with the transcription data, and the corresponding heat map is shown in Figure 4.10.

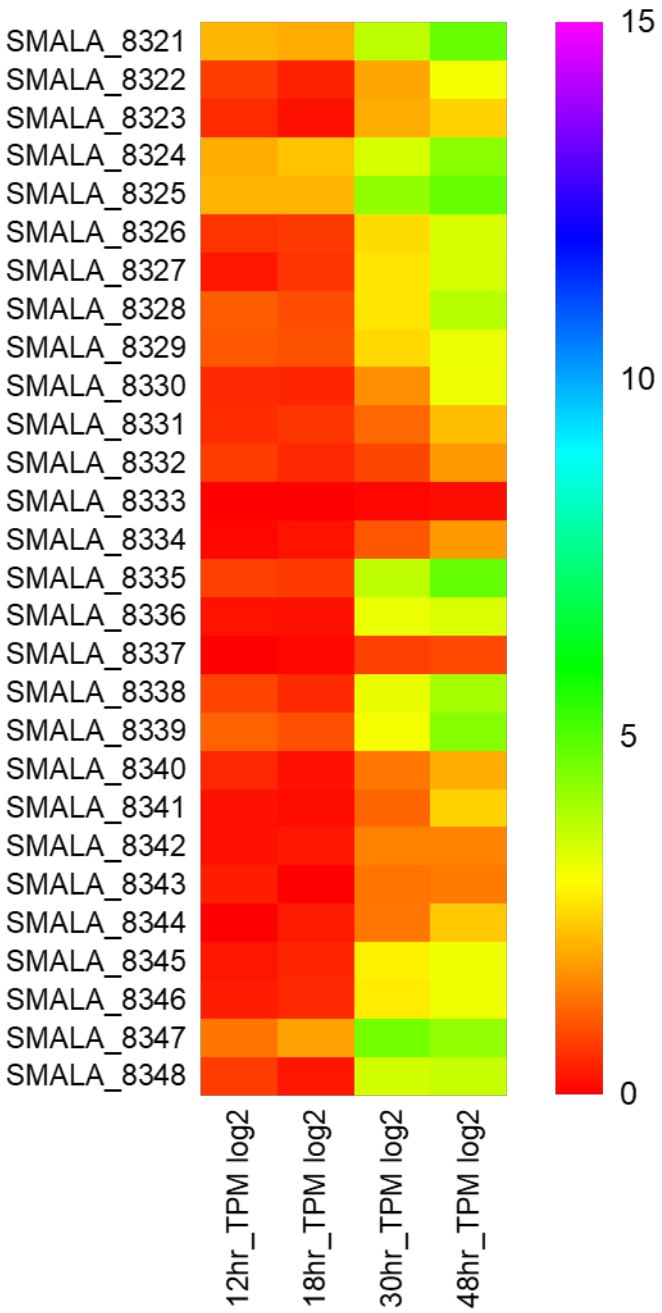


Figure 4.10 Heat map of RNAseq data for the putative prajanamide biosynthetic gene cluster *S. malaysiensis* DSM4137*

Table 4.11 The results of transcriptional analysis of the putative prajinamide biosynthetic gene cluster in *S. malaysiensis* DSM4137*

locus_tag	gene	Annotated name	Length	S1S2_12 hr_TPM	S1S2_18 hr_TPM	S1S2_30 hr_TPM	S1S2_48 hr_TPM
SMALA_8321	avtA4	avtA4 - PLP-dependent aminotransferase	1245	3.4	3.55	3.05	27
SMALA_8322	proA	proA - gamma-glutamyl phosphate reductase	1251	0.65	0.35	0.3	7.8
SMALA_8323	proB	proB - gamma-glutamyl kinase	942	0.4	0.35	0.15	4.5
SMALA_8324		putative transporter	1200	3.1	2.25	4.45	19.9
SMALA_8325		hypothetical protein	252	3.4	3	3.35	27
SMALA_8326		nrps8326	3243	0.55	0.35	0.6	10
SMALA_8327		short-chain dehydrogenase_reductase SDR	780	0.2	0.65	0.65	10.3
SMALA_8328		nrps8328	1575	1.15	0.9	0.85	13.3
SMALA_8329		nrps8329	1266	1.1	0.8	0.95	8.4
SMALA_8330		SARP family transcriptional regulator	2082	0.4	0.35	0.35	8.2
SMALA_8331	asnB-1	asnB-1 - asparagine synthase _glutamine-hydrolyzing_	1845	0.45	0.5	0.55	3.7
SMALA_8332		flavin-nucleotide-binding protein	597	0.65	0.65	0.4	2.5
SMALA_8333		hypothetical protein	1965	0	0	0	0.1
SMALA_8334	metG	metG - methionyl-tRNA synthetase	1590	0.05	0.05	0.15	2.5
SMALA_8335		cupin 2 conserved barrel domain protein	378	0.9	0.75	0.75	27.5
SMALA_8336		L-lysine 6-monooxygenase _NADPH_	1257	0.15	0.25	0.15	9.9
SMALA_8337	dadA8	dadA8 - D-amino-acid dehydrogenase	1083	0	0.05	0.05	0.8
SMALA_8338		dihydrodipicolinate synthase	939	0.8	0.45	0.4	15.5
SMALA_8339		hypothetical protein	390	1.45	0.65	0.9	20.4
SMALA_8340		pyruvate carboxyltransferase	1026	0.4	0.05	0.15	3.1
SMALA_8341		haloacid dehalogenase-like hydrolase	753	0.15	0.1	0.1	4.5
SMALA_8342		short-chain dehydrogenase_reductase SDR	771	0.15	0.1	0.2	1.9
SMALA_8343		short-chain dehydrogenase_reductase SDR	711	0.3	0	0	1.7
SMALA_8344		hypothetical protein	951	0	0.05	0.25	4.1
SMALA_8345		short-chain dehydrogenase_reductase SDR	723	0.2	0.1	0.35	8.1
SMALA_8346		2-hydroxy-3-oxo-5-methylthiopent-2-enoate oxidase	585	0.25	0.25	0.4	8
SMALA_8347		LuxR family transcriptional regulator	798	2.85	1.2	2.7	18.1
SMALA_8348		LuxR family transcriptional regulator	642	0.7	0.15	0.2	11.7

4.3 Concluding remarks

The increasingly fast and cost-effective sequencing of the genomes of antibiotic-producing actinomycetes has revolutionised the way in which the biosynthesis of specialised metabolites is studied. Unfortunately, even sophisticated automated methods for assignment and annotation such as AntiSMASH, when applied to highly-fragmented or poor quality genome sequence assemblies, can give very misleading results. In this chapter AntiSMASH 4.0 has been used on two essentially complete genomes, both from strains of *S. malaysiensis*, to attempt to produce a definitive analysis of the biosynthetic potential of these strains. This analysis has confirmed the unusual wealth of biosynthetic clusters in these bacteria, comprising not only 'generalist' clusters familiar from previous work on model organisms, but also clusters shared with other closely-related species, as well as near-endemic clusters such as that for the peptide malayamycin in DSM14702. It has also underlined the importance of comparative studies of phylogenetically-related strains.

The transcriptomic analysis has shown that, once the hurdle of obtaining high-quality RNA has been overcome, then it is possible, in a rapid and relatively inexpensive way, to obtain valuable information about those clusters that are active and those that are 'silent' under the chosen growth conditions, and about the timing and the coordination of that expression. Such information, especially about the behaviour of putative regulatory genes, can be very useful to guide future attempts to manipulate individual clusters. In the near future it may become cost-effective to carry out RNAseq in parallel on a strain grown in multiple media, and with more replicates and over longer timescales; to monitor the progress of strain improvement; and perhaps also to monitor changes in transcription pattern that take place as a result of targetted mutation. The method does not, of course, supplant the use of RT-PCR for study of specific genes or sets of genes.

The analysis of the putative prajinamide cluster has illustrated how clues from work on a closely-related *Streptomyces* can trigger the search for an unknown gene cluster for an interesting chemotype or scaffold in the genome of interest. For prajinamide, this analysis has led to a prediction that will need to be tested experimentally. In the next chapter, work is described both predicts the novel gene cluster encoding the biosynthesis of a potent nucleoside biocontrol agent, and carries out the necessary experiments to test the hypothesis.

Chapter 5. Mobilisation of the spinosyn gene cluster from *Saccharopolyspora spinosa*

5.1 Introduction

Transcriptomics and genome analysis provide a wealth of insight into a strain of interest and its potential for producing specialised metabolites, and provide impetus for the search for novel compounds and for activating biosynthetic gene clusters. The information is of equal importance for rational improvement in the production of valuable known compounds. The work in this chapter describes efforts to exploit the previously-characterised biosynthetic gene cluster from *S. spinosa* which governs production of the important insecticide known as Spinosad. *S. spinosa* remains a genetically "difficult" strain, so heterologous expression of the spinosyn BCG would provide an attractive new platform for manipulating the cluster to provide new analogues of the spinosyns.

5.1.1 Heterologous expression

Despite the rapid increase in genomic information, as yet only a small fraction of the potential products have been realised. Manipulation of clusters in *S. malaysiensis* DSM4137 in this laboratory, for example, has been restricted by the limited genetic tools available to alter the strain. Also, production of specific metabolites from cultivated streptomycete strains can vary widely for reasons that are not well understood. Many strains being investigated today, especially those from metagenomic screens, are not culturable in standard laboratory settings (Rutledge and Challis, 2015). Although advances like the ichip (Ling et al. 2015), have helped culture strains for the first time under laboratory conditions, strains from extreme environments such as halophilic, marine and cold-adapted Antarctic bacteria may not express their biosynthetic potential without special handling (Barone *et al.*, 2014; Trindade *et al.*, 2015). A technique used to overcome these issues is the heterologous expression of BGCs in hosts that are genetically amenable and well-studied (Gomez-Escribano and Bibb, 2014).

An obvious potential advantage of heterologous expression is that an appropriate host strain provides a 'clean' background for the study of the products of biosynthesis. Also, it may be possible to define the minimum size of the cluster that confers production. In theory, the target metabolite might be produced at higher levels than in the original strains (although this is rarely seen in practice). As an illustration, the heterologous expression of the 80 kbp BGC for the polyketide FK506 in *S. coelicolor* was successful, and the production level at 5.5 mg/L was a real - but modest - improvement on the wild type strain which produced at 1.2 mg/L

(Jones et al,. 2013). Certainly, clusters can be more readily manipulated in *E. coli* using a variety of well-established and specific editing techniques.

In early work on heterologous expression of actinomycete BGCs, the well-studied model organism *S. coelicolor* was the almost inevitable choice, and it remains an attractive one. In particular, mutant strains are available in which the endogenous gene clusters for specialised metabolites have been ablated. *S. coelicolor* M1142, a plasmid free derivative of M145, has had deletions in the regulatory genes *redD* and *actII*-ORFIV. This has abolished production of prodiginine and actinohidrin. M1146 *(*Δ*act,* Δ*red,* Δ*cpk,*Δ*cda*) a derived strain from M1142, has had its four BGCs deleted from the strain giving a cleaner metabolic background. Modifications in *rpoB* (RNA polymerase β-subunit) and *rpsl* (ribosomal protein S12) of *S. coelicolor* have introduced strains M1152 and M1154 which usually allow for higher production of heterologous compounds.

S. lividans is a closely related strain of *S. coelicolor* is another commonly used strain for heterologous expression. The undecylprodigiosin, actinohidrin and calcium-dependent antibiotic are all silent in *S. lividans* but expressed in *S. coelicolor*. *S. lividans* TK24 (*str-6, act+, red+)* has a high transformation efficieny and can accept methylated DNA (Hopwood et al., 1983). Other strains include *S. avermitilis, S. albus* and *S. erythraea. S. avermitilis* ,an industrial producer of avermectins, has been developed as versatile host for heterologous expression with a minimal genome (Komatsu et al., 2010;Komatsu et al., 2013). *S. albus* composes the smallest genome of know sequenced strains. It has an exceptionally fast growth rate and can accept methylated DNA (Zaburannyi et al., 2014). Brady and coworkers use the strain as a host for heterologous expression from metagenomes and example being the Fluostatins (Feng et al., 2010). *S. erythraea*, responsible for the production of erythromcin, is a more closely related strain to *S. spinosa.* It has been used as a heterologous host for the tetronate, RK-682 (Sun et al, 2010). It contains the building blocks required for heterologous expression of PKS and has been used for the heterologous expression of a truncated polyketide synthase (Martin et al., 2003).

5.1.2 The spinosyn insecticides

The spinosyns are an industrially important family of compounds produced by the actinomycete *S. spinosa*. *S. spinosa* was discovered in 1990 (Mertz and Yao, 1990) at the site of a defunct sugar mill in the Virgin Islands. Cell extracts of this strain showed insecticidal activity against lepidopterous insect pests and the spinosyns were purified by monitoring this

activity (Thompson et al., 1997). The spinosyns act by a novel mode of action on the nicotinic acetylcholine receptors (nAChR) of insects, different from receptors targetted by the neonicotinoid and pyrethroid insecticides (Millar and Denholm, 2007). The major metabolites produced by *S. spinosa*, spinosyns A and D, make up the insecticidal chemical marketed as 'Spinosad '. This neurotoxin kills many insect pests and has no known toxic effect on humans (Bond et al. 2004, Pan et al. 2011). In 1999 Dow Agrosciences won the 'Presidential Green Chemistry Award' for the development of Spinosad as a relatively environmentally friendly biocontrol agent against pests of important commercial crops.

5.1.3 The structure of natural spinosyns

The core structure of the spinosyns consists of a 12-membered macrocyclic ring fused to a 5,6,5-*cis-anti-trans* tricyclic ring system (Waldron et al., 2001). The major component of the spinosyn fermentation is spinosyn A (Figure 4.1). The second most abundant compound is spinosyn D, which differs in having an additional methyl group at the C6 position (Figure 4.1).

5.1.4 Spinosyn polyketide biosynthesis

Clegg Waldron and colleagues at Dow AgroSciences reported the sequence of the BGC for spinosyns in *S. spinosa* in 2001 (Waldron et al., 2001). Their bioinformatic analysis of the cluster, combined with analysis of several mutants in which genes has been successfully disrupted, led to a plausible proposal for the biosynthetic pathway involving chain assembly on a canonical modular type I PKS (Figure 4.1). The cluster contains five Type I polyketide synthase genes, which extend over 56 kbp. The KS domain of the loading module lacks the conserved active site cysteine and contains a glutamine residue instead (Waldron et al., 2001), consistent with recruitment of methylmalonyl-CoA and its *in situ* decarboxylation by the KSQ to provide the starter unit (Bisang et al., 1999). The extender units are specifically either acetate or propionate units except for the unit inserted at module 8, where the acyltransferase, despite having active site motifs typical of a methylmalonyl-CoA-specific AT, appears to recruit both malonyl- and methylmalonyl-CoA, leading under standard ferementation conditions to a final mixture of 80% spinosyn A, 20% spinosyn D (Waldron et al., 2001).

Additional genes within the ~80 kbp cluster are predicted to encode for sugar attachment and for some steps (but not all) in deoxyaminohexose (forosamine) biosynthesis. Genes for rhamnose biosynthesis are also absent. Other genes were unassigned but were

hypothesised to be involved in the remarkable intramolecular C-C cross-bridge formation (see below) and in rhamnose attachment and methylation (Waldron et al., 2001).

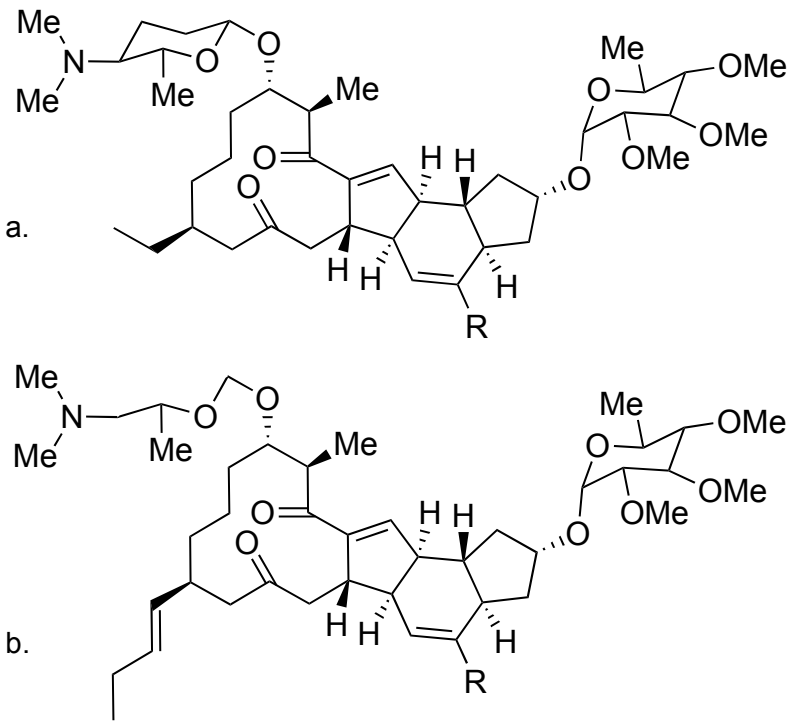


Figure 5.3: The structure of spinosyns A, D and butenyl spinosyn. a. For spinosyn A, R = H and for spinosyn D, R = CH_3 b. Butenyl-spinosyn

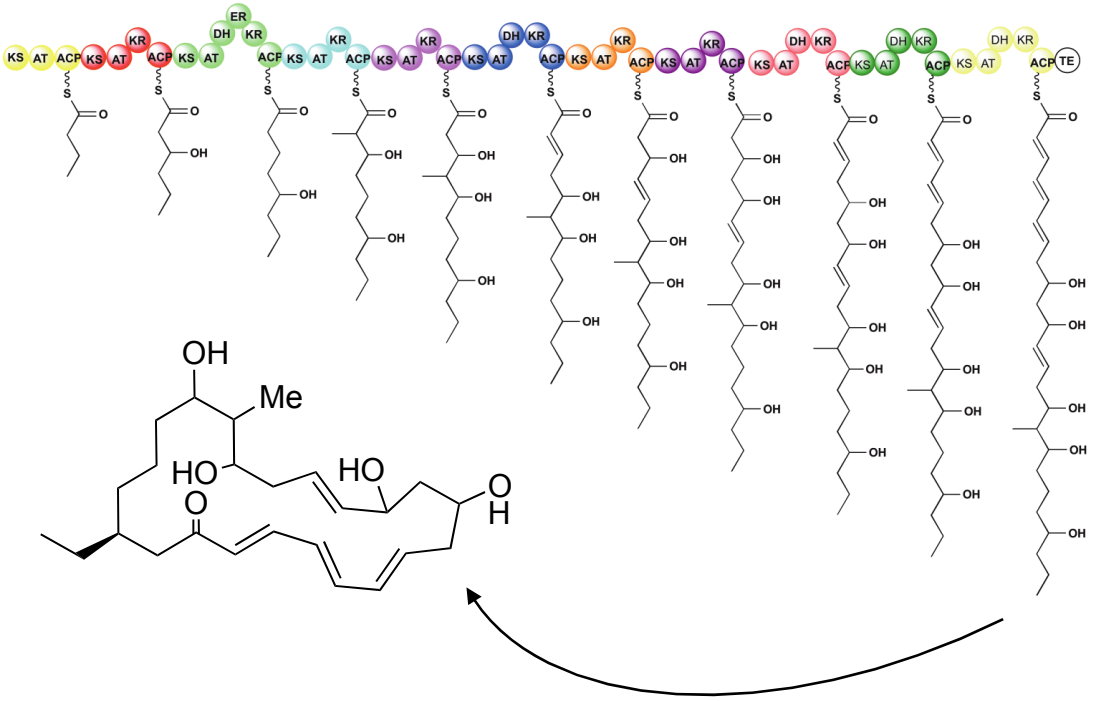


Figure 5. 4 The assembly-line formation of the spinosyn polyketide backbone. The 56 kbp polyketide backbone encodes one loading module and 10 extender modules. Nine acetates and two propionates lead to the formation of a spinosyn aglycone. The KS domain of the loading module lacks the active site cysteins at position 174 and contains a glutamine instead. The AT domain of module 8 is similar to the propionate loading domain but only incorporates propionate 20 % of the time. Adapted from (Waldron et al,. 2001)

5.1.5 Late-stage tailoring reactions in spinosyn biosynthesis

The decaketide polyketide macrolactone produced by the spinosyn PKS undergoes an intramolecular cyclisation catalysed by SpnF, one of the first authenticated enzyme-catalysed Diels-Alder reactions (Kim et al., 2011) (Figure 4.3). The addition of a rhamnose sugar follows, and the appended sugar is then methylated to give a 2,3,4-tri-*O*-methylated rhamnose at the C9 position of the aglycone (Kim et al., 2011) A further cross-bridge is formed between C-7 and C-14 via an Rauhut-Currier reaction catalysed by SpnL (Figure 4.3), which is followed by the addition of the aminosugar forosamine (Hong et al., 2008) to the C17 position of the macrocyclic ring in the pseudoaglycone (Figure 4.2).

Natural minor variants of the spinosyn molecule have been reported, including changes at the C6, C16 and C21 positions of the tetracycle, the methyl position of the

forosamine nitrogen and the 2'-, 3'- and 4'-positions of the rhamnose moiety (Crouse et al., 2001).

5.1.6 Spinosyn analogues

Recently the neonicotinoid class of insecticide has come under scrutiny for their effects on non-target insects, and a total ban is possible in Europe. There is already a Europe-wide ban in place on the triazole pesticides due to their endocrine-disrupting potential. Clearly, new insecticides with a good environmental and toxicological profile are urgently needed, and the spinosyn scaffold offers a promising starting point. There is a well-developed SAR for spinosyn based on chemical modifications made at Dow AgroSciences (Sparks et al., 2008). Notably, removal of the forosamine moiety, for example, results in a complete loss of activity (Geng et al., 2013). There are a number of natural spinosyn variants but unfortunately none have as yet shown significant improvement on the insecticidal activity and environmental profile of spinosyn A. For example in 2006 *Saccharopolyspora pogona* was found to produce butenyl-spinosyns (Hahn et al., 2006).The gene cluster sequence for butenyl-spinosyns shows 94 % identity to that for spinosyn biosynthesis except that the PKS contains an additional polyketide module responsible for the introduction of two extra carbons at the C-21 position (Lewer *et al.*, 2009) (Figure 5.3) The presence of the double bond in butenyl-spinosyns has been used to carry out semisynthesis via chemical cross-metathesis, and encouragingly some at least of the new compounds retained insecticidal activity (Daeuble *et al.*, 2009).

A limited amount of genetic engineering has been carried out on the spinosyn PKS to make alterations at the C21 position of the spinosyn molecule by genetically modifying the loading domain. The new derivatives were as active as spinosyn A, but production levels were too low in the wild type strain for this method to appear commercially viable (Sheehan et al., 2006), Modifications of the 5-6 double bond of the tetracycle allowed greater photolytic stability but internal unsaturation lowered insecticidal activity (Crouse et al., 2009).

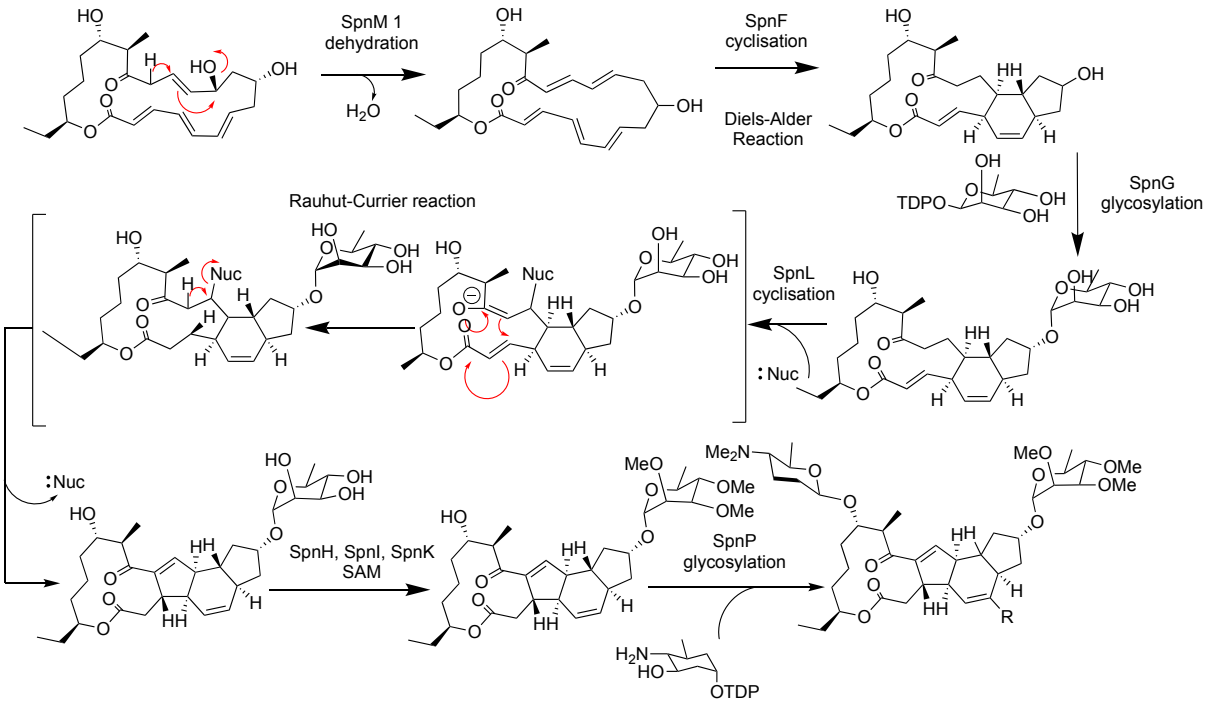


Figure 5.5 The late steps of the spinosyn biosynthetic pathway, (adapted from Kim et al., 2010, 2011). The formation of the aglycone is followed by intramolecular cyclisation that occurs via a catalysed Diels-Alder reaction. Addition of a rhamnose sugar follows which is then methylated to a 2,3,4-tri-O-methylated rhamnose at the C9 position. Further intramolecular cyclisation between C-7 and C-14 occurs via a Rauhut-Currier reaction, which is followed by the addition of the methylated amino sugar forosamine to the C17 position of the macrocyclic ring.

5.1.7 Spinetoram

In 2007 a modified version of Spinosad, called Spinetoram, was launched by Dow Agrosciences. This semi-synthetic compound bears 5,6-dihydro and 3'-O-ethyl substitutions compared to spinosyn A (Sparks et al., 2008). It has been shown to be more active than spinosad and with an expanded spectrum, while maintaining an excellent environmental and toxicological profile. Spinetoram won the 'Presidential Green Chemistry Challenge - Designing Greener Chemicals Award' in 2008.

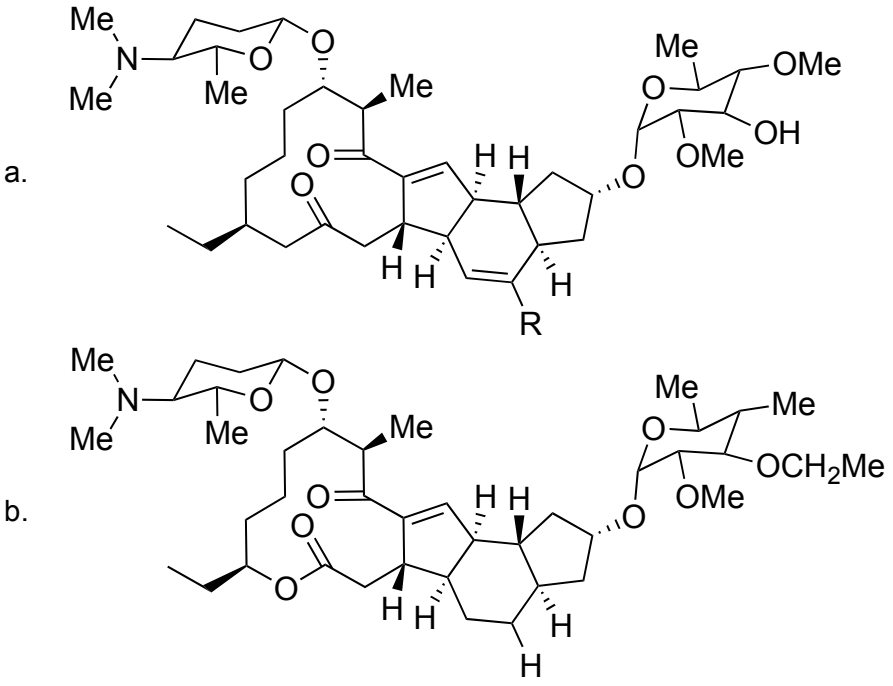


Figure 5.4 The structure of spinetoram. The molecule (b) is derived from the biosynthetic intermediate spinosyn J (a) (R=H) by chemical conversion of the 3'-hydroxyl of rhamnose to the ethyl ether.

5.1.8 Alternative methods for heterologous expression of natural product biosynthetic clusters

Because genetic manipulation of the spinosyn biosynthetic cluster in the wild type *S. spinosa* strain had proved experimentally so difficult over the last 20 years (P. F. Leadlay, personal communication) heterologous expression appeared an attractive alternative to allow

manipulation of the spinosyn BGC in a host that was genetically amenable. As discussed below, the choice of method to achieve this lay between library screening methods and direct cloning or 'direct capture' methods.

A traditional method of cloning BGCs involves the construction of a cosmid library (for inserts up to ~48 kbp) or of a bacterial artificial chromosome (BAC/PAC) library (for inserts up to at least 140 kbp) in *E. coli* and their subsequent screening for the presence of the intact target cluster within the insert. BAC vectors are based on the *E. coli* fertility plasmid (F-factor) origin of replication and PAC vectors on the P1-phage origin of replication, and both are replicated as single copies in *E. coli*. Use of the PAC vector has been shown to produce more stable libraries (Pierce et al., 1992). The size of a BGC can vary from 20 kbp to over 100 kbp, therefore libraries would often only capture the small BCGs or only part of a large one. Libraries with larger inserts are technically more demanding to construct, and although custom-built libraries can be conveniently ordered from specialist suppliers, this is an expensive option. Also, screening library clones by PCR to check for the presence and the integrity of the entire cluster is relatively time-consuming.

Assuming a correct clone is identified, this then must be mobilised from *E. coli* into an appropriate expression host, most conveniently using conjugation. Site-specific integration of the conjugated DNA into the chromosome of the actinomycete expression host is the preferred method of ensuring stability of expression of the cloned genes. An example of this approach is the reported cloning of the 83.5 kbp DNA region containing the biosynthetic gene cluster from *Streptomyces tsukubaensis* for FK506 (tacrolimus), a clinically important immunosuppressant. A PAC library was constructed using the P1-derived artificial chromosome pESAC13 (Sosio et al., 2000). Only trace amounts of FK506 were detected in the *S. coelicolor* heterologous host strain, but subsequent overexpression of a LuxR regulatory protein did increase production (Jones et al., 2013).

At the outset of this part of the present work, alternative targeted approaches were (becoming) available for the heterologous expression of actinomycete BGCs. One method available at the time for direct cloning or "capture" of a desired piece of DNA was transformation-associated recombination (TAR) cloning. TAR cloning is a well-established *in vivo* technique using *Saccharomyces cerevisiae* (Larionov et al. 1996). TAR cloning makes use of the high level of homologous recombination and ease of transformation of *Saccharomyces cerevisiae* (Kouprina and Larionov 2016). The method relies on the homologous recombination of a linearised plasmid which has homologous overlaps to a segment of DNA of interest. This is the method used to make yeast artificial chromosomes

(YACS) (Ramsay 1994). The Brady group introduced TAR cloning for the capture of microbial BGCs from soil-derived cosmid clones in 2010 (Kim et al. 2010).

Meanwhile, Stewart, Müller and colleagues had developed an elegant *in vitro* technique they have named linear linear homologous recombination (LLHR). They showed that full-length RecE and RecT from the Rac prophage together promote homologous recombination between two linear molecules, and used these proteins to clone large regions of DNA directly from purified genomic DNA into *E. coli* expression vectors bypassing library construction and screening. They exemplified the method by direct cloning of nine PKS and NRPS BGCs with sizes ranging from 10 to 37 kbp. However, capture of a 52 kbp BGC was not possible (Fu et al. 2012).

Other techniques can be used to construct BGCs, refactor BGCs, and edit genes *in vitro*. Golden Gate and MASTER (methylation-assisted tailorable ends rational) require the use of type IIS enzymes (cutting away from their specific recognition site). Golden Gate assembly, using *Bsa*I, *Bsm*BI and *Bbs*I, can assembly up to nine fragments in a specific order. However, the efficiency of assembly using type IIS enzymes is low when constructing large fragments. In contrast, several closely-related techniques all use homologous recombination rather than restriction enzymes to manipulate large pieces of DNA efficiently, prominent among which is isothermal assembly (Gibson et al. 2009; Gibson 2011).

Isothermal assembly (Gibson *et al.*, 2009; Gibson, 2011) relies on three enzymes in a one-pot reaction mixture: first, a T5 exonuclease creates single-stranded overhangs on both the vector and the intended insert DNA. The exonuclease is deactivated by heating to 55°C and a thermostable Phusion DNA polymerase enzyme takes over. DNA ligase is present to seal nicks. *E. coli* is transformed with the product mixture and selection occurs on agar containing the appropriate antibiotics. Isothermal assembly has the advantage that it leaves a scarless product. It is said not to work well with high (G+C)% DNA because the reaction is conducted at the relatively low temperature of 50°C (Gibson et al., 2009). On the other hand, Dr Y. Zhou in this laboratory had recently succeeded in using isothermal assembly for the direct cloning of the 42 kbp actinomycete gene cluster governing biosynthesis of the polyketide diolide conglobatin (Zhou *et al.*, 2015), using the *E. coli* – *Streptomyces* shuttle plasmid pSET152.

5.2 Results and Discussion

5.2.1 Capture of the spinosyn biosynthetic gene cluster

In the application of isothermal assembly as used by Zhou et al. (2015), the biosynthetic gene cluster is digested from total genomic DNA using restriction sites outside

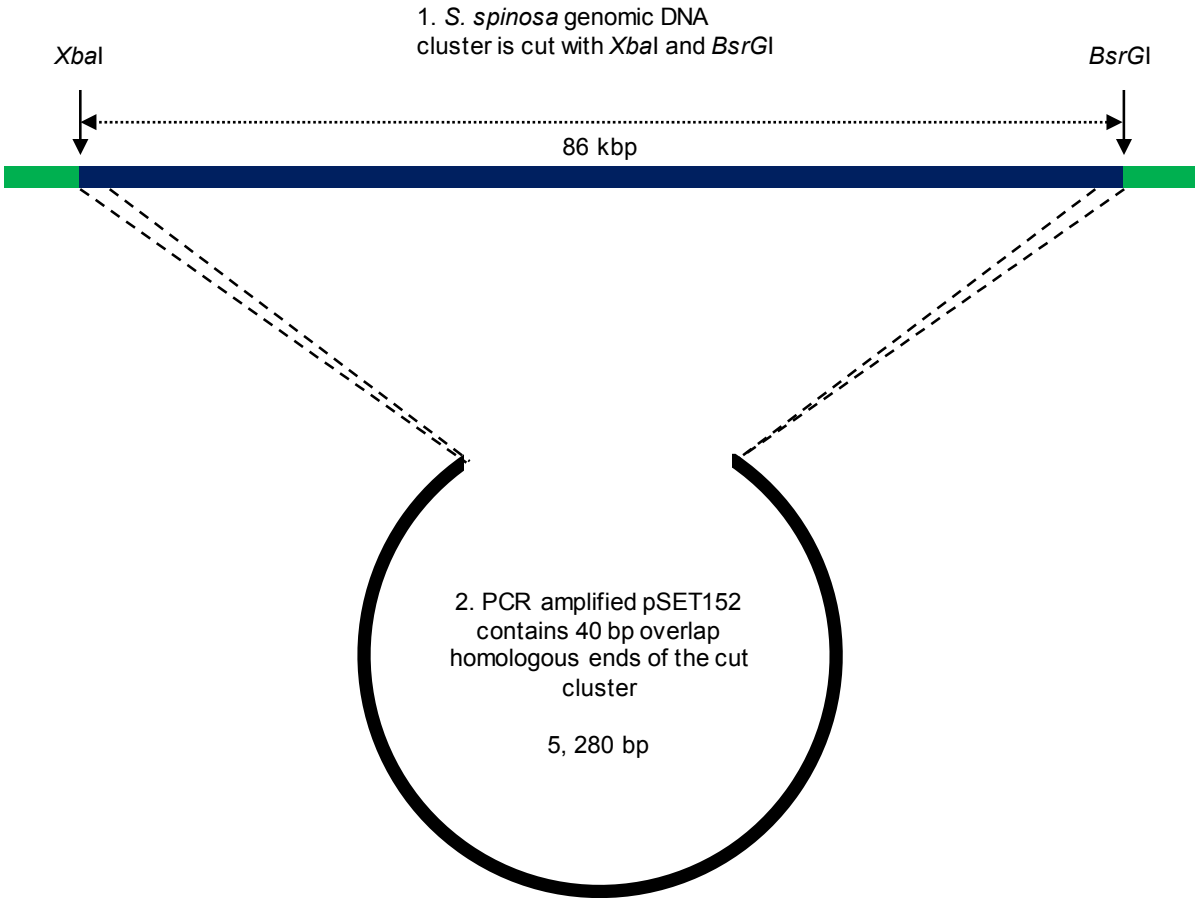


Figure 5.6 The principle of direct cloning of the spinosyn cluster using pSET152 and isothermal assembly.

of the cluster that do not cut within it, and the product mixture, including the cluster-bearing fragment, is submitted to the isothermal assembly three-enzyme process together with a PCR-amplified vector backbone. The PCR primers used to amplify the vector backbone contain overlapping regions of about ~40 bp with perfect homology to the ends of the digested gene cluster. The isothermal assembly reaction products are then used to transform an *E. coli* strain. Colonies are tested for the presence of the expected insert-containing plasmid by colony PCR, positive colonies are grown overnight and plasmid extracted using the alkaline lysis method (Materials and methods section 2.42). Initially, the widely-used cloning vector pSET152

(Bierman *et al.*, 1992) was chosen as the vector, as had been successful for Zhou et al. (2015). It confers resistance to apramycin (Tang *et al.*, 2000; Kumagai *et al.*, 2010). It contains the ΦC31 *attp-int* system for site-specific integration at the phage attachment site in the host genome (Combes *et al.*, 2002). Despite the well-known tendency of such high copy number vectors to undergo unwanted recombination, it would allow a quick initial test of the direct capture method with such a very large insert.

S. spinosa DNA was extracted as previously described (Materials and methods section 2.43). Two restriction enzymes, *Xba*I and *Bsr*GI, were found to cut conveniently outside of the predicted limits of the spinosyn cluster and not within it (Figure 4.5). They were used to digest the genomic DNA for 4 hours at 37 °C. The resulting genomic DNA fragments were treated in two alternative ways, to remove small fragments that might interfere in the isothermal assembly reaction.

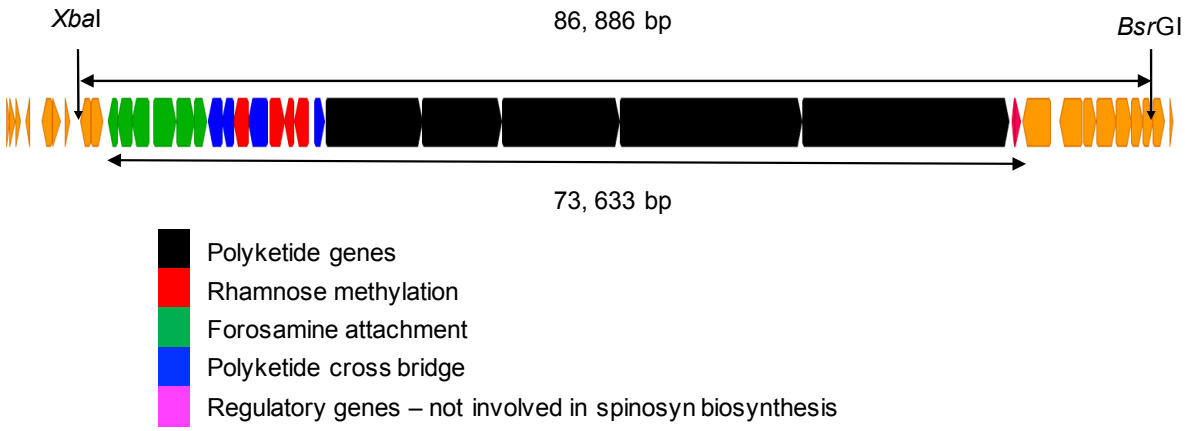


Figure 5.5 The spinosyn biosynthetic pathway. *Xba*I *and Bsr*GI were identified as having appropriate restriction sites outside the cluster, and none within it.

For a portion of the mixture, overnight electrophoresis on agarose gel (0.4%, low melting agarose) was used to separate large and small fragments, and DNA was extracted by using a ß-agarase digestion (Figure 4.6) :

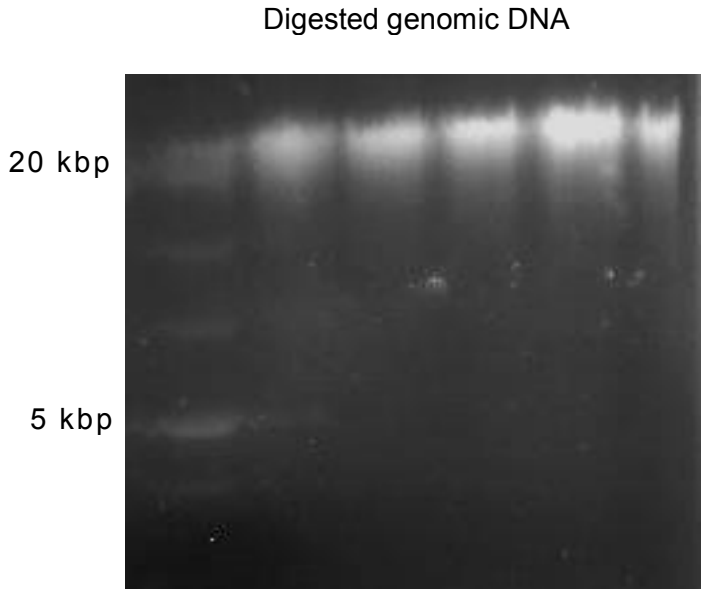


Figure 5.7 Removal of small from large genomic DNA fragments using agarose gel electrophoresis. Genomic DNA was digested with *Xba*I and *BsG*I for four hours. Electrophoresis on a 0.4 % low melting point agarose gel was conducted overnight at 3 V/cm. In this system all fragments larger than ~20 kbp run together on the gel. Large fragments of DNA were extracted from the gel using β-agarase and precipitated with isopropanol.

The rest of the mixture of genomic DNA fragments was purified using phenol/chloroform extraction, precipitated with isopropanol and washed with 70 % ethanol. Purified genomic DNA from either procedure was added to an isothermal assembly reaction mixture and incubated at 37 °C for 15 minutes, and then for between one and four hours at 45 ˚C (Gibson *et al.*, 2009; Gibson, 2011). The isothermal assembly mixture was desalted by drop dialysis (Materials and methods section 2.35) and then 1-10 μl was used to transform *E. coli* DH10B cells by electroporation.

Colony PCR was used to detect the presence of the cluster (Figure 4.7). Ten of these putatively positive colonies were grown overnight and plasmid DNA extracted. A further PCR reaction using the extracted plasmid as template unfortunately gave negative results. DNA sequencing of the insert in one of these plasmids revealed that loss of the insert and a major rearrangement of the vector had occurred. Not entirely surprisingly, the high copy number vector housing the 87 kbp fragment inserted here had proved to be unstable. Nevertheless, this initial trial did show that the gene cluster could be cloned using the isothermal assembly method. This technique using pSET152 therefore appeared eminently suitable for assembling

plasmids of less than 50 kbp. It has in fact become an established method in this laboratory for that purpose.

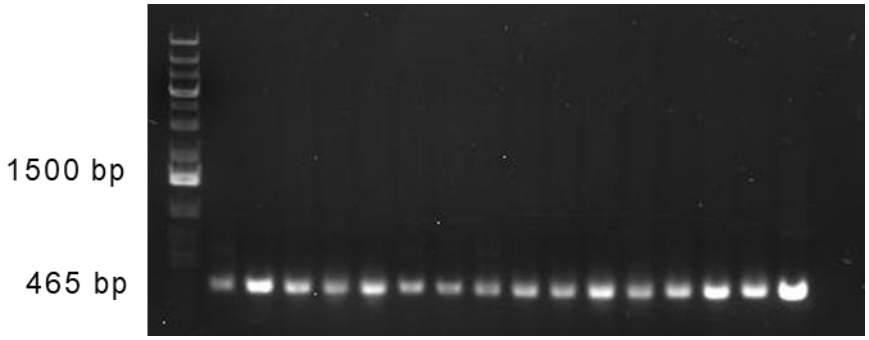


Figure 5.8 Colony PCR of recombinant *E. coli* to test for the presence of the spinosyn biosynthetic gene cluster. All sixteen colonies initially tested proved positive for the presence of the polyketide synthase gene *spnA*. Lanes 1-15, colony PCR; lane 16, genomic DNA positive control; Lane 17, vector negative control.

5.2.2 Capture of large biosynthetic gene clusters using the PAC vector pESAC13

The PAC vector pESAC13 developed by Donadio and colleagues (Sosio et al., 2000) contains the P1 phage origin of replication, and like pSET152 is an *E. coli* – *Streptomyces* shuttle vector. In contrast to pSET152, though, the PAC vector pESAC13 is a single copy number vector in *E. coli* (Jones *et al.*, 2013). However, it contains an inducible *lac* operator for conversion to a multi copy number if required. Importantly, it has been shown stably to contain genomic DNA fragments greater than 150 kbp and it has become the vector of choice for commercial custom-built actinomycete genome libraries. The cloning site for pESAC13 lies between two *Bam*HI restriction sites. A T7 promoter and a *sacB* gene flanking the restriction sites allows selection against re-ligated vectors on agar containing 5 % sucrose. pESAC13 contains the ΦC31 *attp-int* system to allow specific integration into the target genome, and confers kanamycin resistance in *E. coli* and thiostrepton resistance in *Streptomyces*.

pESAC13 at 23.3 kbp in size (20.6 kbp after *Bam*H1 digestion) is considerably larger than pSET152. This makes PCR amplification of the plasmid improbable. A new method was therefore devised to capture large BGCs using pESAC13. To test this new method, the large gene cluster encoding the biosynthetic pathway to the polyether ionophore nigericin was selected. The nigericin BGC from *S. malaysiensis* DSM4137 at ~132 kbp (Harvey et al., 2007) is significantly larger than the spinosyn cluster, so if this test were successful it would be very encouraging. Crucially, in this laboratory the nigericin cluster has previously been cloned from

a PAC library made in pESAC13, and conjugated into *Streptomyces* spp. where nigericin production was seen (J.-O. Frister and Y. Zhou, unpublished), demonstrating that the integrity of the cloned DNA had been preserved.

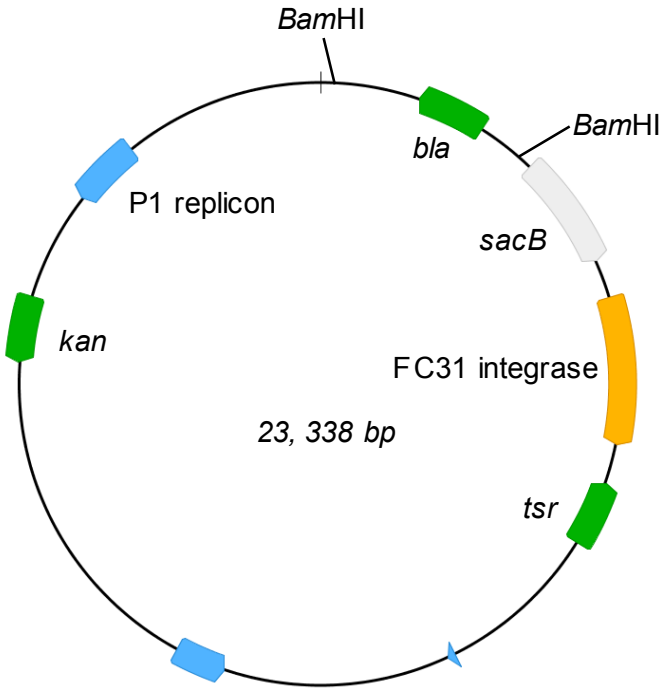


Figure 5.8 The pESAC13 *E. coli* - *Streptomyces* shuttle vector. pESAC13 *contains* a *sacB* gene flanking the *Bam*HI cloning site. This allows selection against *E. coli* containing religated vector on media containing 5 % sucrose**.**

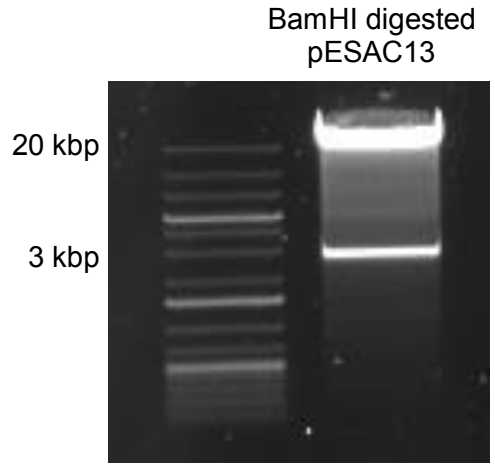


Figure 5.9 The PAC vector pESAC13 (23,338 bp) becomes a linear molecule of 20, 604 bp after digestion at its two *Bam*H1 sites.

One complication with the choice of the nigericin BGC was that no suitable (combination of) restriction enzyme(s) was identified flanking the large cluster that did not also cut inside the cluster. Still, if this new method were to be generally useful this was an issue that would anyway need to be addressed. *Xba*I was chosen as the restriction enzyme to use, even though this enzyme would cut once within the nigericin cluster, as shown in Figure 4.10. A four-part isothermal assembly was designed to attempt the direct capture of the nigericin gene cluster, explained in detail below.

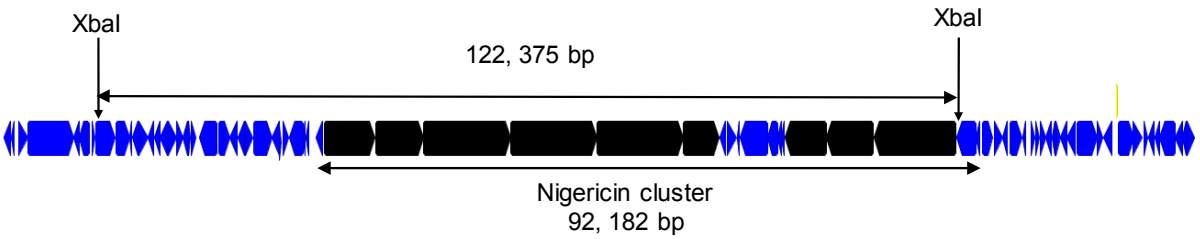


Figure 5.9 The nigericin biosynthetic gene cluster. *Xba*I restriction sites in the region of the 122 kbp cluster are located as shown, with one site located 2, 436 bp inside the cluster boundary. The nigericin modular polyketide synthase genes are indicated by black arrows.

5.2.3 Four-part isothermal assembly for biosynthetic gene cluster capture

Isothermal assembly is normally used to stitch together fragments of DNA that contain overlaps of 15 – 40 bp with high efficiency, but if increased specificity is required it can accommodate overlaps of over 100 bp (Gibson *et al.*, 2009; Gibson, 2011). The GC content of the overlapping regions of the isothermal assembly is crucial: if the GC content is too high, the efficiency will drop considerably and if the overlapping region is too short, nonspecific binding can result. If the overlapping regions create hairpins that are too strong, the efficiency of the reaction will also drop considerably. A balance must therefore be struck between efficiency and selectivity.

The design of the four-part assembly is shown in Figure 5.11. A PCR fragment was amplified either side of the *Xba*I restriction sites. On the left-hand side, a 1.7 kbp fragment of the flanking region was PCR amplified, the PCR product also including 100 bp of the nigericin cluster and at the 5' end overlapping the pESAC13 vector sequence by 40 bp. On the right-hand side, a 3.5 kbp PCR fragment was amplified, 100 bp into the cluster from the *Xba*I cut site creates an overlap with the cluster, outside of the cluster a PCR primer has a 40 bp region overlapping pESAC13.

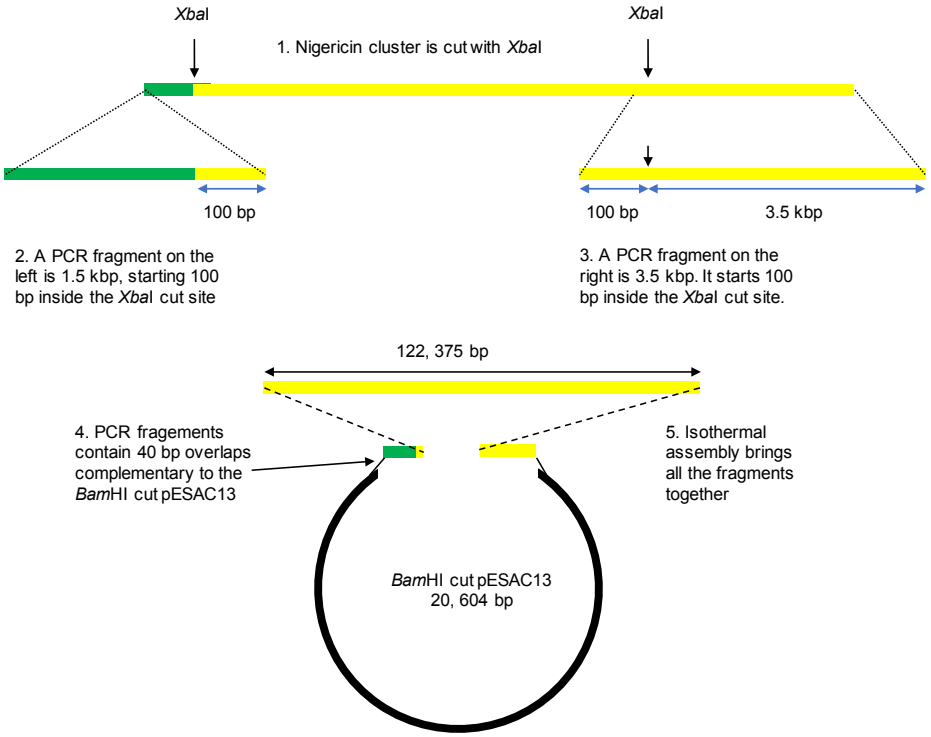


Figure 5.11 Design of an isothermal assembly to capture the nigericin biosynthetic gene cluster. The four fragments to be assembled are the vector, the *Xba*I-cut nigericin gene cluster, and two bridging PCR fragments, as shown.

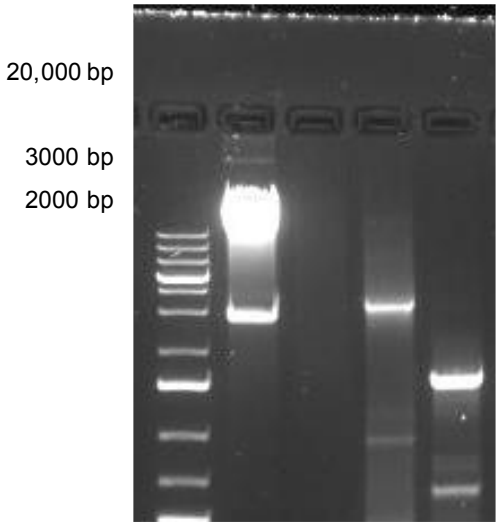


Figure 5.12 Isothermal assembly of the nigericin cluster in pESAC13: the fragments to be combined with the *Xba*I-cut DNA region containing the nigericin cluster. Lane 1, molecular weight standards, Generuler 1kb plus; lane 2, *Bam*HI-cut pESAC13; lane 3, nigericin right-hand side PCR product; lane 4, nigericin left-hand side PCR product.

These purified PCR products provided two of the four fragments required for assembly. Meanwhile, genomic DNA was cut with *Xba*I, taking care at every stage to minimise shearing of the DNA, and pESAC13 was digested with *Bam*HI and purified on a 0.5 % gel. All four fragments were added to an isothermal assembly mixture and incubated together for 15 mins at 37 °C and then 45 °C for 1, 2, 3 and 4 hours (Gibson *et al.*, 2009; Gibson, 2011; Fu *et al.*, 2014). The isothermal assembly reaction mixture was then used to transform *E. coli* NEB DH10B high efficiency electrocompetent cells, and the desired recombinants were selected for on 2TY solid medium containing kanamycin and 5% sucrose.

Seven colonies were obtained from the transformation at the 2 hour reaction time-point. Colony PCR was performed using primers designed to amplify a portion of one of the nigericin PKS genes and yield a product of 483 bp. Positive colonies (Figure 4.13a) were grown overnight in 2TY containing kanamycin. Plasmids were extracted using alkaline lysis and the junction of the plasmid was checked by sequencing. One of the positive clones was named pFD030:Nig and used for further experiments.

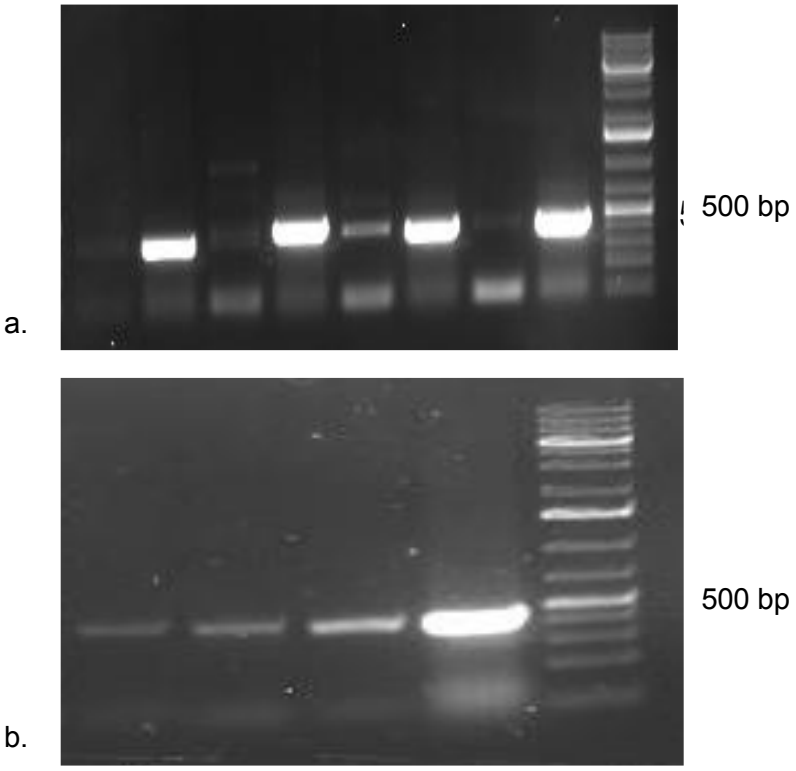


Figure5.13 Colony PCR of candidate recombinants for the presence of the assembled nigericin biosynthetic gene cluster. a, E. coli transformants. All positive recombinants were taken forward for conjugation into S. lividans; lane 8, genomic DNA positive control; b, S. lividans TK24 exconjugants. The recombinants of lanes 1 and 2 were used for further analysis. Lane 4, genomic DNA positive control; DNA ladder – Generuler 1 kb plus.

5.2.4 Transfer of pFD030:Nig into a methylation-deficient strain

As discussed, previous work in this laboratory had led to the heterologous expression of nigericin in *S. lividans* TK24 and *S. coelicolor* M1154 using a PAC clone identified by screening a previously constructed pESAC13 library. *E. coli* DH10B is a strain routinely used for cloning in which large plasmids are known to be stable. However, plasmid DNA propagated in DH10B will become methylated *via* the *dam, dcm* and *hsd* systems (MacNeil et al., 1992).

S. coelicolor M1154 has a strong methylation-specific restriction system preventing entry of foreign methylated DNA. Transfer of the plasmid DNA of interest (here pFD030:Nig) into the methylation deficient ET12567 (*dam*$^-$*, dcm*$^-$*, hsd*$^-$*)* strain is essential before intergeneric

transfer into M1154. The strain ET12567 (Cam^R) normally contains the 'helper plasmid' pUZ8002 (Kan^R) which facilitates conjugation of the *oriT*-containing plasmid into the recipient *Streptomyces* strain. Since pESAC13 (Kan^R) also confers kanamycin resistance in *E. coli* , ET12567 was instead transformed with the self-transmissible helper plasmid pR9604 bearing carbenicillin resistance ($Carb^R$) to allow selection for the PAC clone of interest.

The large size of the PAC plasmids can make transformation difficult. Chemical transformation is not possible therefore transfer is often achieved by "triparental mating". However, this can often lead to rearrangements, especially in plasmids containing regions of high homology. Here, electroporation was used, using a modified protocol (Materials and methods section 2.38), to transform pFD030:Nig (Kan^R) into ET12567 (pR9604) (Kan^R, $Carb^R$).

5.2.5 Transfer of pFD030:Nig into *Streptomyces* host strains

pFD030:Nig was introduced into M1154 and TK24 via mycelial conjugation (Materials and methods section 2.38). After 16 hours growth on SFM solid medium containing 20 mM $MgCl_2$, plates were overlaid with thiostrepton, to select for *Streptomyces* clones that contain pFD030:Nig, and with nalidixic acid, to stop the growth of any remaining *E. coli*. After five days of growth four M1154 exconjugants and five TK24 exconjugants were visible. This is fewer than would normally be expected for conjugations into TK24 and M1154, but the low efficiency may be because of the size of the plasmid being transferred.

Exconjugants were restreaked on SFM agar containing thiostrepton to ensure that selection was for exconjugants in which pFD030:Nig had been stably integrated into the each genome. Genomic DNA was extracted and presence of the nigericin cluster checked by PCR amplification of a region of PKS. Of the nine exconjugants tested, three tested positive for the nigericin gene cluster. These were named M1154:Nig1, TK24:Nig1, and TK24:Nig2 respectively.

5.2.6 Production of nigericin in *Streptomyces coelicolor* M1145 and *S. lividans TK24*.

Strain M1154:Nig1, and strains TK24:Nig1 and Nig2, were grown in 7 ml TSBY for three days in 50 ml conical flasks containing springs and then 5 ml was transferred to a 250 ml flask containing 50 ml TSBY. LC-MS/MS analysis of the supernatant extraction using

selective ion monitoring for the m/z of the sodium salt of nigericin $[M+Na]^+$ = 747.5) revealed a peak with the same retention time as authentic nigericin and with the same MS-MS fragmentation pattern. Production was not seen in all samples in either *S. lividans* or *S. coelicolor*, a variability which was also seen in the previous work with the PAC clone selected from the pEASAC library and cloned into both TK24 and M1154. This variability may well reflect adventitious recombination between the mutually homologous DNA regions encoding modules of the nigericin PKS in the heterologous host strains.

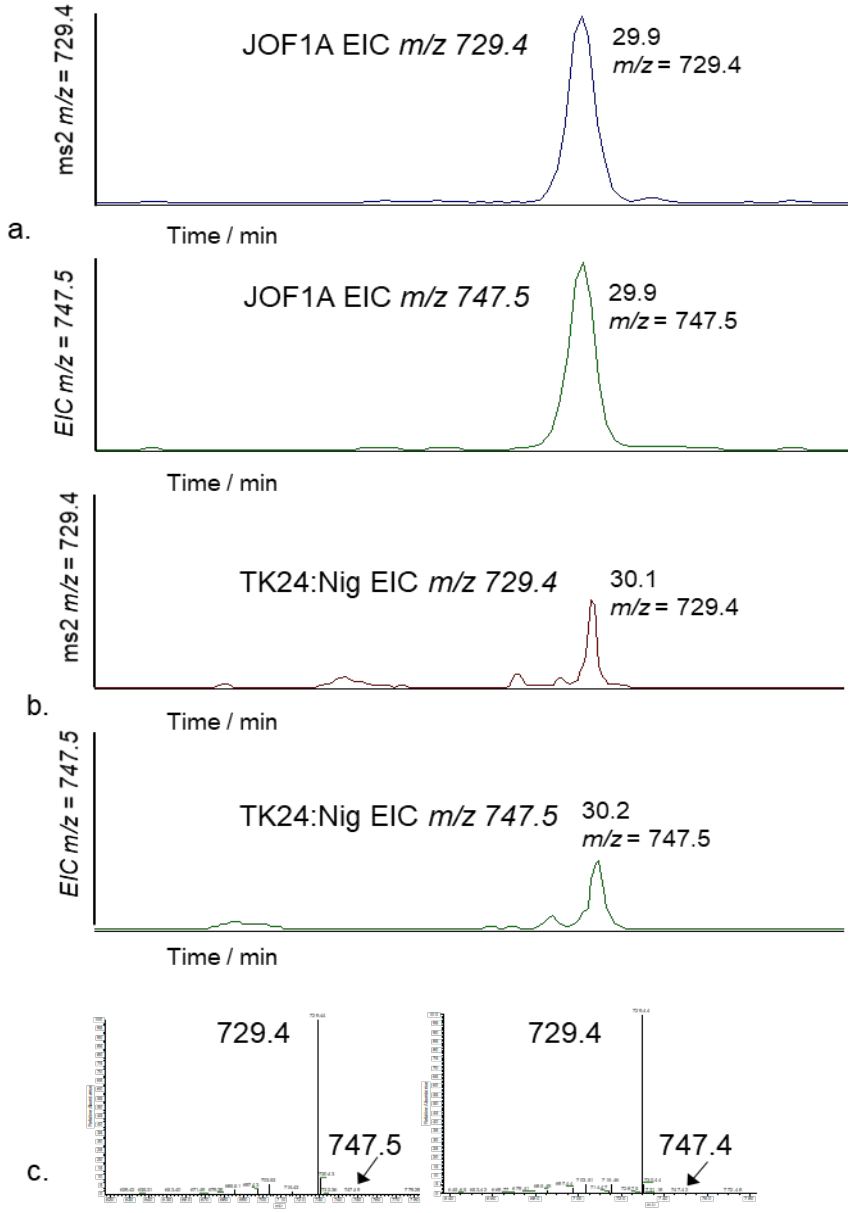


Figure 5.14 LC-MS analysis of nigericin produced in the heterologous host strain *S. coelicolor* M1145. The peak at 30.1 min in the extract from TK24:Nig1 had the same retention time as authentic nigericin, and the expected m/z $(M+Na)^+$ = 747.5 for nigericin. a. Nigericin production from previous heterologous construct by Ole Frister and DY Zhou in the Leadlay Lab. b. Heterologous expression of TK24:Nig1. c. MS/MS analysis – left standard, right TK24:Nig1 confirm production of Nigericin.

No attempt was made to optimise the level of heterologous expression of nigericin in these experiments. It was certainly low compared to production in the native strain S. malaysiensis DSM4137,(in which nigericin production reaches ~100 mg/L (Harvey et al., 2007), but it was at levels similar to those seen after expression in M1145 of the PAC clone isolated from the pEASAC library (Figure 4.14). This result strongly suggested that the four-part isothermal assembly procedure had been successful, and encouraged its application to the heterologous expression of the spinosyn biosynthetic gene cluster.

5.2.7 – Capturing the spinosyn biosynthetic gene cluster

To capture the spinosyn BGC, a four-part isothermal assembly was designed in the same way as for the nigericin BGC. *S. spinosa* genomic DNA was digested using the restriction enzymes *Xba*I and *Bsr*GI. The resulting mixture is predicted to contain a DNA fragment of 86.9 kbp housing the entire cluster (73.2 kbp) and flanking regions of 11.6 and 2.2 kbp respectively. Plasmid pESAC13 was digested with *Bam*HI as previously described and purified on a 0.5 % agarose gel. 100-170 bp PCR fragments were amplified from either side of the digested pESAC13 vector (P1:Spin and P2:Spin). 500 – 700 bp fragments were PCR amplified from *S. spinosa* genomic DNA (S1:Spin and S2:Spin). Each PCR fragment amplified from genomic DNA contained around 150 bp of DNA inside the relevant restriction site of the BGC and the rest was amplified from the flanking region outside of the restriction site. Each PCR fragment amplified from pESAC13 contained a 20 bp tag which corresponds to the appropriate genomic DNA amplified fragment; P1:Spin to S1:Spin and P2:Spin to S2:Spin.

PCR fusion of P1:Spin and S1:Spin and P2:Spin and S2:Spin created two of the fragments for use in a four-part isothermal assembly reaction. These PCR fragments were gel purified and an isothermal assembly reaction set up containing gel-purified *Bam*HI-cut pESAC13, *Xba*I/*Bsr*GI-cut genomic DNA and the two PCR fragments.

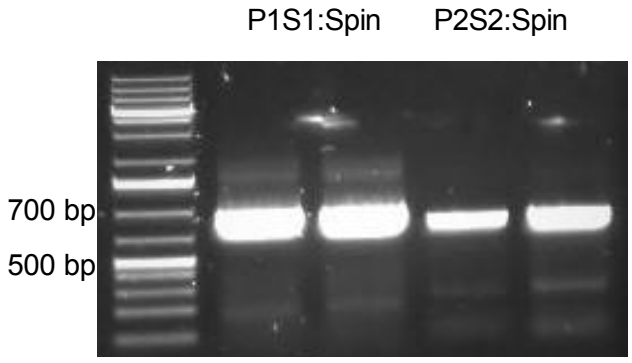


Figure 5.15 Bridging PCR fragments created for use in the four-part isothermal assembly of a pEASAC-based clone housing the spinosyn biosynthetic gene cluster.

The isothermal assembly reaction mixture was incubated for 15 minutes at 37 °C and then for 2 hours at 45 °C. The resultant mixture was transformed into *E. coli* DH10B by electroporation (2.43) and recombinants were selected on 2TY agar containing kanamycin and 5 % sucrose. Colonies were observed 36 hours after electroporation. Twenty-two of the

colonies tested positive, as judged by observation of an approximately 500 bp band after colony PCR using primers to part of the *spnA* PKS gene (Figure 4.16). Ten of these were grown overnight in 2TY media containing kanamycin. Only two grew in liquid media. Plasmid from one of these was named pFD040:Spin and was used to transform the methylation-deficient *E. coli* strain ET12567/pR9604 with selection on 2TY plates containing kanamycin, chloramphenicol and carbenicillin, as previously described for the nigericin plasmid.

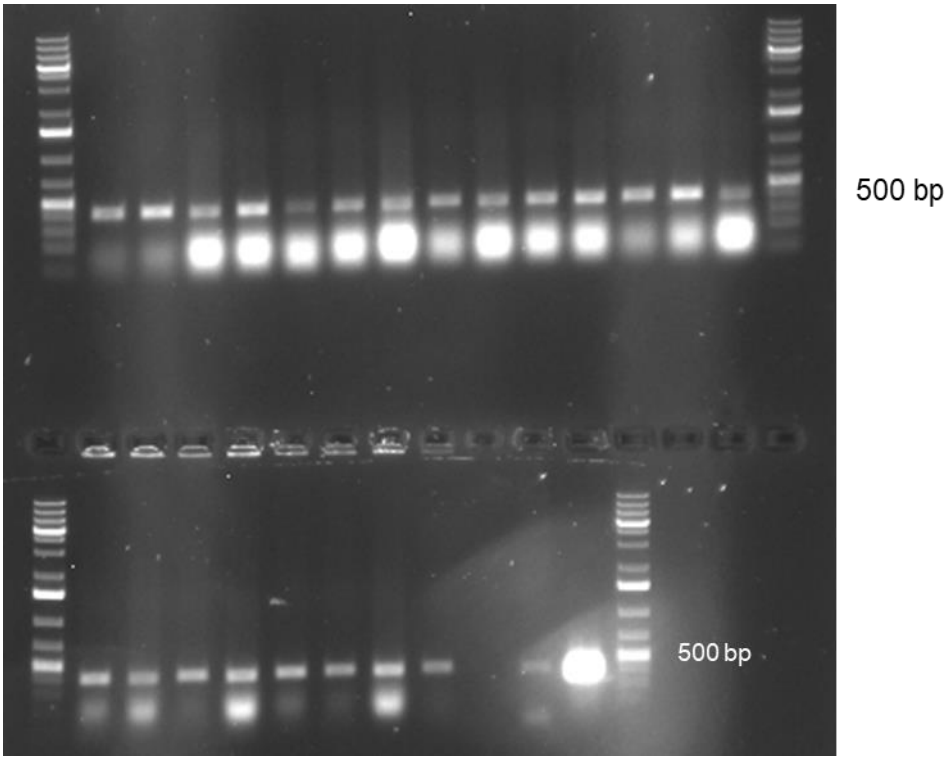


Figure 5.16 Colony PCR screening of *E. coli* recombinants for the presence of the spinosyn biosynthetic gene cluster. Ten of the colonies showing the brightest band at just under 500 bp were selected for growth in liquid medium containing kanamycin. Lane 25, genomic DNA positive control; Ladder – Generuler 1 kb plus.

5.2.8 Conjugation of pFD040:Spin into actinomycete host strains

pFD040:Spin was introduced into M1154, TK24 and *S. erythraea* BIOT2085 via mycelia and spore conjugation (Materials and methods section 2.43). After 16 hours growth for M1154 and TK24, and 18 hours growth for BIOT2085, plates were overlaid with thiostrepton and nalidixic acid. Exconjugants were restreaked on SFM agar containing

thiostrepton to ensure that selection identified exconjugants in which pFD040:Nig was stably integrated into the respective genomes. Genomic DNA was extracted to check for the presence of the *spnA*. From a total of twelve exconjugants, one M1154, three TK24 and no BIOT2085 exconjugants tested positive for the spinosyn gene cluster. These were named M1154:Spin1, TK24:Spin1, TK24:Spin2 and TK24:Spin3 respectively. These strains were grown in 7 ml TSBY for three days in 50 ml flasks containing springs and then 5 ml of the culture was transferred to a 250 ml flask containing 50 ml TSBY or Spinosyn fermentation media (Strobel Jr and Nakatsukasa, 1993).

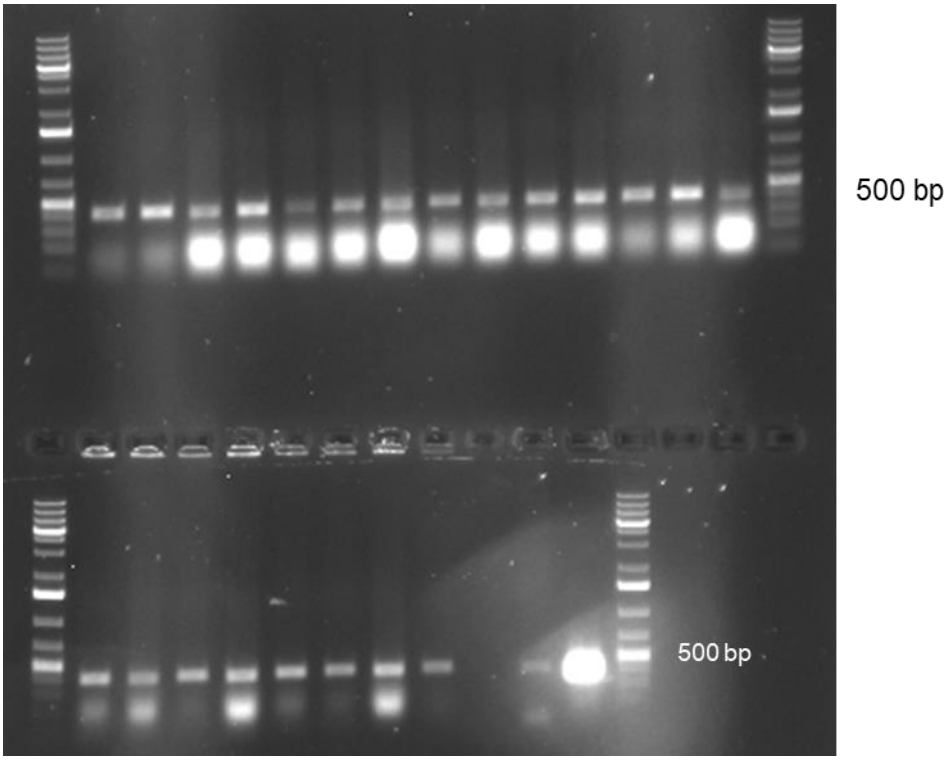


Figure 5.17 Detection of spinosyn in heterologous host; a. M1154 and b. TK24.
a. Lanes 2 to 5 primers testing for SpnE, SpnA, SpnD and SpnK respectively. Lanes 6 to 9, genomic DNA positive control. b. Lanes 2 to 5 primers testing for SpnE, SpnA, SpnD and SpnK respectively. Lanes 10 to 13, genomic DNA positive control.

The four strains M1154:Spin1 and TK24:Spin1-3 were grown in liquid medium for 7 days and the broth supernatants were extracted and subjected to LC-MS/MS analysis, to test for the production of spinosyn or spinosyn-related metabolites. Unfortunately, none of these strains were found to produce such compounds.

At this point, and despite the disappointing initial lack of exconjugants when *S. erythraea* BIOT2085 was used as a recipient in conjugation experiments with pFD040:Spin, it was decided to focus further attention on this strain as a potential expression host, because its cellular machinery can reasonably be expected to be far closer to that of *S. spinosa* itself, and in these experiments the native promoters of the spinosyn BGC are being relied upon. Further, there is very limited understanding of precise modes of regulation of BGCs in *Saccharopolyspora* spp. BIOT2085 is a strain derived from a parent erythromycin-overproducing strain by deletion of the resident erythromycin BGC and it was hoped that this too would favour production of an alternative polyketide product such as spinosyn.

5.2.9 Electroporation of *S. erythraea* BIOT2085 with pFD0040:Spin

Numerous attempts were made to transfer pFD0040:Spin into BIOT2085 by conjugation. Overlay times, medium composition and *E.coli/Streptomyces* cell number ratios were all varied. Conjugation was successful but subsequent PCR analysis of the spinosyn polyketide region returned a negative result. Triparental mating using DH10B containing the helper plasmid pUB307 (Kan^R) was also attempted to test whether modification of the cluster was occurring within ET12567 itself, but with no greater success in the subsequent conjugation step. Therefore, to attempt to bypass any potential issues arising either from the use of ET12567 or from *S. erythraea*'s response to incoming DNA via conjugation, an electroporation protocol was developed, based on a protocol devised for the electroporation of *S. erythraea* spores (Wang *et al.*, 2008).

pFD0040:Spin was incubated in cell-free extract of *S. erythraea* with the addition of SAM (Matsushima and Baltz, 1994; Donahue *et al.*, 2000; Chen *et al.*, 2012). The recovered plasmid was used to transform *S. erythraea* BIOT2085 spores by electroporation and recombinants were selected on ABB13 solid medium containing kanamycin. Four colonies were observed and these were restreaked on ABB13 agar and SFM agar containing thiostrepton. Genomic DNA was extracted to check for the presence of the spinosyn cluster. One tested positive and this isolate was named BIOT2085:Spin.

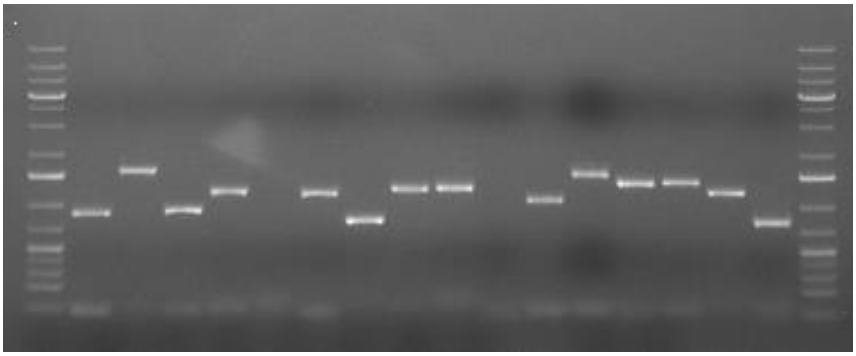


Figure 5.18 PCR analysis of the spinosyn heterologous plasmid. Lanes 2 – 4, cross-bridging genes; Lanes 6 – 10, rhamnose methylation genes, Lanes 12- 16 polyketide genes; Lane 17, *tsr* gene; Ladder used is Generuler 1 kb plus.

5.2.6 Detection of spinosyns in extracts of cultures of recombinant S. erythraea BIOT-2085

BIOT2085:Spin was grown in 7 ml TSBY for three days in 50 ml flasks containing springs and then 5 ml of the culture was transferred to a 250 ml flask containing either 50 ml TSBY or Strobel fermentation medium (Strobel Jr and Nakatsukasa, 1993). After supernatant extraction and LC-MS/MS analysis, production of spinosyn was observed. Production was very low and was not stable.

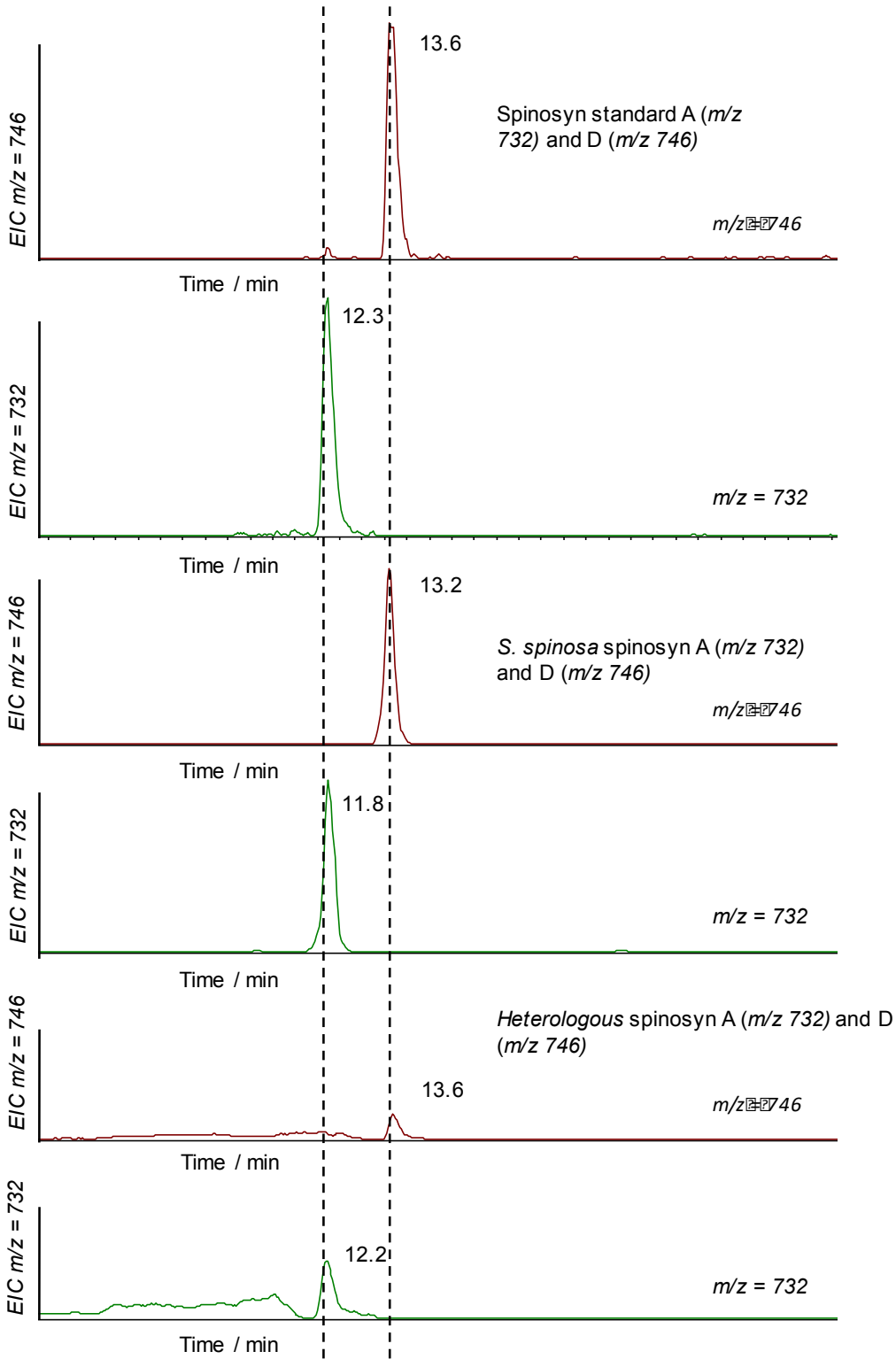


Figure 5.10 Heterologous expression of the spinosyn biosynthetic cluster. Production was always at the same low level: spinosyn-related peaks could not be detected in the total ion current (TIC) trace., However the identity of the putative spinosyn A and spinosyn D signals in ion-selective mode was seen.

An attempt was made to obtain higher and more stable production of spinosyns in BIOT2085:Spin by making serial dilutions of spores, and replacing them on ABB13 solid media

to isolate single spores. Ten such single spore isolates were tested for the production of spinosyn after each round of isolation. However, no further increase in spinosyn production was observed.

5.2.7 Introduction of pFD0040:Spin into *S. spinosa*

It has been reported that introduction of a partial biosynthetic gene cluster into the *S. spinosa* genome led to an increase in production. Introduction of the post PKS genes, *spnF to spnR* into the genome increased titre from 100 to 388 mg L^{-1} (Tang *et al.*, 2011). Intergeneric conjugation of pFD0040:Spin ($Thio^R$) was carried out using mycelia and spores. Also, electroporation of *S. spinosa* spores collected from BHI solid medium was carried out after incubation of pFD0040:Spin in cell-free extract of *S. spinosa* (Matsushima and Baltz, 1994; Wang *et al.*, 2008).

Conjugation following a published protocol (Xue *et al.*, 2013) led to a few exconjugants that acquired thiostrepton resistance. Electroporation produced two colonies that acquired thiostrepton resistance. All strains lost the ability to produce spinosyn. PCR analysis of of the polyketide genes returned negative results, it seems spinosyn cluster had been lost following conjugation.

The genome sequence reveals that this species does not contain a sequence corresponding perfectly to the attB recognition site of the ΦC31integrase. Rather, it contains several mismatched pseudo-attB sites where integration is likely to be less efficient. It is possible that inefficient integration of the psuedo *attB* in *S. spinosa* allows homologous recombination between the plasmid insert and the resident spinosyn BGC to occur, disrupting spinosyn production. Changing the integrase to the recently discovered Chen integrase which contains a 56 bp core sequence in the *attb/attp* site which is contained in the *S. spinosa* genome (Chen *et al.*, 2015).

5.2.9 Addition of genes responsible for forosamine and rhamnose biosynthesis

It had been shown that one bottle-neck for spinosyn production was the genes that are involved in the biosynthesis of rhamnose and forosamine (Pan *et al.*, 2011). The genes *gtt* and *gdh* are involved in the biosynthesis of both sugars. Overexpression of these genes in the spinosyn pathway has shown to increase production of spinosyn in the WT.

gtt/kde and gdh were PCR amplified from S. spinosa genomic DNA. Here they were assembled in pGP9, an integrative plasmid that contains the ΦBT1 integrase site. PGP9 was

cut with *NdeI* and *XbaI* and the genes inserted under the control of the strong heterologous promoter *PactI*. The resultant plasmid pFD0047:Rham was transformed into the ET12567 (pUZ8002) strain. Conjugation into the wild-type strain and BIOT2085:Spin allowed for integration into the partial *attB* ΦBT1 site of the *S. spinosa* chromosome. Potential exconjugants were then restreaked on SFM and ABB13 containing apramycin (*S. spinosa*) or apramycin and thiostrepton (BIOT2085:Spin).

Fermentation in TSBY and Strobel spinosyn fermentation media followed by analysis of the extracts on LC-MS/MS confirmed that production of the wild-type was increased with the addition of these genes. Production in BIOT2085:Spin however showed no increase in production.

While working on the heterologous production of spinosyn, the heterologous expression of spinosyn in *S. erythraea* was published (Huang *et al.*, 2016). Here, it was noted that the addition of the sugar genes to *S. erythraea* does not increase the production of spinosyn if the titer is already low.

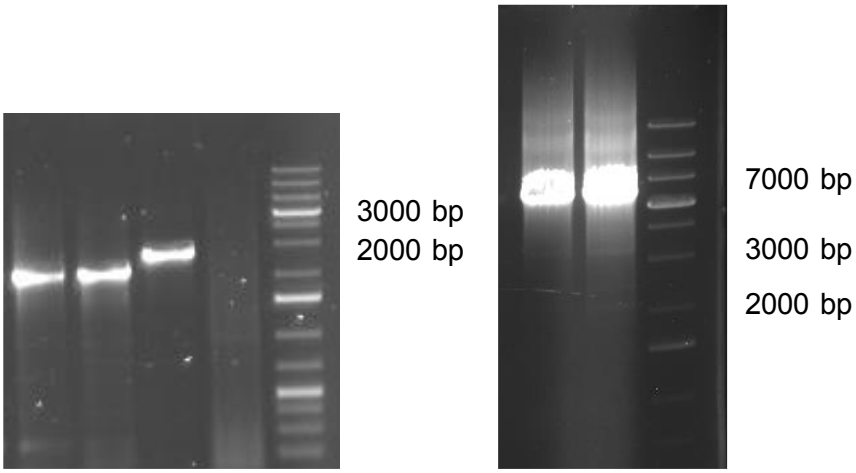


Figure 5.11 Overexpression of the spinosyn sugar genes. The genes were captured on PGP9, a plasmid with the ΦBT1 integrase and PactI promoter, for overexpression in the heterologous host. a. Expression of the individual sugar genes; Lanes 1 to 3, *gtt, epi, gdh/kre*; b. Lanes 1 and 2, PCR amplifies pGP9. DNA ladded used – Generuler 1kb plus

5.4 Discussion and future work

The aim of this project was to capture and then to achieve a stable expression of the spinosyn biosynthetic gene cluster in the heterologous host to enable fast and easy genetic engineering of the cluster. This was reasoned by inability to perform any genetic manipulations of the wild type producer of spinosyns, *S. spinosa*, and by very attractive and potent herbicide activity of these molecules.

When this work was started, only limited examples of direct cloning of the entire biosynthetic pathways were reperted (mainly for clusters not exceeding 20-40 kbp) [refs]. Over the last three years the number of techniques for direct capture of biosynthetic gene clusters was growing. In 2014, the Moore group published direct capture of 60 kbp NRPS gene cluster using TAR cloning, however, the efficiency was at 2 %.(Yamanaka et al., (2014). This was further improved, and in 2015, the Moore group published on an updated TAR plasmid which allowed the capture of BCGs at up to 50 % efficiency. This is a large improvement on the previous plasmid and takes direct capture using yeast homologous recombination a step forward (Tang et al. 2015). In 2015, CRISPR was used with Gibson assembly to capture a 100 kbp biosynthetic gene cluster with what is known as the CATCH technique (Jiang et al. 2015). Both these methods have been trialled alongside the work described here to capture nigericin and spinosyn BGCs.

In the present work the large DNA fragments-capture approach was developed independently and was shown to be adaptable for size of insert and for the availability of restriction sites. These factors along with size of DNA, degree of homology in the sequence, GC content, time and cost are all reasons to have multiple techniques on hand when looking at heterologous expression and capture of the biosynthetic gene cluster of interest. I was able to use this approach to assemble two challenging biosynthetic gene clusters – 133 kbp nigericin PKS and 86 kbp spinosyn PKS. My protocol was successfully used by fellow lab members to assemble multiple BGCs.

Following assembly of spinosyn BGC-containing PAC clone, heterologous expression in industrial *S. erythraea* BIOT strain showed production of spinosyns A and D, although at low levels. On the way to this result I had to optimize conjugation conditions, spore/mycelia preparation protocols, and adapt methods of handling and treatment of large DNA fragments to the capabilities of the lab.

While working on this project, heterologous expression of the spinosyn gene cluster was reported by two other groups using *S. erythraea* (Huang et al. 2016) and later *S. albus*

(Tan et al. 2017) as heterologous hosts. In both cases only low production of spinosyn in heterologous host was initially observed, consistent with the results described in this chapter. For espression in *S. erythraea,* it was shown by Huang et al that the addition of the sugar biosynthetic genes did not have an effect on production, and random mutagenesis was undertaken to enhance production. Sequencing of the genome of the mutant stain and identification of the mutations that caused such effect on the spinosyn biosynthesis would be beneficial for the scientific community. When the cluster was moved into *S. albus host*, a BAC library was ordered to obtain a stable plasmid with the spinosyn BGC. After the cluster was integrated into the host, intensive transcriptomics, proteomics analysis was performed which allowed targeted refactoring of the pathway to improve production. Addition of the rhamnose biosynthetic genes and *spnI,* a forosamine biosynthetic gene further improved the production.

Knowledge from these papers can be applied to the strain created in this work. Specific mutagenesis along with rational refactoring should result in stable spinosyn production and will allow to perform genetic engineering of the cluster with a goal to deliver novel analogues with potent bioactivity.

.

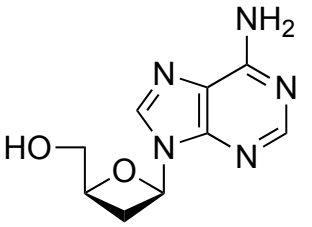
Chapter 6 Biosynthesis of the herbicidal nucleoside albucidin

6.1 Introduction

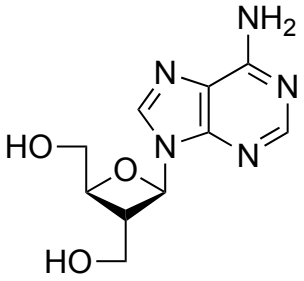
In previous chapters the value has been shown of an essentially complete actinomycete genome sequence in enabling a different approach to the discovery of the biosynthetic genes for both expected and unexpected specialised metabolites; and an exploration of methods for heterologous expression that can both aid initial analysis of a target gene cluster and also overcome technical problems in manipulation of the original strain. My BBSRC-sponsored PhD project is associated with, although separately funded from, the major BBSRC- and Syngenta-sponsored programme of collaborative research (sLoLa) being carried out in this and three other Universities from 2013-2018. I have been able to take advantage of the wealth of genome sequence information accumulated within the sLoLa programme during this period to explore the biosynthesis of albucidin, a nucleoside with a highly intriguing chemical structure and a potent herbicidal activity of potential industrial interest.

6.1.1 The structure of albucidin and related compounds

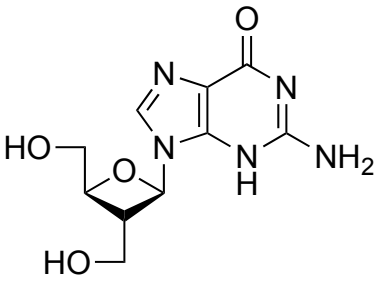
Albucidin is a novel nucleoside (Figure 6.1) produced by *Streptomyces albus* subsp. *chlorinus*, a strain that also produces a characteristic brilliant green-yellow pigment in its aerial mycelium when grown on oatmeal agar (Hahn *et al*., 2009). Albucidin is also reported to be a potent herbicide that induces bleaching and chlorosis in grass and broadleaf weeds (Hahn *et al*., 2009). It shares with oxetanocin A (9-[(2*R*,3*R*,4*S*)-3,4-bis(hydroxymethyl)-2-oxetanyl] adenine) (Figure 6.1), a bioactive metabolite produced by *Bacillus megaterium* NK84-0218 (Shimada *et al*., 1986) the distinction that they are the only two known examples of a naturally-occurring four-membered ring nucleoside. The antiviral activity of oxetanocin A stimulated a major synthetic effort in many laboratories to explore this class of modified nucleoside as anti-viral (and especially anti-HIV) compounds (Kitagawa *et al*., 1991). Among the compounds synthesised at the time as analogues of oxetanocin A were oxetanocin-G in which the base was switched, and carboxetanocin-G, in which the oxygen atom of the oxetan ring was replaced (Sato et al., 1989) (Figure 6.1).



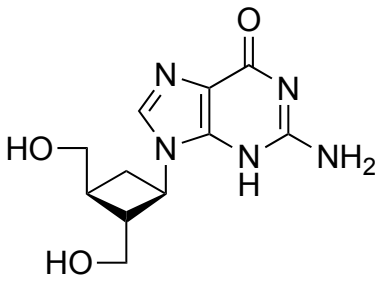
Albucidin



Oxetanocin A



Oxetanocin G



Carboxetanocin G

Figure 6.1 The structure of albucidin and oxetanocin A, and of typical synthetic analogues tested as anti-viral compounds

Figure 6.2 The genes in *Bacillus megaterium* NK84-0218 putatively responsible for the biosynthesis of the antiviral nucleoside oxetanocin A. The length of the DNA region shown is 6.8 kbp.

6.1.2 The biosynthesis of oxetanocin

The genes responsible for oxetanocin A biosynthesis are borne on a plasmid in *Bacillus megaterium* NK84-0218. It was shown that transfer of a 6.8 kbp fragment of plasmid DNA into the heterologous host *Bacillus megaterium* ATTC6459 conferred oxetanocin A production (Morita *et al.*,1999). This 6.8 kbp fragment contains four ORFs: *osxA, oxsB, oxrA* and *oxrB* (Figure 6.2) and it was proposed that these genes might comprise a biosynthetic cluster with structural and regulatory genes (Morita et al. 1999). Analysis of their sequence now suggests that OxsA and OxrB are HD-domain phosphohydrolases, OxrA is a pentapeptide repeat protein of unknown function, and, most tellingly, OxsB is a cobalamin-dependent S-adenosylmethionine (AdoMet) radical (radical SAM) enzyme. Radical SAM enzymes are a rapidly-expanding and versatile class of enzymes, using radical intermediates to accomplish catalytic conversions that are difficult via conventional enzyme catalysis (Bandarian, 2012; Ruszczycky *et al*., 2012), for example the ring contraction that leads to 7-deazapurine antibiotics in *Streptomyces rimosus* (Bandarian, 2012)

The aims of the work in this chapter were: first, to use genome mining of the near-complete genome sequence of the albucidin producer to identify candidate genes for albucidin biosynthesis, in which it was hoped that sequence homologies with genes of the oxetanocin cluster would be helpful; secondly, to obtain experimental evidence to support such candidate genes being implicated in the biosynthesis; and, thirdly, to define the minimal cluster that is sufficient to govern albucidin biosynthesis. It was also hoped that production of albucidin, which as explained below is erratic and low in the producing strain, might be improved.

6.2 Results and Discussion

6.2.1 Albucidin detection

Albucidin was originally reported to be produced at levels of up to 2 mg/L, production being detectable after six days and maximum yield being attained after 12-13 days of fermentation (Hahn *et al.*, 2009). To test for production, *S. albus* subsp. *chlorinus* was grown for 12 days in production medium G (Hahn et al. 2009). Initial experiments showed that albucidin could be indeed be detected in *n*-butanol extracts of the culture by using LCMS, but the production was significantly lower than described in the literature, at around 0.1 – 0.5 mg/l. For this part of the work, Diane Urwin (Syngenta, Jealott's Hill, UK) kindly supplied details of methods she used for detection, and a standard sample of albucidin.

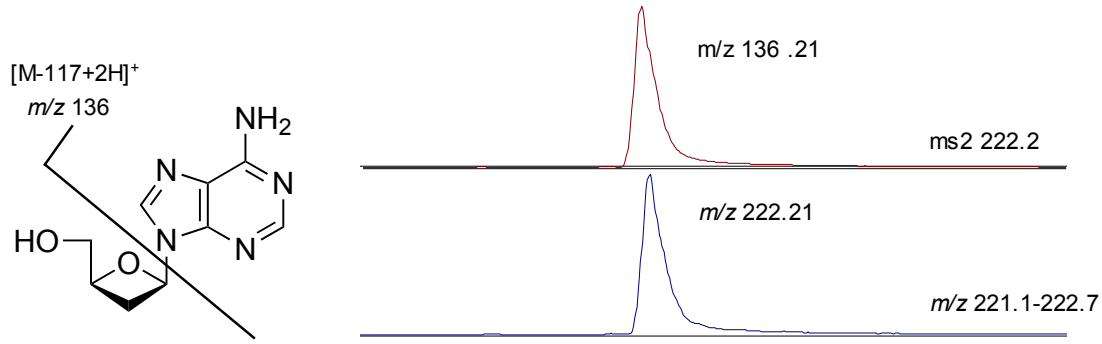


Figure 6.3 LC-MS/MS analysis of an albucidin standard sample supplied from Syngenta. MS/MS and TIC showing albucidin standard. Both oxetanocin and albucidin give rise to a prominent ms^2 peak at m/z 136 .

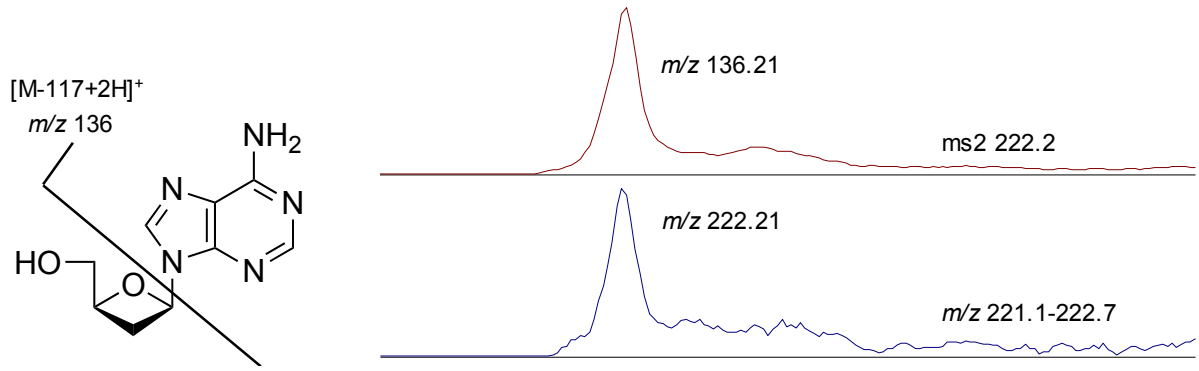


Figure 6.4 LC-MS/MS analysis of butanol extracts from a fermentation of *S. albus* subsp. *chlorinus* showing albucidin production. MS/MS analysis of albucidin produces a prominent ms^2 peak at m/z 136. .

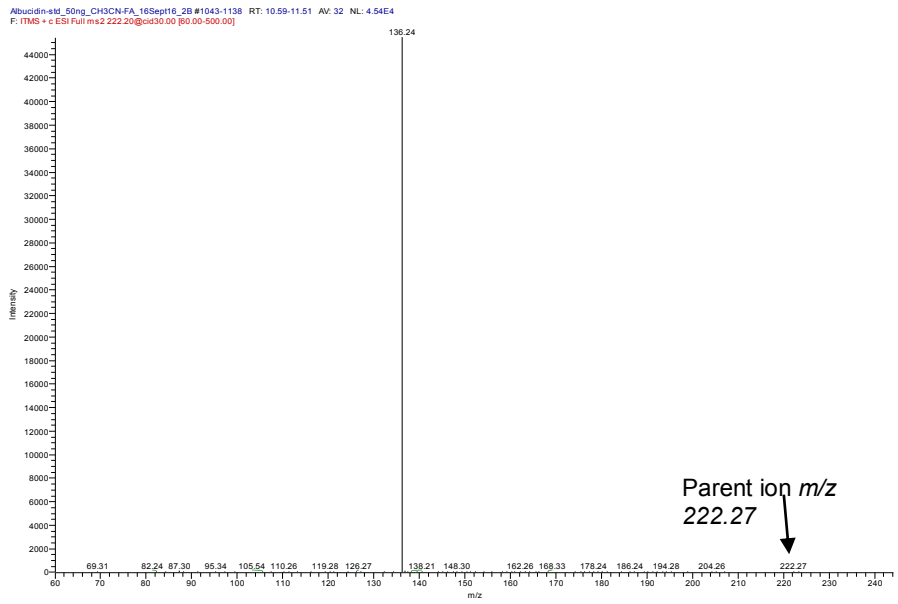


Figure 6.5 MS/MS analysis of albucidin. The arrow marks the position of the expected molecular ion. Production was very low so MS/MS analysis was vital for the identification of albucidin. A dominant peak at *m/z* 136 was seen.

6.2.2 Attempts to improve albucidin production

Single spore isolation was used in an initial attempt to identify an isolate that would give more consistent and higher production. Several rounds of single spore isolation were performed on the wild type strain and tested for albucidin production. Spores were collected and a serial dilution plated on oatmeal agar. Ten individual isolates were selected for each generation of spore isolation and these were tested by LC-MS/MS analysis of the supernatant after growth in medium G for 12 days. In total four generations of spore isolates were tested. However, no increase in the amount or the consistency of production of albucidin was observed.

A media screen was also undertaken in an attempt to improve albucidin production. Fermentation media used previously for oxetanocin and nikkomycin production were included (Shimada *et al*., 1986; Zeng *et al.*, 2002). Other liquid media investigated included a gentamicin production medium (which contains adenosine), medium G (the original medium used for albucidin production) (Hahn *et al.*, 2009), SFM, TSBY and Tap Water Medium. For media components, see section 2.2. None of the media tested resulted in increased production of albucidin compared to medium G. The production of albucidin in SFM was similar to that observed in medium G. SFM was later used for heterologous expression and was adopted as the routine production medium.

Production of albucidin on solid media was also tested, using SFM, TSB, ABB13, oatmeal, YS, MDM and VYL agar-based media. YS, MDM and VYL media had been previously tested for oxetanocin production (Shimada *et al*., 1986). Albucidin was detected in SFM medium, but production was lower than in liquid SFM and medium G cultures. No other media showed better production than SFM.

Because the chemical structures of albucidin and oxetanocin are so similar, and their biosynthetic pathways appear so similar, it was plausible that *S. albus* subsp. *chlorinus* would also produce oxetanocin. When analysing the fermentation of *S. albus* subsp. *chlorinus* for albucidin production, a molecular ion with m/z 252 was also observed, corresponding to the $[M+H]^+$ ion expected for oxetanocin. Fragmentation results obtained from LC-MS/MS analysis of this ion, and from an albucidin standard provided by Syngenta, further supported its assignment as oxetanocin, as both nucleosides yielded a dominant fragment ion of m/z 136, corresponding to protonated adenine. High resolution mass spectrometry (HRMS) confirmed the molecular ion as m/z 252.1096, corresponding to that expected (± 1.0 ppm) for protonated oxetanocin.

6.2.3 Analysis of the *S. albus* subspecies *chlorinus* genome for a homologue of OxsB

The *S. albus* subsp. *chlorinus* genome sequence was obtained in this laboratory and assembled by Dr Markiyan Samborskyy into an initial 11 scaffolds, in a genome sequence of 7.6 Mbp in total. This was the assembly that was initially examined to search for the albucidin genes. Subsequently this draft has been improved by Dr Samborskyy into essentially a single scaffold, with no alteration to the region of interest. The amino acid sequence of the OxsB protein was used to probe the *S. albus* subsp. *chlorinus* genome using BLAST searches. A single hit was obtained to an ORF that was annotated as a radical SAM enzyme. This ORF was tentatively named *acdB*. Blastx screening of the NCBI protein database using *acdB* returned OxsB as the top hit, as expected.

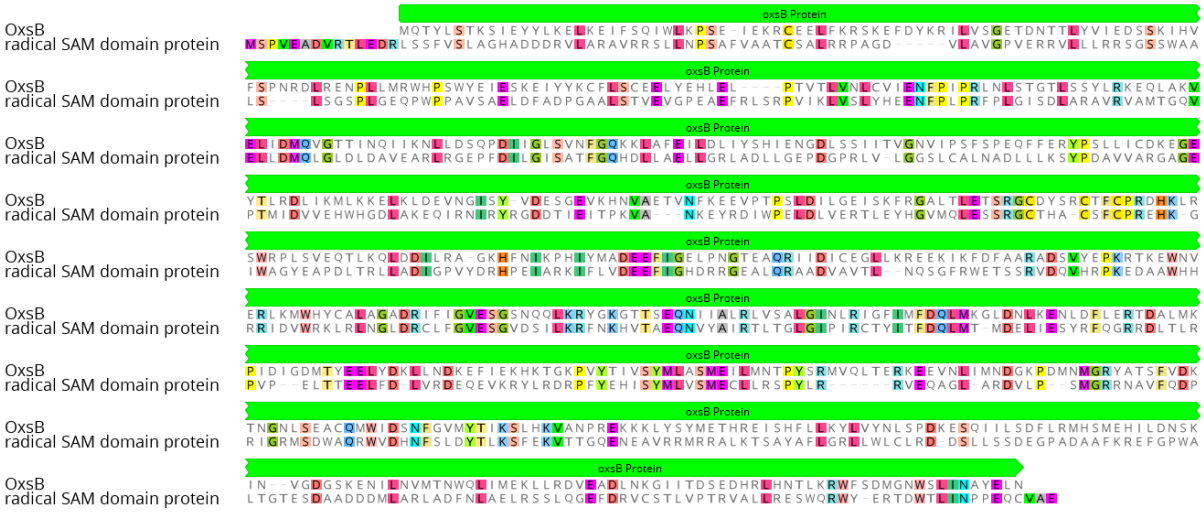


Figure 6.6 Alignment of OxsB and the putative radical SAM enzyme AcdB from *Streptomyces albus* subsp. *chlorinus*.

6.2.3 Annotation of a putative biosynthetic cluster (*acd*) for albucidin in the genome of *Streptomyces albus* subsp. *chlorinus* and comparison with oxetanocin genes

The genes flanking the *acdB* genes are listed in **Table 6.1**. These genes were included in a large region identified by AntiSMASH analysis as a cluster associated with an NRPS, but were not identified in their own right. Surprisingly at first sight, *acdB* is flanked by different genes to those flanking OxsB in *B. megaterium*.

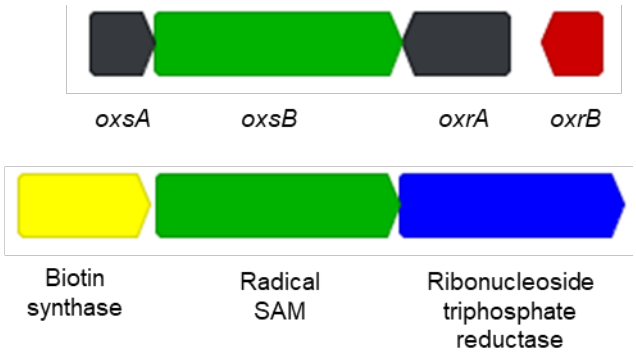


Figure 6.7 Comparison of the genes flanking the albucidin radical SAM AcdB below and the oxetanocin radical SAM OxsB above. The region of *S. albus* subsp. *chlorinus* DNA shown extends for 5, 498 bp.

Table 6.1 The putative albucidin cluster lies within an AntiSMASH-predicted region centred on an NRPS candidate. The predicted B_{12}-dependent radical SAM enzyme, situated in the middle of the cluster has been highlighted in yellow flanked either side by a putative biotin synthase (yellow) and a ribonucleoside triphosphate reductase (blue). There is a phosphohydrolase some distance away (red) but this is not a homologue of OxsA, the oxetanocin phosphohydrolase.

ORF	blast_score	Length	Product
SALBC_551	83.2	363	MazG nucleotide pyrophosphohydrolase
SALBC_550	192	1503	amino acid adenylation domain-containing protein
SALBC_549	186	606	hypothetical protein
SALBC_548	189	939	ribonucleoside-diphosphate reductase beta chain
SALBC_547	214	1644	AMP-dependent synthetase and ligase
SALBC_546		330	hypothetical protein
SALBC_545	50.8	816	acyl-CoA thioesterase
SALBC_544	100	1239	hypothetical protein
SALBC_543	137	840	putative S-adenosyl-L-methionine (SAM)-dependent methyltransferase
SALBC_542	49.7	225	phosphopantetheine-binding
SALBC_541	64.3	1218	amino acid adenylation domain protein
SALBC_540	49.3	249	non-ribosomal peptide synthetase
SALBC_539	293	2019	ribonucleoside-triphosphate reductase
SALBC_538	507	2217	radical SAM domain protein
SALBC_537	193	1191	Biotin synthase
SALBC_536		522	hypothetical protein
SALBC_535		993	hypothetical protein
SALBC_534	199	924	hypothetical protein
SALBC_533		2040	hypothetical protein
SALBC_532	68.9	594	thymidylate kinase-like protein
SALBC_531	106	720	hypothetical protein
SALBC_530		1038	hypothetical protein
SALBC_529		705	hypothetical protein
SALBC_528	283	1107	carboxylesterase
SALBC_527	35	549	hypothetical protein
SALBC_526	209	459	putative cytoplasmic protein
SALBC_525	71.6	837	hypothetical protein
SALBC_524	269	762	two component transcriptional regulator, LuxR family
SALBC_523	471	2535	ABC transporter integral membrane protein
SALBC_522	109	249	putative gualylate cyclase protein
SALBC_521	72.4	303	hypothetical protein
SALBC_520	245	696	hydrolase
SALBC_519	491	1575	putative integral membrane protein
SALBC_518	300	1116	secreted protein

6.2.4 Confirmation of the involvement of AcdB in the biosynthesis of albucidin

To confirm that the radical SAM enzyme AcdB was indeed involved in the biosynthesis of albucidin, an in-frame deletion mutant was made in the *acdB* gene. A knockout mutant was prepared via gene editing based on CRISPR, using the plasmid pCRISPomyces2 (Apr^r) (Cobb *et al*., 2015). The pCRISPomyces2 system allows easy introduction of gRNA through Golden Gate assembly and the introduction of template DNA via a designated *Xba*I site. A thiostrepton cassette was added to pCRISPomyces2 (Apr^R) because initial selection using apramycin (Cobb *et al*., 2015) gave false exconjugants.

The gRNA was designed to target the iron sulphur cluster of the radical SAM protein, while the template DNA was designed to introduce a 51 base pair in-frame deletion to excise the CX3CX2C motif required for iron sulfur cluster binding. The resulting plasmid, pAM51Δrad ($Thio^R$) containing the gRNA and the template DNA, was transferred into the methylation deficient *E. coli* strain ET12567 (Cam^R) (*dam*$^-$, *dcm*$^-$, *hsdM*$^-$) by chemical transformation. The

ET12567 strain contains the “helper” plasmid pUZ8002 (Kan^R) which provides the machinery needed for conjugation into *Streptomyces*. Successful transformants were selected on 2TY solid media containing all three antibiotics (Cam, Kan, Thio).

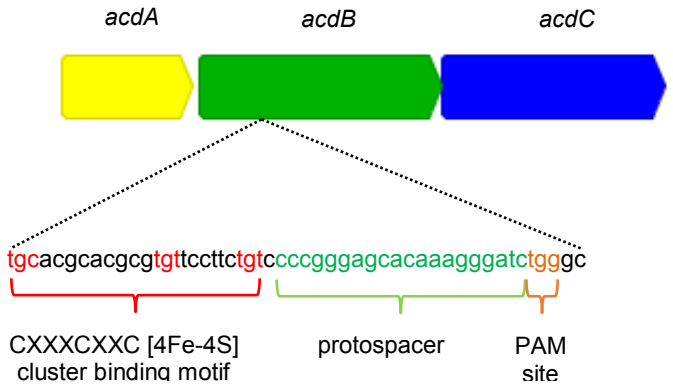


Figure 6.8 The design of the CRISPR experiment. gRNA targeted the [4Fe-4S] cluster, and the template DNA in the pCRISPomyces2 plasmid contained homology regions around the [4Fe-4S] cluster. The plasmid was designed by Dr Annabel Murphy.

pAM51Δrad was introduced into *S. albus* subsp. *chlorinus* via conjugation (section 2.43). After 18 hours growth, plates were overlaid with thiostrepton and nalidixic acid. Successful exconjugants were selected on SFM agar containing thiostrepton and nalidixic acid. Two further rounds of growth on SFM agar at 37 °C were intended to clear the plasmid from the strain. However false positive results were observed by PCR for the ex-conjugants at this stage, and further rounds of growth at 37 °C were required to allow removal of the vector. The construction and use of this knockout plasmid was performed with the help of Dr Annabel Murphy.

PCR analysis followed by Sanger sequencing confirmed that the desired knockout mutant had been obtained (Figure 6.9). This strain was then tested for production of albucidin via butanol extraction followed by HPLC MS/MS analysis. This confirmed that albucidin production had been abolished, supporting the essential role of *acdB* in the production of albucidin.

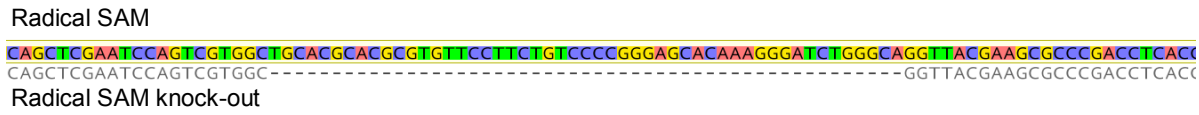


Figure 6.9 Confirmation of the in-frame deletion in *acdB*. Sequencing of the PCR products from deletion mutant and wild type confirmed that the *acdB* gene had been inactivated in-frame, with the loss of 51 bp. Fermentation and LC-MS/MS confirmed that albucidin production had been abolished.

6.2.5 Heterologous expression of the albucidin cluster

The advantages of heterologous expression have previously been discussed in Chapter 4. Here, confirmation of the cluster identity, improved production, clean metabolic background and clarification of the cluster boundaries are all attractive potential outcomes for heterologous expression. For heterologous expression of the albucidin cluster, three potential hosts were considered that have all been used successfully for the heterologous expression of natural products: *S. lividans* TK24, *S. coelicolor* M1154 and *S. albus* J1074. Of these, *S. albus* J1074 was the obvious choice as the most closely related strain to the wildtype albucidin producer strain *Streptomyces albus* subsp. *chlorinus*. No homologues of the albucidin genes were found in the genome sequence of *S. albus* J1074.

The strategy for heterologous expression was designed with help from Dr Annabel Murphy. Constructs based on various cluster boundaries were designed to identify the genes needed for albucidin production. The vector pIB139 (Apr^r) is a modified version of pSET152 (Bierman *et al*., 1992) incorporating a ΦC31 *attP* site for integration into the host genome. To create pIB139 (Wilkinson *et al.*, 2002), the *permE** promoter had been added in front of the multiple cloning site of pSET152 to allow robust transcription of the inserted cluster in various streptomycete hosts. Five constructs were designed, and PCR was used to amplify the region required (Figure 6.10). The PCR primers were designed with overlaps to allow for introduction into pIB139 (Apr^r) by Hot Fusion (*Fu et al.*, 2014). Following PCR and assembly, *E. coli* DH10B was transformed with the resultant Hot Fusion mixture, colonies were selected on 2TY agar containing apramycin and the incorporation of the correct insert was confirmed by PCR sequencing of the isolated plasmid. Work with plasmid F1R3 was discontinued after heterologous expression had been obtained with plasmid F1R2 (Figure 6.10).

Each of the plasmids was transformed into *E. coli* ET12567 (pUZ8002) (Cam^R, Kan^R). ET12567 is used to provide unmethylated DNA for strains such as *S. coelicolor* M1154 which have a restriction modification system that degrades incoming methylated DNA. All constructs were transferred by conjugation to TK24, M1154 and J1074 respectively. Later, conjugation involving *S. albus* J1074 was achieved using triparental mating. Here, DH10B containing the

correct plasmid was conjugated alongside DH10B containing the self-transmissible plasmid, pUB307 (Kan^R). J1074 can accept methylated DNA, and conjugates efficiently using the triparental method. Potential exconjugants were selected and restreaked on SFM solid medium containing apramycin. Validated exconjugants were fermented in SFM liquid medium for 7 days followed by analysis of the supernatant by LC-MS/MS.

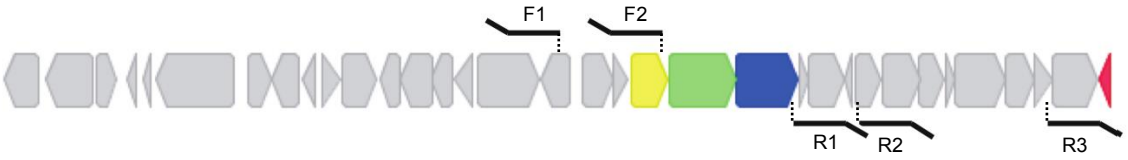


Figure 6.10 The design of primers to amplify the putative albucidin cluster. Constructs containing F1R1:Alb, F1R2:Alb, F2R1:Alb, and F2R2:Alb were created. The PCR product for F1R3 was attempted but after initial success with the smaller constructs, was discontinued. The region of DNA shown here extends for 30.7 kbp.

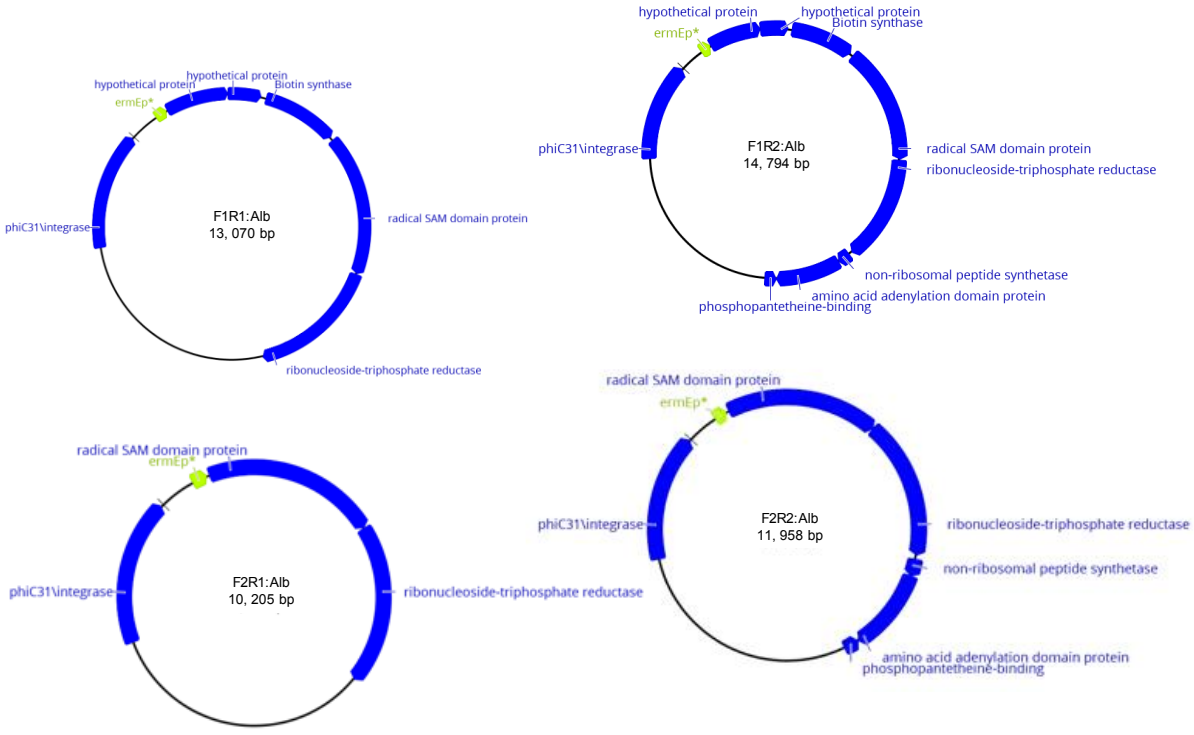


Figure 6.11 Design of plasmids to probe the biosynthetic gene cluster of albucidin. Each pIB139-based plasmid (Wilkinson *et al.*, 2002) houses a different complement of genes from the region around *acdB*, the homologue of *oxsB* in oxetanocin A biosynthesis in *B. megaterium*.

The heterologous host *S. albus* J1074 expressing either the F1R2:Alb or the F1R1:Alb construct was found to produce both albucidin and oxetanocin in approximately equal quantities, and at similar titres. Heterologous strains were grown at 30 ˚C in 7 ml TSBY media for 48 hours before 5 ml was transferred to a 250 ml flask containing 30 ml SFM. After 7 days further growth at 30 ˚C, the supernatant was analysed by LC-MS/MS.

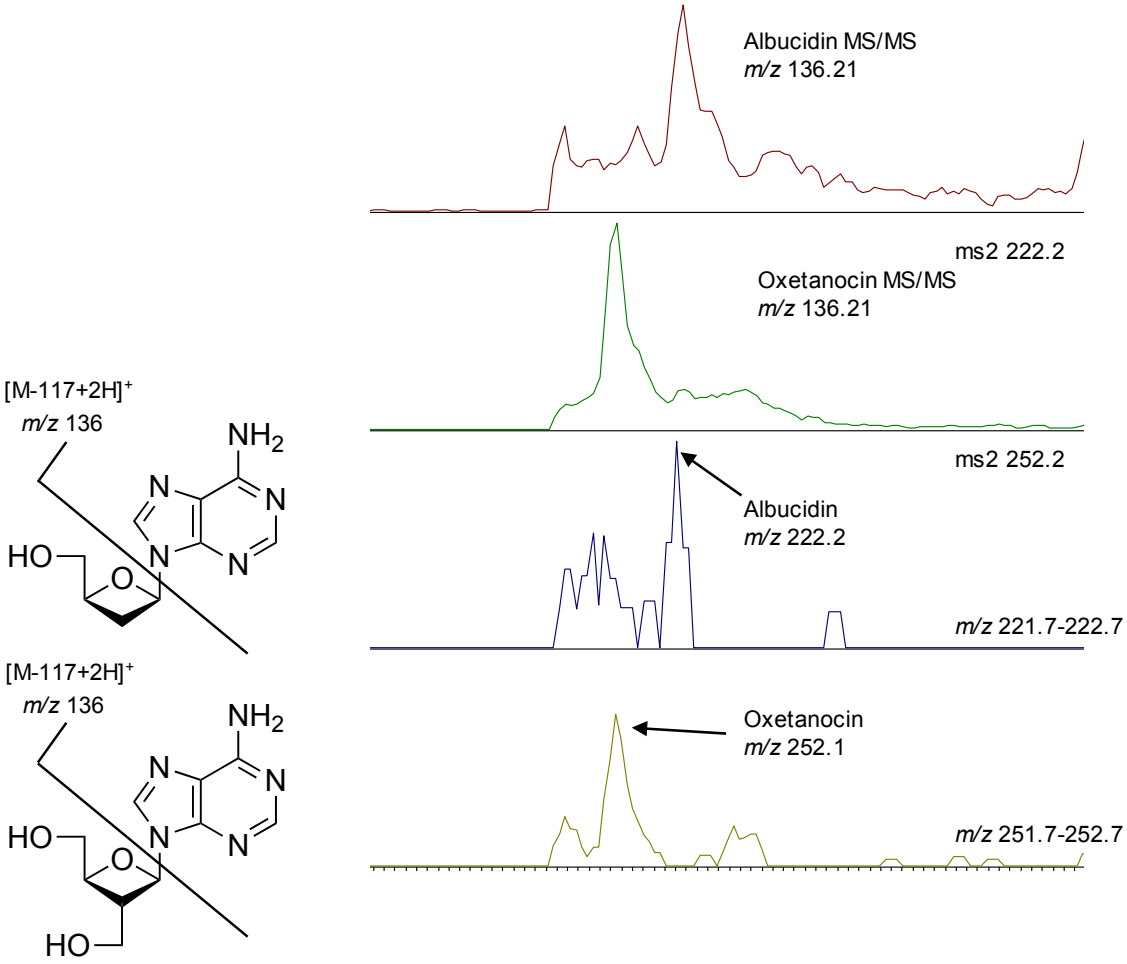


Figure 6.12 MS/MS data showing the production of oxetanocin and albucidin by J1074:F1R1

In contrast, the heterologous expression of either F2R1-Alb or F2R2-Alb resulted in the detection of oxetanocin, but not of albucidin.

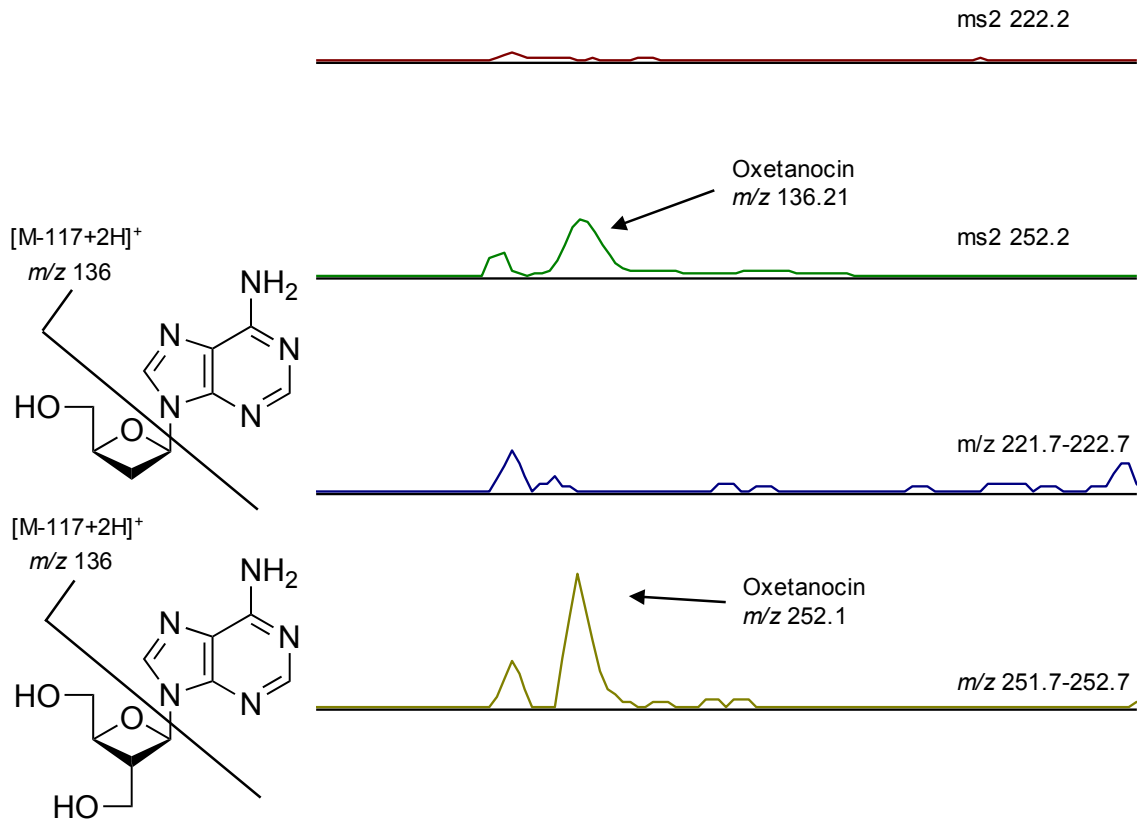


Figure 6.13 MS/MS analysis of the supernatant of J1074:F2R2. Both J1074:F2R1 and J1074:F2R2 showed production of oxetanocin but not albucidin. Neither plasmid contained the gene annotated as a biotin synthase.

6.2.7 Media screen for albucidin production in *S. albus* J1074

A media screen to test for improved production of albucidin in the heterologous host was carried out. A seed culture of J1074:F1R1 was prepared using 7 ml TSBY in a baffled flask containing metal springs for 48 hours. This seed culture (5 ml) was then added to 50 ml media in a 250 ml flask containing springs and incubated at 30 ˚C for 7 days. The supernatant was then examined for the presence of albucidin by LC-MS/MS.

The media tested were the same as those tested with the native producer *S. albus* subsp. *chlorinus* (section 5.2.2). The heterologous host was found to produce albucidin when

grown in Tap Water medium, although no albucidin had been detected for this medium using the native producer. Otherwise, only decreased production of albucidin was observed when compared to the SFM.

6.2.8 Albucidin production in recombinant *S. albus* J1074 housing the entire albucidin region identified by AntiSMASH

From the results of heterologous expression using the plasmids of Figure 5.12 as described in section 5.2.8, and especially the production of albucidin by *S. albus* J1074:F1R1, it became possible to narrow down the genes essential to produce albucidin and oxetanocin. The experiments also demonstrate that oxetanocin production requires only two genes: a radical SAM enzyme and a ribonucleoside triphosphate reductase, *acdB* and *acdC* respectively. The production of albucidin requires one or more of the three genes upstream of *acdB/C*, and it seems clear that the biotin synthase (*acdA*) is likely to play a key role.

It was nevertheless possible that additional neighbouring genes apart from the five genes tested in F1R1:Alb do contribute in some way to the albucidin pathway. Production of albucidin in the heterologous strain was higher than that of the native producer, but still low at around 2 mg/l. To check the potential influence of flanking genes in the S. albus J1074 background, direct *in vitro* capture of the entire cluster identified by AntiSMASH was performed using the Isothermal assembly capture technique devised in Chapter 4.

Genomic DNA from *S. albus* subsp. *chlorinus* was cut with the restriction enzymes HindIII and MfeI to create a 52,154 bp fragment. The vector pIB139 was PCR amplified with primers containing 40 bp overlaps relating to the restriction enzyme cut end of the fragment. After isothermal assembly, the isothermal assembly mixture was used to transform *E. coli* DH10B by electroporation and then colonies were selected on 2TY agar containing apramycin. Colonies were tested by PCR and two colonies out of 50 were found to contain the appropriate insert. The resulting plasmid was named pFD54:Alb.

Intergenic transfer of pFD54:Alb (Apr^R) into J1074 was by triparental mating using DH10B (pUB307) (Kan^R) and overlaid with apramycin and nalidixic acid. Exconjugants were restreaked on selective SFM agar. Successful exconjugants were inoculated in TSBY media and then grown in liquid SFM and TSBY as fermentation media. While albucidin and oxetanocin were both observed, no increase in production was seen compared to the smaller constructs and no further compounds were observed. No further genes outside of the 5 identified genes appear to contribute to albucidin production under these conditions.

6.2.9 Heterologous expression of *acdB* alone in *S. albus* J1074 is sufficient for oxetanocin production

At this stage the close relationship between oxetanocin and albucidin biosynthesis was firmly established. A plausible scheme would involve initial action by the ribonucleoside reductase AcdC on a metabolite in primary metabolism, either the mono-, di- or triphosphate of adenosine, to create a 2'-deoxyadenosyl phosphate, and this would undergo ring contraction catalysed by AcdB to give a precursor to both oxetanocin and albucidin. The oxetanocin cluster in B. megaterium does not house a ribonucleoside triphosphate reductase equivalent to *acdC*, so it seemed possible that *acdC* might also be dispensed with for albucidin production in *S. albus* subsp. *chlorinus* and *S. albus* J1074. Therefore, *acdB* was expressed on its own in J1074. The construct was designed using PCR to amplify the *acdB* gene and insert this into pIB139 under the control of the constitutive *permE** promoter. Conjugation was carried out as above and successful exconjugants restreaked, grown in TSBY and then fermented in liquid SFM.

Oxetanocin production was observed in this heterologous host, demonstrating that the nucleoside reductase *acdC* is not required *in vivo* to produce oxetanocin. It is likely that the 2'deoxygenation step is catalysed by another ribonucleoside reductase gene encoded on the *S. albus* chromosome.

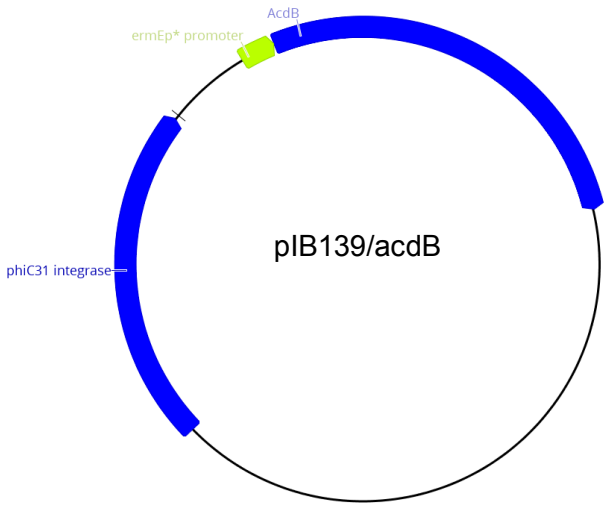


Figure 6.14. The plasmid for expression of the putative ring contraction enzyme AcdB in S. albus J1074

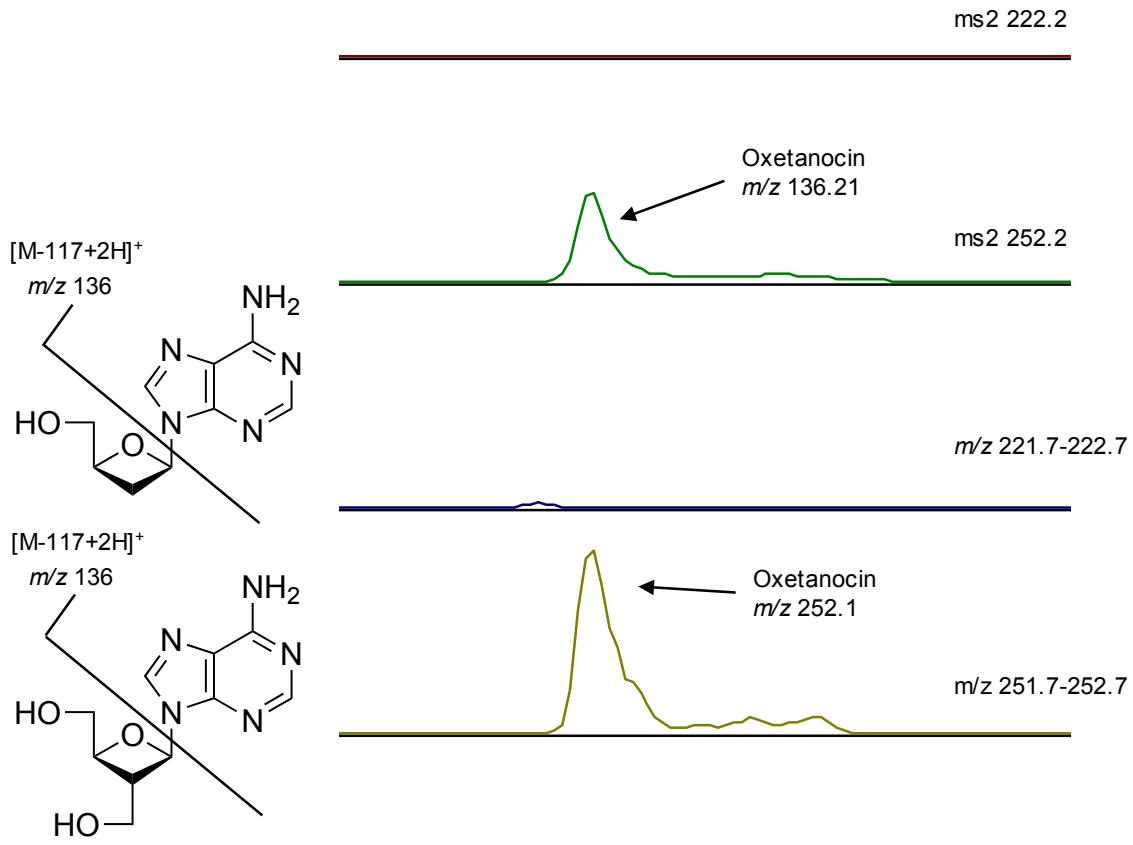


Figure 6.15 MS/MS analysis of the supernatant of J1074:AcdB showing production of oxetanocin.

6.3 Concluding remarks

The work described in this chapter has identified the biosynthetic gene cluster for the remarkable nucleoside albucidin in *S. albus* subsp. *chlorinus*, and defined three key genes which all appear to encode remarkable enzymes utilising free radical chemistry: first, the ribonucleoside reductase AcdC carries out the well-understood step of 2'-deoxygenation in the ribose ring; then ring contraction is catalysed by the B_{12}-dependent radical SAM enzyme AcdB; and finally the removal of the hydroxymethyl sidechain which differentiates albucidin from oxetanocin is catalysed by the "biotin synthase" AcdA.

As these experiments were being concluded, an important contribution to our understanding of the mechanism of the OxsB-catalysed ring contraction, and the structure of the enzyme, was published by Hung-wen (Ben) Liu and Cathy Drennan and their colleagues (Bridwell-Rabb *et al*., 2017). They used a synthetic gene for OxsB expressed as a fusion

protein in *E. coli* to establish the nature of the substrate for this enzyme, and managed to recapitulate the ring-contraction process *in vitro*. After seven years of work, and three years after the completion of the X-ray crystal structure of OxsB in the Drennan laboratory (H.-w. Liu, personal communication via P. F. Leadlay), they were able to determine the substrate specificity of OxsB, and to define the *in vitro* products, and could propose the plausible mechanism shown in Figure 6.16 (Bridwell-Rabb et al., 2017). Conditions had to be carefully optimised to preserve the activity of the enzyme with its embedded 4Fe-4S iron sulfur cluster. They showed that *in vitro*, 2'deoxyadenosine mono-, di- and triphosphates are all competent substrates for purified recombinant OxsB. When OxsB was incubated with dAMP, an aldehyde form of oxetanocin 5'-monophosphate was the observed product. The authors proposed that the resulting aldehyde group is reduced by a cellular component, most likely an alcohol dehydrogenase, to produce oxetanocin (Bridwell-Rabb *et al*., 2017).

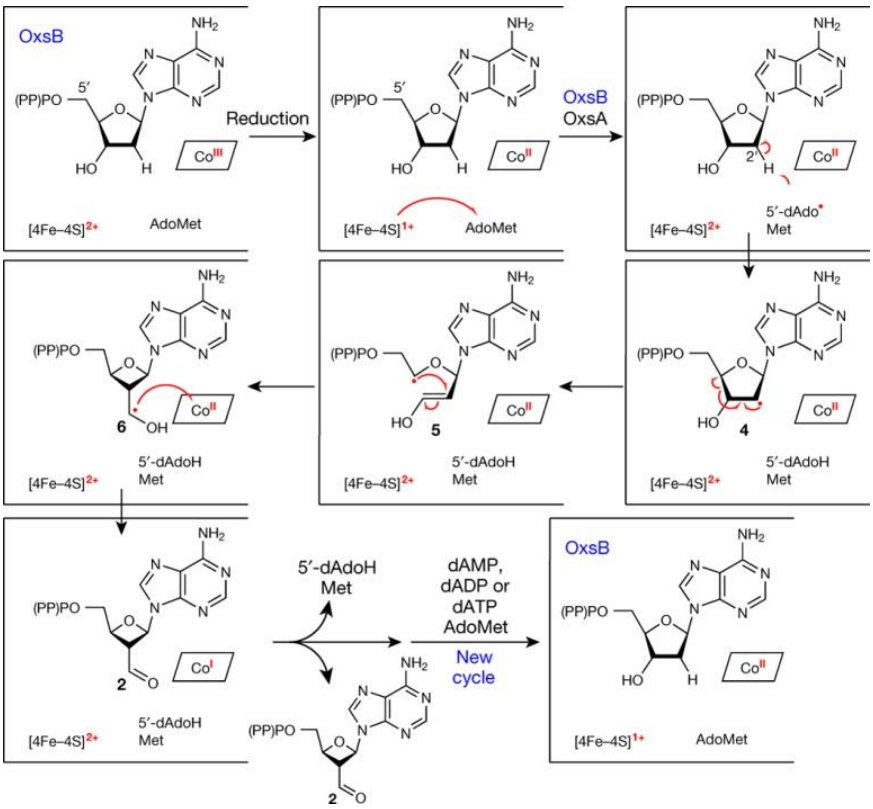


Figure 6.16 The mechanism of ring contraction catalysed by OxsB in oxetanocin biosynthesis (adapted from Bridwell-Rabb *et al*., 2017). The reaction is initiated by using the embedded iron-sulfur cluster to reductively cleave the carbon-sulfur bond of SAM and generate the reactive 5-deoxyadenosyl radical.

Separately, these authors have also presented the X-ray crystal structure of the metalloenzyme OxsA (Bridwell-Rabb et al., 2016), and have confirmed its activity as a phosphohydrolase that can successively remove phospho groups from a nucleoside

triphosphate to give the nucleoside. The active site is apparently adapted to be selective for the oxetan nucleoside triphosphate over larger substrates.

The mechanism proposed by Liu and colleagues for oxetanocin biosynthesis provides a satisfying explanation for the production of the remarkable oxetan ring system, and it is highly probable that AcdB catalyses the same reaction. No albucidin has been reported as being produced *in vitro* in the experiments with OxsB, and the evidence of the experiments in this chapter suggests that AcdA is the enzyme that must divert the aldehyde intermediate of the mechanism in Figure 6.16 towards the removal of the sidechain, and the production of albucidin. The appearance of oxetanocin as a significant side-product during the heterologous production of albucidin in J1074 suggests that optimisation of the AcdA-catalysed reaction is a key goal to increase albucidin production in that host strain. In addition, the experiments presented in this chapter provide a platform for future *in vitro* exploration of the mechanism of this reaction, using enzymes heterologously expressed either in *S. alb*us J1074 or in *E. coli*.

The radical SAM superfamily of enzymes is easily one of the largest known, consisting of more than 16,000 members in all living organisms (Frey *et al*., 2008; Landgraf *et al*., 2016). Although the annotation of AcdA is as a biotin synthase, detailed *in vitro* functional and structural studies will be needed to define the exact mechanism. Biotin synthesis from the dethiobiotin precursor catalysed by biotin synthase involves cleavage of two unactivated C-H bonds and the insertion of sulfur, apparently derived from an iron-sulfur cluster (Fugate and Jarrett, 2012). Perhaps one pointer towards the role of AcdA is that a number of B_{12}-dependent radical SAM enzymes are known that catalyse decarboxylation, such as the coproporphyrinogen III oxidase HemN, which catalyzes protoporphyrinogen IX production by sequential decarboxylation of the two propionate side chains of coproporphyrinogen III (Layer *et al*., 2006), and BlsE in the biosynthetic pathway to the peptidyl nucleoside antibiotic produced by *Streptomyces griseochromogenes* (Liu *et al*., 2017).

Chapter 7 Conclusions and suggestions for future work

The work described in this thesis has addressed three different aspects of the current paradigm for discovery, analysis and manipulation of pathways to specialised metabolites in *Streptomyces* bacteria. The chief conclusions that can be drawn from the genome mining of the two *S. malaysiensis* strains DSM4137 and DSM14702, described in the first part of Chapter 3, are first, that these strains are strikingly rich in biosynthetic potential; and secondly, that an essentially complete genome sequence allows a significant saving in time and effort in detecting, annotating and decoding biosynthetic gene clusters, compared to the still-widespread use of fragmented genome sequences for initial searches. This is particularly true for clusters that are organised in new ways, or involve novel combinations of enzymes. The use of complete genome sequence also improves the accuracy of automated analysis using programs such as antiSMASH. A further conclusion to be drawn is that such programs cannot be used uncritically: not only are some target clusters still not detected in the latest version (for example the malayamycin cluster in DSM14702, Dr H. Hong, personal communication), but the program tends to predict wider limits than would be allocated manually. It is also interesting that some clusters have been found, such as that for padanamide, that were first described from strains isolated from oceanic sources.

The transcriptomic analysis described in Chapters 3 and 4 has deliberately focused on selected clusters, that appeared best to answer the chief questions being asked. The first question was whether it is possible to distinguish a cluster being expressed, perhaps at low levels, from one that is 'silent', and the data in these Chapters encourage the view that it is. Secondly, it was important to determine whether the output of the RNAseq analysis is consistent with the behaviour of known clusters whose products have been previously characterised. For most of the clusters examined in Chapter 3, this seems to be true. For a few clusters, the relatively early sampling period during the fermentation and the use of a single medium probably underestimates the potential for that cluster to be expressed. If the cost of RNAseq continues to fall, such transcriptomic analyses can be undertaken on a more ambitious scale, and extended (for example) to analysing alterations in the transcriptome caused by genetic manipulation of global regulators or CSRs.

The work in Chapter 5 on the heterologous expression of the spinosyn cluster led to the further development of a direct cloning method first used in this laboratory by Dr Yongjun Zhou. It was shown here that a ~90 kbp spinosyn cluster was successfully 'captured' as an insert in a PAC vector by *in vitro* Gibson isothermal assembly. This approach is sufficiently

flexible that it makes a worthwhile alternative to yeast-based methods for cluster capture. Clearly, given my results and those of two recently published reports using essentially the same methods, of the modest levels of heterologous expression of the spinosyn genes in *S. erythraea*, much remains to be learnt about the regulation of this cluster and of biosynthetic gene clusters in *Saccharopolyspora* generally (neither spinosyn nor erythromycin clusters have CSRs). It has to be admitted that at present rational approaches to titre improvement in heterologous host strains starting from wild type strains are not yet faring well compared to conventional "mutate and select" campaigns of strain improvement in the original producer. In contrast, further improvement of an already over-producing industrial strain by targetted approaches has been highly effective, as shown in recent work on increasing avermectin production by *S. avermitilis* (Zhuo et al., 2014). The aim of future work on spinosyn heterologous expression remains, instead, to obtain a level of expression that allows engineering of the biosynthesis to be carried out in this background and for successful structural modifications later to be copied in the harder-to-manipulate original strain.

The work in Chapter 6 has uncovered and characterised the gene cluster for the unusual and potentially useful nucleoside albucidin. The cluster is small and has been successfully expressed in a heterologous strain. Further work is clearly warranted on AcdA, the fascinating radical SAM enzyme that catalyses the removal of a hydroxymethyl group from the oxetane ring, to establish the identity of the substrate and to determine the mechanism; and if possible to obtain X-ray crystal structure, as done recently for the oxetanocin ring-contracting enzyme OxsB by Liu and colleagues. A second aim must be to attempt to increase the titre, and this would be helped if the entire albucidin pathway could be reconstituted *in vitro* using purified recombinant enzymes, and the true substrate for this pathway (not necessarily the same as for OxsB) were identified

References

Achenbach, H., Mühlenfeld, A., Fauth, U., and Zähner, H. (1988) The galbonolides. Novel, powerful antifungal macrolides from *Streptomyces galbus* ssp. *eurythermus*. *Ann. N. Y. Acad. Sci.* **544**, 128–140.

Ahlert, J., Shepard, E., Lomovskaya, N., Zazopoulos, E., Staffa, A., Bachmann, B. O., Huang, K., Fonstein, L., Czisny, L., Whitmam, R. E., Farnet, C. M., and Thorson, J. S. (2002) The calicheamicin gene cluster and its iterative type I enediyne PKS. *Science* **297**, 1173–1176.

Albright, J. C., Goering, A.W., Doroghazi, J. R., Metcalf, W. W., and Kelleher, N. L. (2014) Strain-specific proteogenomics accelerates the discovery of natural products via their biosynthetic pathways. *J. Ind. Microbiol. Biotechnol.* **41**, 451–459.

Aldridge, S., Parascandola, J., and Sturchio, J. L. (1999). The discovery and development of penicillin 1928-1945. American Chemical Society and Royal Society of Chemistry.

Altschul, S. F. Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.

Anarat-Cappillino, G., and Sattely, E. S. (2014) The chemical logic of plant natural product biosynthesis. *Curr. Opin. Plant Biol.* **19**, 51–58.

Angert, E. R. (2005) Alternatives to binary fission in bacteria. *Nat. Rev. Microbiol*. **3**, 214–224.

Aoyagi, T., Aoyama, T., Kojima, F., Hattori, S., Honma, Y, Hamada, M., and Takeuchi, T. (1992) Cyclooctatin, a new inhibitor of lysophospholipase, produced by Strepto*myces melanosporofaciens* MI614-43F2. Taxonomy, production, isolation, physico-chemical properties and biological activities. *J. Antibiot*. **45**, 1587–1591.

Banik, J. J., and Brady,S. F. (2010) Recent application of metagenomic approaches toward the discovery of antimicrobials and other bioactive small molecules. *Curr. Opin. Microbiol.* **13**, 603–609.

Bandarian, V. (2012) Radical SAM enzymes involved in the biosynthesis of purine-based natural products. *Biochim. Biophys. Acta* **1824**, 1245–1253.

Barka, E. A., Vatsa, P., Sanchez, L., Gaveau-Vaillant, N., Jacquard, C., Klenk, H. P., Clément, C., Ouhdouch, Y. and van Wezel, G. P. (2016) Taxonomy, physiology, and natural products of actinobacteria. *Microbiol. Mol. Biol. Rev*. **80**, 1–43.

Barone, R., De Santi, C., Palma Esposito, F., Tedesco, P., Galati, F., Visone, M., Di Scala, A. and De Pascale, D. (2014). Marine metagenomics, a valuable tool for enzymes and bioactive compounds discovery. *Front. Mar. Sc*i, 1.

Bedford D. J., Lewis, C. G., and Buttner, M. J. (1991) Characterization of a gene conferring bialaphos resistance in *Streptomyces coelicolor* A3(2). *Gene* **104**, 39–45.

Behnken, S., and Hertweck, C. (2012). Anaerobic bacteria as producers of antibiotics. *Appl. Microbiol. Biotechnol*. **96**, 61–67.

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**, 141–147.

Bevitt, D. J., Cortes, J., Haydock, S. F., and Leadlay, P. F. (1992) 6-deoxyerythronolide B synthase-2 from *Saccharopolyspora erythraea*. Cloning of the structural gene, sequence analysis and inferred domain structure of the multifiunctional enzyme. *Eur. J. Biochem.* **204**, 39–49.
*Bevitt, D.J., Staunton, J., and Leadlay, P.F. (1993) Mutagenesis of the dehydratase active site in the erythromycin-producing polyketide synthase. Biochem. Soc. Trans. **21**, 30S.*

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

Bierman, M., Logan, R., O'Brien, K., Seno, E.T., Rao, R.N. and Schoner, B.E. (1992). Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. *Gene* **116**, 43–49.

Bister, B., Bischoff, D., Ströbele, M., Riedlinger, J., Reicke, A., Wolter, F., Bull, A. T., Zähner, H., Fiedler, H. P., and Süssmuth, R. D. (2004) Abyssomicin C: a polycyclic antibiotic from a marine *Verrucosispora* strain as an inhibitor of the p-aminobenzoic acid/tetrahydrofolate biosynthesis pathway. *Angew. Chem. Int. Ed.* **43**, 2574–2576.

Blin, K., Medema, M. H., Kazempour, D., Fischbach, M. A., Breitling, R., Takano, E., and Weber, T. (2013) AntiSMASH 2.0--a versatile platform for genome mining of secondary metabolite producers. *Nucl. Acids Res*. **41**(Web Server issue), W204–212.

Blin, K., Wolf, T., Chevrette, M.G., Lu, X., Schwalen, C.J., Kautsar, S.A., Suarez Duran, H.G., Los Santos, E.L.C. de, Kim, H.U., Nave, M., Dickschat, J.S., Mitchell, D.A., Shelest, E., Breitling, R., Takano, E., Lee, S.Y., Weber, T. and Medema, M.H. (2017). AntiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucl. Acids Res*. **45**, W36¬W41.

Boddy, C. N. (2014) Bioinformatics tools for genome mining of polyketide and non-ribosomal peptides. *J. Ind. Microbiol. Biotechnol.* **41**, 443-50.

Bode, H. B. (2009) Entomopathogenic bacteria as a source of secondary metabolites. *Curr. Opin. Chem. Biol.* **13**, 224–230.

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

Boecker, S., Zobel, S., Meyer, V., and Süssmuth, R. D. (2016) Rational biosynthetic approaches for the production of new-to-nature compounds in fungi. *Fungal Genet. Biol.* **89**, 89–101.

Brady, S. F., Wright, S. A., Lee, J. C., Sutton, A. E., Zumoff, C. H., Wodzinsky, R. S., Beer, S.V., and Clardy, J. (1999) Pantocin B, an antibiotic from *Erwinia herbicola* discovered by heterologous expression of cloned genes. *J. Am. Chem. Soc.* **121**, 11912–11913.

Brahmachari, G. (ed.) (2011) *Bioactive Natural Products: Opportunities and Challenges in Medicinal Chemistry*. World Scientific Publishing Co., Singapore.

Bridwell-Rabb, J., Kang, G., Zhong, A., Liu, H.-w., and Drennan, C. L. (2016) An HD domain phosphohydrolase active site tailored for oxetanocin-A biosynthesis. *Proc. Natl. Acad. Sci USA* **113**, 13750–13755.

Bridwell-Rabb, J., Zhong, A., Sun, H. G., Drennan, C. L., and Liu, H. W. (2017)

A B12-dependent radical SAM enzyme involved in oxetanocin A biosynthesis.

Nature **544**, 322–326.Brown, E. D., and Wright, G. D. (2016) Antibacterial drug discovery in the resistance era. *Nature* **529**, 336–43.

Bu'Lock, J. D. (1965) The biosynthesis of natural products. McGraw-Hill, Maidenhead, UK.

Bumpus, S. B., and Kelleher, N.L. (2008) Accessing natural product biosynthetic processes by mass spectrometry. *Curr. Opin. Chem. Biol*.**12**, 475–482.

Bumpus, S. B., Evans, B. S., Thomas, P. M., Ntai, I., and Kelleher, N. L. (2009) A proteomics approach to discovering natural products and their biosynthetic pathways. *Nat. Biotechnol.* **27**, 951–956.

Caffrey, P. (2003) Conserved amino acid residues correlating with ketoreductase stereospecificity in modular polyketide synthases. *ChemBiochem*. **4**, 654–657.

Caffrey, P., Bevitt, D., Staunton, J., and Leadlay, P.F. (1992). Identification of DEBS 1, DEBS 2 and DEBS 3, the multienzyme polypeptides of the erythromycin-producing polyketide synthase from *Saccharopolyspora erythraea*. *FEBS Lett*. **304**, 225–228.

Callahan, B., Mukund, T., and Shraiman, B. I. (2009) Emergent gene order in a model of modular polyketide synthases. *Proc. Natl. Acad. Sci. USA* **106**, 19410–19415.

Cane D. E., and Ikeda, H. (2012) Exploration and mining of the bacterial terpenome. *Acc. Chem. Res.* **45**, 463-472.

Cantrell, C. L., Dayan, F. E., and Duke, S. O. (2012) Natural products as sources for new pesticides. *J. Nat. Prod*. **75**, 1231–1242.

Carney, J., Santi, D. V., Hutchinson, C. R., and McDaniel, R. (2003) A model of structure and catalysis for ketoreductase domains in modular polyketide synthases. *Biochemistry* **42**, 72–79.

Carver, T., Harris, S. R., Berriman, M., Parkhill, J., and McQuillan, J. A. (2012) Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* **28**, 464–469.

Casida, J. E., and Durkin, K. A. (2017) Pesticide chemical research in toxicology: lessons from Nature. *Chem. Res. Toxicol*. **30**, 94–104.

Challis, G. L., and Ravel, J. (2000) Coelichelin, a new peptide siderophore encoded by the *Streptomyces coelicolor* genome: structure prediction from the sequence of its non-ribosomal peptide synthetase. *FEMS Microbiol. Lett.* **187**, 111–114.

Challis, G. L. (2014) Exploitation of the *Streptomyces coelicolor* A3(2) genome sequence for discovery of new natural products and biosynthetic pathways. *J. Ind. Microbiol. Biotechnol*. **41**, 219–232.

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

Challis, G. L., Ravel, J., and Townsend, C. A. (2000) Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem. Biol*. **7**, 211–224.

Chandra, C. and Chater, K. F. (2012) Developmental biology of Streptomyces from the perspective of 100 actinobacterial genome sequences. *FEMS Microbiol. Rev.* **38**, 345–379.

Chang, F. Y., and Brady, S. F. (2013) Discovery of indolotryptoline antiproliferative agents by homology-guided metagenomic screening. *Proc. Natl. Acad. Sci. USA* **110**, 2478-2483.

Chater, K. F. (2006) Streptomyces inside-out: a new perspective on the bacteria that provide us with antibiotics. *Phil. Trans. Roy. Soc. B: Biol. Sci.* **361**, 761–768.

Chen, J., Xia, H., Dang, F., Xu, Q., Li, W. & Qin, Z. (2015). Characterization of the chromosomal integration of *Saccharopolyspora* plasmid pCM32 and its application to improve production of spinosyn in *Saccharopolyspora spinosa*. *Appl. Microbiol. Biotechnol.*, **99**, 10141–10149.

Chen, L., Wu, D., Cai, X., Guo, F., Blackall, P.J., Xu, X. & Chen, H. (2012). Electrotransformation of *Haemophilus parasuis* with in vitro modified DNA based on a novel shuttle vector. *Vet. Microbiol*., **155**, 310–316.

Chen, S., Kinney, W. A, and Van Lanen, S. (2017) Nature's combinatorial biosynthesis and recently engineered production of nucleoside antibiotics in *Streptomyces*. *World J. Microbiol. Biotechnol.* **33**, 66.

Chen, X. H., Vater, J., Piel, J., Franke, P., Scholz, R., Schneider, K., Koumoutsi, A., Hitzeroth, G., Grammel, N., Strittmatter, A. W., Gottschalk, G., Süssmuth, R. D., and Borriss, R. (2006) Structural and functional characterization of three polyketide synthase gene clusters in *Bacillus amyloliquefaciens* FZB 42. *J. Bacteriol.* 188, 4024–4036.

Chen, Y., McClure, R. A., Zheng, Y., Thomson, R. J., and Kelleher, N. L. (2013) Proteomics guided discovery of flavopeptins: anti-proliferative aldehydes synthesized by a reductase domain-containing non-ribosomal peptide synthetase. *J. Am. Chem. Soc.* **135**, 10449–10456.

Cheng, Y. Q., Coughlin, J. M., Lim, S. K., and Shen, B. (2009) Type I polyketide synthases that require discrete acyltransferases. *Meth. Enzymol.* **459**, 165-186.

Cheng, Y., Tang, G. L., and Shen, B. (2003) Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis. *Proc. Natl Acad. Sci. USA* **100**, 3149–3154.

Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters.

Cimermancic, P., Medema, M. H., Claesen, J., Kurita, K., Wieland Brown, L. C., Mavrommatis, K., Pati, A., Godfrey, P. A., Koehrsen, M., Clardy, J., Birren, B. W., Takano, E., Sali, A., Linington, R. G., and Fischbach, M. A. (2014) Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **158**, 412–421

Claesen, J., and Bibb, M. (2010) Genome mining and genetic analysis of cypemycin biosynthesis reveal an unusual class of post-translationally modified peptides. *Proc. Natl. Acad. Sci. USA* **107**, 16297–16302.

Cobb, R.E., Wang, Y. and Zhao, H. (2015). High-efficiency multiplex genome editing of *Streptomyces* species using an engineered CRISPR/Cas System. *ACS Synth. Biol.* **4**, 723–728.

Combes, P., Till, R., Bee, S. and Smith, M.C. (2002). The *Streptomyces* genome contains multiple pseudo-attB sites for the φC31-encoded site-specific recombination system. *J. Bacteriol*. **184**, 5746–5752.

Conti, E. Stachelhaus, T., Marahiel, M. A., and Brick, P. (1997) Structural basis for

control of antibiotic biosynthesis. *In*: Wang, X., Chen, J., and Quinn, P. (eds) *Reprogramming microbial metabolic pathways*. Springer Verlag, Berlin. pp. 115–138.

*Cortés, J., Haydock, S.F., Roberts, G.A., Bevitt, D.J. and Leadlay, P.F. (1990). An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of Saccharopolyspora erythraea. Nature **348**, 176–178.*

Cortés, J., Velasco, J., Foster, G., Blackaby, A. P., Rudd, B. A., and Wilkinson, B. (2002) Identification and cloning of a type III polyketide synthase required for diffusible pigment biosynthesis in *Saccharopolyspora erythraea*. *Mol. Microbiol.* **44**, 1213–1224.

Cox R. J. (2007) Polyketides, proteins and genes in fungi: programmed nano-machines begin to reveal their secrets. *Org. Biomol. Chem.* **5**, 2010–2026.

Cragg, G. M., and Newman, D. J. (2013) Natural products: a continuing source of novel drug leads. *Biochim. Biophys. Acta* 1**830**, 3670-3695.

Croucher, N. J. and Thomson, N. R. (2010) Studying bacterial transcriptomes using RNA-seq. *Curr. Opin. Microbiol.* **13**, 619–624.

Crouse, G. D., Sparks, T. C., Schoonover, J., Gifford, J., Dripps, J., Bruce, T., Larson, L. L., Garlich, J., Hatton, C., Hill, R. L., Worden, T. V., and Martynow, J. G. (2001) Recent advances in the chemistry of spinosyns. *Pest Manag. Sci.* **5**7, 177–185.

Daeuble, J., Sparks, T.C., Johnson, P., and Graupner, P.R. (2009). Modification of the butenyl-spinosyns utilizing cross-metathesis. *Bioorg. Med. Chem*. **17**, 4197–4205.

Dairi, T. (2005) Studies on biosynthetic genes and enzymes of isoprenoids produced by actinomycetes. *J. Antibiot.* **58**, 227-243.

Daniel, R. (2005) The metagenomics of soil. *Nat. Rev. Microbiol*. **3**, 470–478.

Daum, M., Peintner, I., Linnenbrink, A., Frerich, A., Weber, M., Paululat, T., and Bechthold, A. (2009) Organisation of the biosynthetic gene cluster and tailoring enzymes in the biosynthesis of the tetracyclic quinone glycoside antibiotic polyketomycin. *ChemBiochem.* **10**, 1073–83.

Davies, J. (2006) Where have all the antibiotics gone? *Can. J. Infect. Dis. Med. Microbiol.* **17**, 287–290.

Davies, J., von Ahsen, U., Wank, H., and Schroeder, R. (1992) Evolution of secondary metabolite production: potential roles for antibiotics as prebiotic effectors of catalytic RNA reactions. *Ciba Found. Symp.* **171**, 24–32.

Dawlaty, J., Zhang, X., Fischbach, M. A., and Clardy, J. (2010) Dapdiamides, tripeptide antibiotics formed by unconventional amide ligases. *J. Nat. Prod.* **73**, 441–446.

DeBoer, C., Meulman, P. A., Wnuk, R. J., and Peterson, D. H. (1970) Geldanamycin, a new antibiotic. *J. Antibiot.* **23**, 442–447.

Delcher, A. L., Bratke, K. A., Powers, E. C., and Salzberg, S. L. (2007) Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* **23**, 673–679.

De Schrijver, A., and De Mot, R. (1999) A subfamily of MalT-related ATP-dependent regulators in the LuxR family. *Microbiology* **145**, 1287–1288.

Del Vecchio, F., Petkovic, H., Kendrew, S. G., Low, L., Wilkinson, B., Lill, R., Cortés, J., Rudd, B. A., Staunton, J., and Leadlay, P. F. (2003) Active-site residue, domain and module swaps in modular polyketide synthases. *J. Ind. Microbiol. Biotechnol.* 30, 489–94.

den Hengst, C. D., Tran, N. T., Bibb, M. J., Chandra, G., Leskiw, B. K., and Buttner, M. J. (2010) Genes essential for morphological development and antibiotic production in *Streptomyces coelicolor* are targets of BldD during vegetative growth. *Mol. Microbiol.* **78**, 361–379.

Dewick, P. M. (2002) The biosynthesis of C5-C25 terpenoid compounds. *Nat. Prod. Rep.* **19**, 181–222.

Dewick, P. M. (2009) Medicinal natural products: a biosynthetic approach, 3rd edn. (West Sussex, John Wiley and Sons. pp. 187–234.

D'Halluin, K., De Block, M., Denecke, J., Janssens, J., Leemans, J., Reynaerts, A., and Botterman, J. (1992) The *bar* gene as selectable and screenable marker in plant engineering. *Meth. Enzymol.* **216**, 415-426.

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

Dickschat, J. S. (2016) Bacterial terpene cyclases. *Nat. Prod. Rep.* **33**, 87-110.

Donadio, S., Staver, M. J., McAlpine, J. B., Swanson, S. J., and Katz, L. (1991) Modular organization of genes required for complex polyketide biosynthesis. *Science* **252**, 675–679.

Donadio, S., Monciardini, P., and Sosio, M. (2007) Polyketide synthases and nonribosomal peptide synthetases: the emerging view from bacterial genomics. *Nat. Prod. Rep.* **24**, 1073–1109.

Donahue, J.P., Israel, D.A., Peek, R.M., Blaser, M.J. & Miller, G.G. (2000) Overcoming the restriction barrier to plasmid transformation of *Helicobacter pylori*. *Molec. Microbiol.* **37**, 1066–1074.

Duong, C. T., Lee, H. N., Choi, S. S., Lee, S. Y. & Kim, E. S. (2009) Functional expression of SAV3818, a putative TetR-family transcriptional regulatory gene from *Streptomyces avermitilis*, stimulates antibiotic production in *Streptomyces* species. *J. Microbiol. Biotechnol*. **19**, 136–139.

Doroghazi, J. R., Albright, J. C., Goering, A. W., Ju, K. S., Haines, R. R., Tchalukov, K. A., Labeda, D. P., Kelleher, N. L., and Metcalf, W. W. (2014) A roadmap for natural product discovery based on large-scale genomics and metabolomics. *Nat. Chem. Biol.* **10**, 963–968.

Du, Y. L., Dalisay, D. S., Andersen, R. J., Ryan, K. S. (2013) N-carbamoylation of 2,4-diaminobutyrate reroutes the outcome in padanamide biosynthesis. *Chem. Biol.* **20**, 1002–1011.

Duncan, K. R., Crüsemann, M., Lechner, A., Sarkar, A., Li, J., Ziemert, N., Wang, M., Bandeira, N., Moore, B. S., Dorrestein, P. C., and Jensen, P. R. (2015) Molecular networking and pattern-based genome mining improves discovery of biosynthetic gene clusters and their products from *Salinispora* species. *Chem. Biol.* **22**, 460–471.

Eisenreich, W., Schwarz, M., Cartayrade, A., Arigoni, D., Zenk, M. H., and Bacher, A. (1998) The deoxyxylulose phosphate pathway of terpenoid biosynthesis in plants and microorganisms. *Chem. Biol.* **5**, R221–R233.

El-Sayed, A. K., Hothersall, J., Cooper, S. M., Stephens, E., Simpson, T. J., and Thomas, C. M. (2003) Characterization of the mupirocin biosynthesis gene cluster from *Pseudomonas fluorescens* NCIMB 10586. *Chem. Biol.* **10**, 419–430.

Embley, T. M., and Stackebrandt, E. (1994) The molecular phylogeny and systematics of the actinomycetes. *Annu. Rev. Microbiol.* **48**, 257–289.

Eschenmoser, A., Ruzicka, L., Jeger, O., and Arigoni, D. (1955) 226. On terpenes: a stereochemical interpretation of the biogenetic isoprene rule for terpenes. *Helv. Chim. Acta* **38**, 1890–1904.

Fair, R. J. and Tor, Y. (2014) Antibiotics and bacterial resistance in the 21st century. *Perspect. Med. Chem.* **6**, 25–64.

Feng, Z., Kim, J. H., and Brady, S. F. (2010) Fluostatins produced by the heterologous expression of a TAR reassembled environmental DNA derived type II PKS gene cluster. *J. Am. Chem. Soc.* **132**, 11902-11903.

Feng, Z., Kallifidas, D., and Brady, SF. (2011) Functional analysis of environmental DNA-derived type II polyketide synthases reveals structurally diverse secondary metabolites. *Proc. Natl. Acad. Sci. USA* **108**, 12629–12634

Fernández-Moreno, M. A., Martínez, E., Boto, L., Hopwood, D. A., and Malpartida, F. (1992) Nucleotide sequence and deduced functions of a set of cotranscribed genes of *Streptomyces coelicolor* A3(2) including the polyketide synthase for the antibiotic actinorhodin. *J. Biol. Chem.* **267**, 19278–19290.

Firn, R. D. and Jones, C. G. (2000) The evolution of secondary metabolism - a unifying model. *Mol. Microbiol.* **37**, 989–994.

Firn, R.D., and Jones, C.G. (2003) Natural products – a simple model to explain chemical diversity. *Nat. Prod. Rep.* **20**, 382–391.

Flett, F., Mersinias, V. & Smith, C.P. (1997) High efficiency intergeneric conjugal transfer of plasmid DNA from Escherichia coli to methyl DNA-restricting streptomycetes. FEMS Microbiology Letters, 155, 223–229.

Frey, P. A., Hegeman, A. D., and Ruzicka, F. J. (2008) The Radical SAM Superfamily. *Crit. Rev. Biochem. Mol. Biol.* **43**, 63–88.

Fu, J., Bian, X., Hu, S., Wang, H., Huang, F., Seibert, P. M., Plaza, A., Xia, L., Müller, R., Stewart, A. F., Zhang, Y. (2012) Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nat. Biotechnol.* **30**, 440-446.

Fu, C., Donovan, W. P., Shikapwashya-Hasser, O., Ye, X. and Cole, R. H. (2014). Hot Fusion: An Efficient Method to Clone Multiple DNA Fragments as Well as Inverted Repeats without Ligase. *PLoS One* **9**, e115318.

Fugate, C. J., and Jarrett, J. T. (2012) Biotin synthase: insights into radical-mediated carbon-sulfur bond formation. *Biochim. Biophys. Acta.* **1824**, 1213–1222.

Funa, N., Ohnishi, Y., Fujii, I., Shibuya, M., Ebizuka, Y., and Horinouchi, S. (1999) A new pathway for polyketide synthesis in microorganisms. *Nature* **400**, 897–899.

Funk, M. A., and van der Donk, W. A. (2017) Ribosomal natural products, tailored to fit. *Acc. Chem. Res.* **50**, 1577–1586.

Furumai, T., Yamakawa, T., Yoshida, R. & Igarashi, Y. (2003) Clethramycin, a new inhibitor of pollen tube growth with antifungal activity from *Streptomyces hygroscopicus* TP-A0623. I. Screening, taxonomy, fermentation, isolation and biological properties. *J. Antibiot.* **56**, 700–704.

Gallin, J. I. (2012) *A Historical Perspective on Clinical Research*. *In*: *Principles and Practice of Clinical Research* (3rd Edition) (Gallin, J. I., and Ognibene, F. eds.)

Galm, U., Wendt-Pienkowski, E., Wang, L., George, N. P., Oh, T. J., Yi, F., Tao, M., Coughlin, J. M., and Shen, B. (2009) The biosynthetic gene cluster of zorbamycin, a member of the bleomycin family of antitumor antibiotics, from *Streptomyces flavoviridis* ATCC 21892. *Mol. Biosyst*. **5**, 77–90.

Gardete, S., and Tomasz, A. (2014) Mechanisms of vancomycin resistance in *Staphylococcus aureus*. *J. Clin. Invest.* **124**, 2836–2840.

Gatewood, M. L., Bralley, P., Weil, M, R., and Jones, G. H. (2012) RNA-Seq and RNA immunoprecipitation analyses of the transcriptome of *Streptomyces coelicolor* identify substrates for RNase III. *J. Bacteriol.* 194, 2228–2237.

genomics. *Nat. Prod. Rep.* **24**, 1073–1109.

Gibson, D. G. (2011). Enzymatic assembly of overlapping DNA fragments. *Methods Enzymol.* **498**, 349–361.

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C.A. 3rd, and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345.

Ginsberg, J. (2005) *Selman Waksman and antibiotics*. American Chemical Society.

Goering A. W., McClure, R. A., Doroghazi, J. R., Albright, J. C., Haverland, N. A., Zhang, Y., Ju, K. S., Thomson, R. J., Metcalf, W. W., and Kelleher, N. L. (2016) Metabologenomics: Correlation of microbial gene clusters with metabolites drives discovery of a nonribosomal peptide with an unusual amino acid monomer.*ACS Cent. Sci.* **2**, 99–108.

Goh, E. B., Yim, G., Tsui, W., McClure, J., Surette, M. G. and Davies, J. (2002) Transcriptional modulation of bacterial gene expression by subinhibitory concentrations of antibiotics. *Proc. Natl. Acad. Sci. USA.* **99**, 17025–17030.

Gomez-Escribano, J.P. and Bibb, M.J. (2011). Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters. *Microb. Biotechno*l. **4**, 207–215.

Gomez-Escribano, J.P., Alt, S., and Bibb, M.J. (2016). Next Generation Sequencing of Actinobacteria for the Discovery of Novel Natural Products. Mar Drugs **14**, 19.

Goodfellow, M., Kumar, Y., Labeda, D. P., and Sembiring, L. (2007) The *Streptomyces violaceusniger* clade: a home for Streptomycetes with rugose ornamented spores. *Antonie Van Leeuwenhoek* **92**, 173–199.

Goodwin, S., Gurtowski, J., Ethe-Sayers, S., Deshande, P., Schatz, M. C., and McCombie, W. R. (2015) Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. *Genome Res.* **25**, 1–7.

Grabley, S., Hammann, P., Raether, W., Wink, J., and Zeeck, A. (1990) Secondary metabolites by chemical screening: II. Amycins A and B two novel niphimycin analogs isolated from a high producer strain of elaiophylin and nigericin. *J. Antibiot*. **43**, 639–647.
Graupner, K., Scherlach, K., Bretschneider, T., Lackner, G., Roth, M., Gross, H., and Hertweck, C. (2012) Imaging mass spectrometry and genome mining reveal highly antifungal virulence factor of mushroom soft rot pathogen. *Angew. Chem. Int. Ed.* **51**, 13173–13177.

Gregory, M.A., Till, R. and Smith, M.C.M. (2003). Integration site for *Streptomyces* phage phiBT1 and development of site-specific integrating vectors. *J. Bacteriol*. **185**, 5320–5323.

Gross, H., and Loper, J. E., (2009) Genomics of secondary metabolite production by *Pseudomonas* spp. *Nat. Prod. Rep*. **26**, 1408–1446.

Gutiérrez, S., Díez, B., Montenegro, E., and Martín, J. F. (1991) Characterization of the *Cephalosporium acremonium* pcbAB gene encoding alpha-aminoadipyl-cysteinyl-valine synthetase, a large multidomain peptide synthetase: linkage to the pcbC gene as a cluster of early cephalosporin biosynthetic genes and evidence of multiple functional domains. *J. Bacteriol.***173**, 2354–2365.

Hahn, D. R., Gustafson, G., Waldron, C., Bullard, B., Jackson, J. D., and Mitchell, J. (2006) Butenyl-spinosyns, a natural example of genetic engineering of antibiotic biosynthetic genes. *J. Ind. Microbiol. Biotechnol*. **3**3, 94–104.

Hahn, D. R., Graupner, P. R., Chapin, E., Gray, J., Heim, D., Gilbert, J. R., Gerwick, B. C. (2009) Albucidin: a novel bleaching herbicide from *Streptomyces albus* subsp. *chlorinus* NRRL B-24108. *J. Antibiot.* **62**, 191–194.

Harned, R. L., Hidy, P. H., Corum, C. J., and Jones, K. L. (1951) Nigericin a new crystalline antibiotic from an unidentified *Streptomyces*. *Antibiot. Chemother.* **1**, 594-596.

Harvey, B. M., Mironenko, T., Sun, Y., Hong, H., Deng, Z., Leadlay, P. F., Weissman, K. J., and Haydock, S. F. (2007) Insights into polyether biosynthesis from analysis of the nigericin biosynthetic gene cluster in *Streptomyces* sp. DSM4137. *Chem. Biol.* **14**, 703–714.

Haslam, E. (1994) Secondary metabolism — evolution and function: products or processes? *Chemoecology* **5**, 89–95.

Haydock, S. F., Aparicio, J. F., Molnár, I., Schwecke, T., Khaw, L.E., König, A., Marsden, A. F. A., Galloway, I. S., Staunton, J., and Leadlay, P. F. (1995). Divergent sequence motifs

correlated with the substrate specificity of (methyl)malonyl-CoA: ACP transacylase domains in modular polyketide synthases. *FEBS Lett*. **374**, 246–248.

Haydock, S. F., Appleyard, A. N., Mironenko, T., Lester, J., Scott, N., and Leadlay., (2005) Organization of the biosynthetic gene cluster for the macrolide concanamycin A in *Streptomyces neyagawaensis* ATCC 27449. *Microbiology* **151**, 3161–3169.

Haydock, S. F., Mironenko, T., Ghoorahoo, H. I., and Leadlay, P. F. (2004) The putative elaiophylin biosynthetic gene cluster in *Streptomyces sp*. DSM4137 is adjacent to genes encoding adenosylcobalamin-dependent methylmalonyl CoA mutase and to genes for synthesis of cobalamin. *J. Biotechnol.* **113**, 55–68.

Heather, J. M. and Chain, B. (2016) The sequence of sequencers: the history of sequencing DNA. *Genomics* **107**, 1–8.

Helfrich, E. J., and Piel, J. (2016) Biosynthesis of polyketides by *trans*-AT polyketide synthases. *Nat. Prod. Rep.* 33, 231–316.

Herrmann, J., Fayad, A. A., and Müller R. (2017) Natural products from myxobacteria: novel metabolites and bioactivities. *Nat. Prod. Rep.* **34**, 135–160.

Hertweck, C. (2009) The biosynthetic logic of polyketide diversity. *Angew. Chem. Int. Ed.* **48**, 4688–4716.

Hesketh, A., Chen, W. J, Ryding, J., Chang, S., and Bibb, M. (2007) The global role of ppGpp synthesis in morphological differentiation and antibiotic production in *Streptomyces coelicolor* A3(2). *Genome Biol.* **8**, R161.

Hien, A. S., Soma, D. D., Hema, O., Bayili, B., Namountougou, M., Gnankiné, O., Baldet, T., Diabaté, A., and Dabiré K. R. (2017) Evidence that agricultural use of pesticides selects pyrethroid resistance within *Anopheles gambiae* s.l. populations from cotton growing areas in Burkina Faso, West Africa. *PLoS One* **12**, e0173098.

Hong, L., Zhao, Z., Melançon, C. E. 3rd, Zhang, H., Liu H.-w. (2008) In vitro characterization of the enzymes involved in TDP-D-forosamine biosynthesis in the spinosyn pathway of *Saccharopolyspora spinosa*. *J. Am. Chem. Soc.* **130**, 4954–4967Hong, H. J., and Hesketh, A. (2016) Microarray analysis to monitor bacterial cell wall homeostasis. *Methods Mol. Biol.* **1440**, 31-46.

Hong, H., Sun, Y., Zhou, Y., Stephens, E., Samborskyy, M., and Leadlay, P. F. (2016) Evidence for an iterative module in chain elongation on the azalomycin polyketide synthase. *Beilstein J. Org. Chem.* **12**, 2164–2172.

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

Hopwood, D.A., Kieser, T., Wright, H.M. and Bibb, M.J. (1983). Plasmids, recombination and chromosome mapping in *Streptomyces lividans* 66. *J. Gen. Microbiol*. **129**, 2257–2269.

Hou, W., Liu, Q., Tian, L., Wu, Q., Zhang, Y., Xie, W., Wang, S., Miguel, K. S., Funderburk, J., and Scott J. G. (2014) The α6 nicotinic acetylcholine receptor subunit of *Frankliniella occidentalis* is not involved in resistance to spinosad. *Pestic. Biochem. Physiol.* **111**, 60-67.

Huang, J., Lih, C. J., Pan, K. H., and Cohen, S. N. (2001) Global analysis of growth phase responsive gene expression and regulation of antibiotic biosynthetic pathways in *Streptomyces coelicolor* using DNA microarrays. *Genes Dev.* **15**, 3183–3192.

Huang, J., Shi, J., Molle, V., Sohlberg, B., Weaver, D., Bibb, M. J., Karoonuthaisiri, N., Lih, C. J., Kao, C. M., Buttner, M. J., and Cohen, S. N. (2005) Cross-regulation among disparate antibiotic biosynthetic pathways of *Streptomyces coelicol*or. *Mol. Microbiol.* **58**, 1276–1287.

Huang, J., Yu, Z., Li, M.-H., Wang, J.-D., Bai, H., Zhou, J. & Zheng, Y.-G. (2016). High level of spinosad production in the heterologous host *Saccharopolyspora erythraea*. *Appl. Environ. Microbiol.,* **82**, 5603–5611.

Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M. and Omura, S. (2003) Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat. Biotechnol.* **21**, 526–3531.

Igarashi, Y., Iwashita, T., Fujita, T., Naoki, H., Yamakawa, T., Yoshida, R., Furumai, T. (2003) Clethramycin, a new inhibitor of pollen tube growth with antifungal activity from *Streptomyces hygroscopicus* TP-A0623. II. Physico-chemical properties and structure determination. *J. Antibiot.* **56**, 705–708.

Igarashi, Y., Tanaka, Y,. Ikeda, M,. Oikawa, T,. Kitani, S,. Nihira, T,. Mongkol, P,. Janhom, M,. and Panbangred, W,. (2012). Prajinamide, a new modified peptide from a soil-derived Streptomyces. *J. antibiot* **65**,157–159.

Imker, H. J., Walsh, C. T., and Wuest, W. M. (2009) SylC catalyzes ureido-bond formation during biosynthesis of the proteasome inhibitor syringolin A. *J. Am. Chem. Soc.* **131**, 18263–18265.

Jayapal, K. P., Lian, W., Glod, F., Sherman, D. H., and Hu, W. S. (2007) Comparative genomic hybridizations reveal absence of large *Streptomyces coelicolor* genomic islands in *Streptomyces lividans*. *BMC Genomics* **8**, 229.

Jensen, K., Niederkrüger, H., Zimmermann, K., Vagstad, A. L., Moldenhauer, J., Brendel. N., Frank, S., Pöplau, P., Kohlhaas, C., Townsend, C. A., Oldiges, M., Hertweck, C., and Piel, J. (2012) Polyketide proofreading by an acyltransferase-like enzyme. *Chem. Biol.* **19**, 329–339.

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

Ji, H.-F., Li, X.-J., and Zhang, H.-Y. (2009) Natural products and drug discovery. Can thousands of years of ancient medical knowledge lead us to new and powerful drug combinations in the fight against cancer and dementia? *EMBO Rep.* 10, 194–200.

Jia, X. Y., Tian, Z., Shao, L., Qu, X., Zhao, Q., Tang, J., Tang, G., Liu, W. (2006) Cloning, sequencing and genetic characterization of the chlorothricin gene cluster from *Streptomyces antibioticus* DSM 40725 as a model for spirotetronate antibiotic biosynthesis. *Chem. Biol.* **13**, 575–585.

Jiang, M., Yin, M., Wu, S., Han, X., Ji, K., Wen, M., Lu, T. (2017) GdmRIII, a TetR family transcriptional regulator, controls geldanamycin and elaiophylin biosynthesis in *Streptomyces autolyticus* CGMCC0516. *Sci. Rep.* **7**, 4803.

Jiang, W., Zhao, X., Gabrieli, T., Lou, C., Ebenstein, Y. and Zhu, T.F. (2015). Cas9-Assisted targeting of CHromosome segments CATCH enables one-step targeted cloning of large gene clusters. *Nat. Commun*. **6**, ncomms9101.

Jiang, Y., Wang, H., Lu, C., Ding, Y., Li, Y., and Shen, Y. (2013) Identification and characterization of the cuevaene A biosynthetic gene cluster in streptomyces sp. LZ35. *ChemBiochem.* **14**, 1468–1475.

Jin M., Fischbach, M. A., and Clardy, J. (2006) A biosynthetic gene cluster for the acetyl-CoA carboxylase inhibitor andrimid. *J. Am. Chem. Soc.* **128**, 10660–10661.

Jones, A.C., Gust, B., Kulik, A., Heide, L., Buttner, M.J. & Bibb, M.J. (2013). Phage P1-Derived Artificial Chromosomes Facilitate Heterologous Expression of the FK506 Gene Cluster. *PLoS One* **8**, e69319.

Ju, K. S., Gao, J., Doroghazi, J. R., Wang, K. K., Thibodeaux, C. J., Li, S., Metzger, E., Fudala, J., Su, J., Zhang, J. K., Lee, J., Cioni, J. P., Evans, B. S., Hirota, R., Labeda, D. P., van der Donk, W. A., and Metcalf, W. W. (2015) Discovery of phosphonic acid natural products by mining the genomes of 10,000 actinomycetes.

Proc. Natl. Acad. Sci. USA **112**, 12175–12180.

Kanchanabanca, C., Tao, W., Hong, H., Liu, Y., Hahn, F., Samborskyy, M., Deng, Z., Sun, Y., and Leadlay, P. F. (2013) Unusual acetylation-elimination in the formation of tetronate antibiotics. *Angew. Chem. Int. Ed.* **52**, 5785–5788.

Katsuyama, Y., Kita, T., Funa, N., and Horinouchi, S. (2009) Curcuminoid biosynthesis by two type III polyketide synthases in the herb *Curcuma longa*. *J. Biol. Chem*. **284**, 11160–11170.

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

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. and Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649.

*Keatinge-Clay, A .T. (2007) A tylosin ketoreductase reveals how chirality is determined in polyketides. Chem. Biol. **14**, 898–908.*

Kelly W. L., Pan, L., and Li, C. (2009) Thiostrepton biosynthesis: prototype for a new family of bacteriocins. *J. Am. Chem. Soc.* **131**, 4327–4334.

Kevany, B. M., Rasko, D. A., and Thomas, M. G. (2008) Characterization of the complete zwittermicin A biosynthesis gene cluster from *Bacillus cereus*. *Appl. Environ. Microbiol.* **75**, 1144–1155.

Kumbhar, C., Mudliar, P., Bhatia, L., Kshirsagar, A., and Watve M. (2014) Widespread predatory abilities in the genus *Streptomyces*. *Arch. Microbiol*.**196**, 235-248.

Kim S. Y., Zhao, P., Igarashi, M., Sawa, R., Tomita, T., Nishiyama, M., and Kuzuyama, T. (2009) Cloning and heterologous expression of the cyclooctatin biosynthetic gene cluster afford a diterpene cyclase and two P450 hydroxylases. *Chem. Biol.* **16**, 736–743.

Kim, H. J., Ruszczycky, M. W., Choi, S. H., Liu, Y. N., and Liu, H.-w. (2011) Enzyme-catalysed [4+2] cycloaddition is a key step in the biosynthesis of spinosyn A. *Nature* **473**, 109–112.

Kim, H. J., Karki, S., Kwon, S. Y., Park, S. H., Nahm, B. H., Kim, Y. K., and Kwon, H. J. (2014) A single module type I polyketide synthase directs de novo macrolactone biogenesis during galbonolide biosynthesis in Streptomyces galbus. J. Biol. Chem. 289, 34557–34568.

Kim, J. H., Feng, Z., Bauer, J. D., Kallifidas, D., Calle, P.Y., Brady, S. F. (2010) Cloning large natural product gene clusters from the environment: piecing environmental DNA gene clusters back together with TAR. *Biopolymers* **93**, 833–844.

Kitagawa, M., Hasegawa, S., Saito, S., Shimada, N., and Takita, T. (1991) Synthesis and antiviral activity of oxetanocin derivatives. *Tetrahedron Lett*. **32**, 3531-3534.

Kittilä, T., Mollo, A., Charkoudian, L. K., and Cryle, M. J. (2016) New structural data reveal the motion of carrier proteins in nonribosomal peptide synthesis. *Angew. Chem. Int. Ed.* **55**, 9834–9840.

Kleinkauf, H., and Koischwitz, H. (1980) Gramicidin S-synthetase: on the structure of a polyenzyme template in polypeptide biosynthesis. *Mol. Biol. Biochem. Biophys.* **32**, 205–216.

Kleinkauf, H., and von Döhren, H. (1987) Biosynthesis of peptide antibiotics. *Ann. Rev. Microbiol.* **41**, 259–289.

Knappe, T. A., Linne, U., Zirah, S., Rebuffat, S., Xie, X., and Marahiel, M A. (2008) Isolation and structural characterization of capistruin, a lasso peptide predicted from the genome sequence of *Burkholderia thailandensis* E264. *J. Am. Chem. Soc.* **130**, 11446–11454.

Komaki, H., Hosoyama, A., Ichikawa, N., Panbangred, W., Igarashi, Y,. (2016) Draft genome sequence of Streptomyces sp. SPMA113, a prajinamide producer. Genome Announc **4**, 01126-16.

Komatsu, M., Tsuda, M., Omura, S., Oikawa, H., and Ikeda, H. (2008) Identification and functional analysis of genes controlling biosynthesis of 2-methylisoborneol.

Proc. Natl. Acad. Sci. USA **105**, 7422-7427.

Komatsu, M., Uchiyama, T., Omura, S., Cane, D. E., Ikeda, H. (2010) Genome-minimized *Streptomyces* host for the heterologous expression of secondary metabolism. *Proc. Natl. Acad. Sci. USA* **107**, 2646–2651.

Komatsu, M., Komatsu, K., Koiwai, H., Yamada, Y., Kozone, I., Izumikawa, M., Hashimoto, J., Takagi, M., Omura, S., Shin-ya, K., Cane, D. E., and Ikeda, H. (2013) Engineered *Streptomyces avermitilis* host for heterologous expression of biosynthetic gene cluster for secondary metabolites. *ACS Synth. Biol.* **2**, 384–396.

Kouprina, N., Larionov, V. (2016) Transformation-associated recombination (TAR) cloning for genomics studies and synthetic biology. *Chromosoma* **125**, 621–632.

Krug, D., Zurek, G., Revermann, O., Vos, M., Velicer, G. J., and Müller, R. (2008) Discovering the hidden secondary metabolome of *Myxococcus xant*hus: a study of intraspecific diversity. *Appl. Environ. Microbiol.* **74**, 3058-3068.

Krug, D., and Müller, R. (2014) Secondary metabolomics: the impact of mass spectrometry-based approaches on the discovery and characterization of microbial natural products. *Nat. Prod. Rep.* **31**, 768–783.

Kumagai, T., Koyama, Y., Oda, K., Noda, M., Matoba, Y. & Sugiyama, M. (2010). Molecular cloning and heterologous expression of a biosynthetic gene cluster for the antitubercular agent D-cycloserine produced by *Streptomyces lavendulae*. *Antimicrob. Agents Chemother*. **54**, 1132–1139.

Kumbhar, C., Mudliar, P., Bhatia, L., Kshirsagar, A., and Watve, M. (2014) Widespread predatory abilities in the genus *Streptomyces*. *Arch. Microbiol.* **196**, 235-248.

Kumar, Y., Aiemsum-Ang, P., Ward, A. C., and Goodfellow, M. (2007) Diversity and geographical distribution of members of the Strep*tomyces violaceusniger* 16S rRNA gene clade detected by clade-specific PCR primers. *FEMS Microbiol. Ecol.* 62, 54-63.

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., Salzberg, S. L. (2004) Versatile and open software for comparing large genomes. *Genome Biol.* **5**, R12.

Kusari, S., Hertweck, C., and Spiteller, M. (2012a) Chemical ecology of endophytic fungi: origins of secondary metabolites. *Chem. Biol.* **19**, 792–798.

Kusari, S., Verma, V.C., Lamshoft, M., and Spiteller, M. (2012b). An endophytic fungus from *Azadirachta indica* A. Juss. that produces azadirachtin. *World J. Microbiol. Biotechnol.* **28**, 1287–1294.

*Kwan, D.H., Sun, Y., Schulz, F., Hong, H., Popovic, B., Sim-Stark, J.C., Haydock, S.F., Leadlay, P.F. (2008) Prediction and stereochemistry of enoylreduction in modular polyketide synthases. Chem. Biol. **15**, 1231-1240.*

Kwok, R. (2010) Five hard truths for synthetic biology. *Nature* **463**, 288–290.

Lamberth, C. (2016) Naturally occurring amino acid derivatives with herbicidal, fungicidal or insecticidal activity. *Amino Acids* **48**, 929–940.

Landgraf, B. J., McCarthy, E. L., and Booker, S. J. (2016) Radical S-Adenosylmethionine Enzymes in Human Health and Disease. *Annu. Rev. Biochem*. **85**, 485–514.

Larionov, V., Kouprina, N., Graves, J., Chen, X. N., Korenberg, J. R., and Resnick,

M. A. (1996) Specific cloning of human DNA as yeast artificial chromosomes by transformation-associated recombination. *Proc. Natl. Acad. Sci. USA* **93**, 491–496

Ramsay, M. (1994) Yeast artificial chromosome cloning. *Mol. Biotechnol.* **1**, 181–201.

Laskaris P., Tolba, S, Calvo-Bado, L., and Wellington, E. M. (2010) Coevolution of antibiotic production and counter-resistance in soil bacteria. *Environ. Microbiol.* **12**, 783-796.

Laureti, L., Song, L., Huang, S., Corre, C., Leblond, P., Challis, G. L., and Aigle, B. (2011) Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in *Streptomyces ambofaciens*. *Proc. Natl. Acad. Sci. USA* **108**, 6258–6263.

Lautru, S., Deeth, R. J., Bailey, L. M., and Challis, G. L. (2005) Discovery of a new peptide natural product by *Streptomyces coelicolor* genome mining. *Nat. Chem. Biol*. **1**, 265–269.

Layer, G., Pierik, A. J., Trost, M., Rigby, S. E., Leech, H. K., Grage, K., Breckau, D. Astner, I. Jansch, L. Heathcote, P., Warren, M. J., Heinz, D. W., and Jahn, D. (2006) The substrate radical of *Escherichia coli* oxygen-independent coproporphyrinogen III oxidase HemN. *J. Biol. Chem*. **281**, 15727–15734.

Lechartier, B., Rybniker, J., Zumla, A., and Cole, S. T. (2014) Tuberculosis drug discovery in the post-post-genomic era. *EMBO Mol. Med.* **6**,158–68.

Letzel, A. C., Li, J., Amos, G. C. A., Millán-Aguiñaga, N., Ginigini, J., Abdelmohsen, U. R., Gaudêncio, S. P., Ziemert, N., Moore, B. S., and Jensen, P. R. (2017) Genomic insights into specialized metabolism in the marine actinomycete *Salinispora*. *Environ. Microbiol*. **19**, 3660–3673.

Lewer, P., Hahn, D. R., Karr, L. L., Duebelbeis, D. O., Gilbert, J. R., Crouse, G. D., Worden, T., Sparks, T. C., Edwards, P. M., and Graupner, P. R. (2009) Discovery of the butenyl-spinosyn insecticides: novel macrolides from the new bacterial strain *Saccharopolyspora pogona*. *Bioorg. Med. Chem*. **17**, 4185–4196.

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

Li, S., Li, Y., Lu, C., Zhang, J., Zhu, J., Wang, H., Shen, Y. (2015) Activating a cryptic ansamycin biosynthetic gene cluster to produce three new naphthalenic octaketide ansamycins with n-pentyl and n-butyl side chains. *Org. Lett*. **17**, 3706-3709.

Li, T.-L., Choroba, O. W., Hong, H., Williams, D. H., and Spencer, J. B. (2001) Biosynthesis of the vancomycin group of antibiotics: characterisation of a type III polyketide synthase in the pathway to (S)-3,5-dihydroxyphenylglycine. *Chem. Commun.,* 2156–2157.

Li, W., Csukai, M., Corran, A., Crowley, P., Solomon, P. S., and Oliver, R. P. (2008)

Malayamycin, a new streptomycete antifungal compound, specifically inhibits sporulation of *Stagonospora nodo*rum (Berk) Castell and Germano, the cause of wheat glume blotch disease. *Pest Manag. Sci.* **64**, 1294–1302.

Ling L. L., and 20 other authors. (2015) A new antibiotic kills pathogens without detectable resistance. *Nature* **517**, 455–459.

Liu, C., Zhang, J., Lu, C., and Shen, Y. (2015a) Heterologous expression of galbonolide biosynthetic genes in *Streptomyces coelicolor*. *Antonie Van Leeuwenhoek* **107**, 1359–1366.

Liu, C., Zhu, J., Li, Y., Zhang, J., Lu, C., Wang, H., and Shen, Y. (2015b) In vitro reconstitution of a PKS pathway for the biosynthesis of galbonolides in *Streptomyces* sp. LZ35. *ChemBiochem*. **16**, 998–1007.

Liu, L., Ji, X., Li, Y., Ji, W., Mo, T., Ding, W., and Zhang, Q. (2017) A mechanistic study of the non-oxidative decarboxylation catalyzed by the radical S-adenosyl-L-methionine enzyme BlsE involved in blasticidin S biosynthesis. *Chem. Commun*. **53**, 8952–8955.

Liu, W., Nonaka, K., Nie, L., Zhang, J., Christenson, S. D., Bae, J., Van Lanen, S. D., Zazopoulos, E., Farnet, C. M., Yang, C. F., and Shen, B. (2005) The neocarzinostatin biosynthesis gene cluster from *Streptomyces carzinostaticus* ATCC15944 involving two iterative type I polyketide synthases. *Chem. Biol*. **12**, 293–302.

Liu, Y., Tao, W., Wen, S., Li, Z., Yang, A., Deng, Z., and Sun, Y. (2015). In vitro CRISPR/cas9 system for efficient targeted DNA editing. *mBio* **6**, e01714-01715.

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

Lohman, J. R., Ma, M., Osipiuk, J., Nocek, B., Kim, Y., Chang, C., Cuff, M., Mack, J., Bigelow, L., Li, H., Endres, M., Babnigg, G., Joachimiak, A., Phillips, G. N. Jr., and

Louis, P., and Galinski, E. A. (1997). Characterization of genes for the biosynthesis of the compatible solute ectoine from *Marinococcus halophilus* and osmoregulated expression in *Escherichia coli*. *Microbiology* **143**, 1141–1149.

Macheleidt, J., Mattern, D. J., Fischer, J., Netzker, T., Weber, J., Schroeckh, V., Valiante V., and Brakhage A. A. (2016) Regulation and role of fungal secondary metabolites. *Annu. Rev. Genet.* **50**, 371–392.

MacNeil, D.J., Gewain, K.M., Ruby, C.L., Dezeny, G., Gibbons, P.H. and MacNeil, T. (1992). Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene* **111**, 61–68.

Mahenthiralingam, E., Song, L., Sass, A., White, J., Wilmot, C., Marchbank, A., Boaisha, O., Paine, J., Knight, D., and Challis, G. L. (2011) Enacyloxins are products of an unusual hybrid modular polyketide synthase encoded by a cryptic *Burkholderia ambifaria* genomic island. *Chem. Biol.* **18**, 665–677.

Maksimov, M. O., and Link, A. J. (2014) Prospecting genomes for lasso peptides. *J. Ind. Microbiol. Biotechnol.* **41**, 333–344.

Malpartida, F., and Hopwood, D. A. (1986) Physical and genetic characterisation of the gene cluster for the antibiotic actinorhodin in *Streptomyces coelicolor* A3(2). *Mol. Gen. Genet.* 205, 66–73.

Mandala, S. M., Thornton, R. A., Milligan, J., Rosenbach, M., Garcia-Calvo, M., Bull, H. G., Harris, G., Abruzzo, G. K., Flattery, A. M., Gill, C. J., Bartizal, K., Dreikorn, S., and Kurtz, M. B. (1998) Rustmicin, a potent antifungal agent, inhibits sphingolipid synthesis at inositol phosphoceramide synthase. *J. Biol. Chem*. **273**, 14942–14949.

Mantione K. J., Kream, R. M., Kuzelova, H., Ptacek, R., Raboch, J., Samuel, J. M., and Stefano, G. B. (2014) Comparing bioinformatic gene expression profiling methods: microarray and RNA-Seq. *Med. Sci. Monit. Basic Res.* **20**, 138–142.

Marahiel, M.A. (2016) A structural model for multimodular NRPS assembly lines *Nat. Prod. Rep.* **33**, 136–140.

Margulies, M. and 30 other authors (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376-380.

Martin, C. J., Timoney, M. C., Sheridan, R. M., Kendrew, S. G., Wilkinson, B., Staunton, J., and Leadlay, P. F. (2003) Heterologous expression in *Saccharopolyspora erythraea* of a pentaketide synthase derived from the spinosyn polyketide synthase. *Org. Biomol. Chem.* **1**, 4144–4147.

Martín, J. F., Casquiero, J., and Liras, P. (2005) Secretion systems for secondary metabolites: how producer cells send out messages of intercellular communication. *Curr. Opin. Microbiol.* **8**, 282–293.

Martín, J., and Liras, P. (2012) Cascades and networks of regulatory genes that control antibiotic biosynthesis. *Subcell. Biochem.* **64**, 115–138.

Matsushima, P. and Baltz, R.H. (1994). Transformation of *Saccharopolyspora spinosa* protoplasts with plasmid DNA modified in vitro to avoid host restriction. *Microbiology* **140**, 139–143.

McClure, R. A., Goering, A. W., Ju, K. S., Baccile, J. A., Schroeder, F. C., Metcalf, W. W., Thomson, R. J., and Kelleher, N. L. (2016) Elucidating the rimosamide-detoxin natural product families and their biosynthesis using metabolite/gene cluster correlations. *ACS Chem. Biol*. **11**, 3452–3460.

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

antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences.

Medema, M. H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M. A., Weber, T., Takano, E., Breitling, R. (2011) *Nucl. Acids Res.* **39** (Web Server issue): W339-346.

Medema, M. H., Trefzer, A., Kovalchuk, A., van den Berg, M., Müller, U., Heijne, W., Wu, L., Alam, M. T., Ronning, C. M., Nierman, W. C., Bovenberg, R. A. L., Breitling, R., and Takano, E. (2010) The sequence of a 1.8-Mb bacterial linear plasmid reveals a rich evolutionary reservoir of secondary metabolic pathways. *Genome Biol. Evol.* **2**, 212–224.

Meier, J. L., Niessen, S., Hoover, H. S., Foley, T. L., Cravatt, B. F., and Burkart, M. D. (2009) An orthogonal active site identification system (OASIS) for proteomic profiling of natural product biosynthesis. *ACS Chem. Biol.* **4**, 948–957.

Menche, D., Arikan, F., Perlova, O., Horstmann, N., Ahlbrecht, W., Wenzel, S. C., Jansen, R., Irschik, H., and Müller, R. (2008) Stereochemical determination and complex biosynthetic assembly of etnangien, a highly potent RNA polymerase inhibitor from the myxobacterium *Sorangium cellulosum*. *J. Am. Chem. Soc.* **130**, 14234–14243.

Mendéz, C, and Salas, J. A. (2001) The role of ABC transporters in antibiotic-producing organisms: drug secretion and resistance mechanisms. *Res. Microbiol.* **152**, 341–350.

Millar, N. S., and Denholm, I. (2007) Nicotinic acetylcholine receptors: targets for commercially important insecticides. *Invert. Neurosci.* **7**, 53–66.

Moldenhauer, J., Chen, X. H., Borriss, R., and Piel, J. (2007) Biosynthesis of the antibiotic bacillaene, the product of a giant polyketide synthase complex of the *trans*-AT family. *Angew. Chem. Int. Ed.* **46**, 8195–8197.

Mootz, H. D., and Marahiel, M. A. (1997) The tyrocidine biosynthesis operon of *Bacillus brevis*: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains. *J. Bacteriol.* **179**, 6843-6850.

Morita, M., Tomita, K., Ishizawa, M., Takagi, K., Kawamura, F., Takahashi, H., and Morino, T. (1999) Cloning of oxetanocin A biosynthetic and resistance genes that reside on a plasmid of *Bacillus megaterium* strain NK84-0128. *Biosci. Biotechnol. Biochem.* **63**, 563–566.

Motoyama T., and Osada, H. (2016) Biosynthetic approaches to creating bioactive fungal metabolites: Pathway engineering and activation of secondary metabolism. Bioorg. Med. Chem. Lett. 26, 5843–5850.

Motoyama T., and Osada, H. (2016) Biosynthetic approaches to creating bioactive fungal metabolites: Pathway engineering and activation of secondary metabolism. *Bioorg. Med. Chem. Lett.* **26**, 5843–5850.

Muliandi, A., Katsuyama, Y., Sone, K., Izumikawa, M., Moriya, T., Hashimoto, J.,, Kozone, I., Takagi, M., Shin-ya, K., and Ohnishi, Y. (2014) Biosynthesis of the 4-methyloxazoline-containing nonribosomal peptides, JBIR-34 and -35, in *Streptomyces* sp. Sp080513GE-23. *Chem. Biol.* **21**, 923–934.

Müller, W. M., Schmiederer, T., Ensle, P., and Süssmuth, R. D. (2010) In vitro biosynthesis of the prepeptide of type-III lantibiotic labyrinthopeptin A2 including formation of a C-C bond as a post-translational modification. *Angew. Chem. Int. Ed.* **49**, 2436–2440.

Musiol, E. M., Greule, A., Härtner, T., Kulik, A., Wohlleben, W., and Weber, T. (2013) The AT_2 domain of KirCI loads malonyl extender units to the ACPs of the kirromycin PKS. *ChemBiochem.* **14**, 1343–1352.

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

Nguyen, T., Ishida, K., Jenke-Kodama, H., Dittmann, E., Gurgui, C., Hochmuth, T., Taudien, S., Platzer, M., Hertweck, C., and Piel, J. (2009) Exploiting the mosaic structure of *trans*-acyltransferase polyketide synthases for natural product discovery and pathway dissection. *Nat. Biotechnol*. **26**, 225–233.

Nielsen, J, and Keasling, J. D. (2016) Engineering Cellular Metabolism. *Cell* **164**, 1185**–** 1197.

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

Niu, G., and Tan, H. (2015) Nucleoside antibiotics: biosynthesis, regulation, and biotechnology. *Trends Microbiol.* **23**, 110-119.

Niu, G., Zheng, J., and Tan, H. (2017) Biosynthesis and combinatorial biosynthesis of antifungal nucleoside antibiotics. Sci. China. Life Sci. 60, 939-947.

Noh, J. H., Kim, S. H., Lee, H.N., Lee, S. Y., and Kim, E. S. (2010) Isolation and genetic manipulation of the antibiotic down-regulatory gene, wblA ortholog for doxorubicin-producing strain improvement. *Appl. Microbiol. Biotechnol.* **86**, 1145–1153.

Olano, C., Wilkinson, B., Sánchez, C., Moss, S. J., Sheridan, R., Math, V., Weston, A. J., Braña, A. F., Martin, C. J., Oliynyk, M., Méndez, C., Leadlay, P. F., and Salas, J. A. (2004) Biosynthesis of the angiogenesis inhibitor borrelidin by *Streptomyces parvulus* Tü4055: cluster analysis and assignment of functions. *Chem. Biol.* **11**, 87–97.

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. *J. Bacteriol*. **190**, 4050-4060.

Oldfield, E., and Lin, F. Y. (2012) Terpene biosynthesis: modularity rules. *Angew. Chem. Int. Ed.* **51**, 1124-1137.

Oliynyk, M., Samborskyy, M., Lester, J. B., Mironenko, T., Scott, N., Dickens, S., Haydock, S. F., Leadlay, P.F. (2007) Complete genome sequence of the erythromycin-producing bacterium *Saccharopolyspora erythraea* NRRL23338. *Nat. Biotechnol.* **25**, 447–453.

Omrane, S., Sghyer, H., Audéon, C., Lanen, C., Duplaix, C., Walker, A. S., and Fillinger, S. (2015) Fungicide efflux and the MgMFS1 transporter contribute to the multidrug resistance phenotype in *Zymoseptoria tritici* field isolates. *Environ. Microbiol.* **17**, 2805-2823.

O'Rourke, S., Widdick, D., and Bibb, M. (2017) A novel mechanism of immunity controls the onset of cinnamycin biosynthesis in *Streptomyces cinnamoneus* DSM 40646. *J. Ind. Microbiol. Biotechnol.* **44**, 563-572.

Ortega, M. A., and van der Donk, W. A. (2016) New insights into the biosynthetic logic of ribosomally synthesized and post-translationally modified peptide natural products. *Cell Chem. Biol.* **23**, 31-44.

Osbourn, A. (2010) Secondary metabolic gene clusters: evolutionary toolkits for chemical innovation. *Trends Genet.* **26**, 449–457.

Owen, J. G., Charlop-Powers, Z., Smith, A. G., Ternei, M. A., Calle, P. Y., Reddy, B. V., Montiel, D., and Brady, S. F. (2015) Multiplexed metagenome mining using short DNA sequence tags facilitates targeted discovery of epoxyketone proteasome inhibitors. *Proc. Natl. Acad. Sci. USA* **112**, 4221–4226.

Owen, M. D., and Zelaya, I. A. (2005) Herbicide-resistant crops and weed resistance to herbicides. *Pest Manag. Sci.* **61**, 301-311.

Pachlatko, J. P. (1998) Natural products in crop protection. *Chimia* **52**, 29–47.

Paget, M.S.B., Chamberlin, L., Atrih, A., Foster, S.J., and Buttner, M.J. (1999). Evidence that the extracytoplasmic function sigma factor E Is required for normal cell wall structure in *Streptomyces coelicolor* A3(2). *J. Bacteriol.* **181**, 204–211.

Painter, R. E., Adam, G. C., Arocho, M., DiNunzio, E., Donald, R. G., Dorso, K., Genilloud, O., Gill, C., Goetz, M., Hairston, N. N., Murgolo, N., Nare, B., Olsen, D. B., Powles, M., Racine, F., Su. J., Vicente. F., Wisniewski, D, Xiao, L., Hammond, M., and Young, K. (2015) Elucidation of DnaE as the antibacterial target of the natural product, nargenicin. *Chem. Biol.* **22**, 1362–1373.

Pan, H.-X., Li, J.-A., He, N.-J., Chen, J.-Y., Zhou, Y.-M., Shao, L., and Chen, D.-J. (2011). Improvement of spinosad production by overexpression of gtt and gdh controlled by promoter PermE* in *Saccharopolyspora spinosa* SIPI-A2090. *Biotechnol. Lett.* **33**, 733–739.

Partida-Martinez, L. P., and Hertweck, C. (2005) Pathogenic fungus harbours endosymbiotic bacteria for toxin production. *Nature* **437**, 884–888.

Peano, C., Bicciato, S., Corti, G., Ferrari, F., Rizzi, E., Bonnal, R. J., Bordon, R., Albertini, A., Bernardi, L. R., Donadio, S., and De Bellis, G. (2007) Complete gene expression profiling of *Saccharopolyspora erythraea* using GeneChip DNA microarrays. *Microb. Cell Fact.* **6**, 37.

Piel, J. (2002) A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of *Paederus* beetles. *Proc. Natl. Acad. Sci. USA* **99**, 14002–14007.

Piel, J. (2011) Approaches to capturing and designing biologically active small molecules produced by uncultured microbes. *Annu. Rev. Microbiol*. **65**, 431–453.

Pierce, J.C., Sauer, B., and Sternberg, N. (1992) A positive selection vector for cloning high molecular weight DNA by the bacteriophage P1 system: improved cloning efficacy. *Proc. Natl. Acad. Sci. USA* **89**, 2056–2060.

Poralla, K., Muth, G., and Härtner, T. (2000) Hopanoids are formed during transition from substrate to aerial hyphae in *Streptomyces coelicolor* A3(2). *FEMS Microbiol. Lett.* 189, 93–95.

Qu, X., Lei, C., and Liu, W. (2011) Transcriptome mining of active biosynthetic pathways and their associated products in *Streptomyces flaveolus*. *Angew. Chem. Int. Ed.* **50**, 9651-9654.

Rabe, P., Samborskyy, M., Leadlay, P. F., and Dickschat, J. S. (2017) Isoafricanol synthase from *Streptomyces malaysiensis*. *Org. Biomol. Chem.* **15**, 2353-2358.

Rackham, E. J., Grüschow, S., and Goss, R. J. (2011) Revealing the first uridyl peptide antibiotic biosynthetic gene cluster and probing pacidamycin biosynthesis. *Bioeng. Bugs* 2, 218-221.

Rackham, E. J., Grüschow, S., Ragab, A. E., Dickens, S., and Goss, R. J. (2010) Pacidamycin biosynthesis: identification and heterologous expression of the first uridyl peptide antibiotic gene cluster. *ChemBiochem.* **11**, 1700-1709.

Ragab, A. E., Grüschow, S., Tromans, D. R., and Goss, R. J. (2011) Biogenesis of the unique 4',5'-dehydronucleoside of the uridyl peptide antibiotic pacidamycin. *J. Am. Chem. Soc.* ***133***,15288-15291.

Ramos, J. L., Martinez-Bueno, M., Molina-Henares, A. J., Terán, W., Watanabe, K., Zhang, X., Gallegos, M. T., Brennan, R., and Tobes, R. (2005) The TetR family of transcriptional repressors. *Microbiol. Mol. Biol Rev.* **69**, 326–356.

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

Reed, J., Stephenson, M. J., Miettinen, K., Brouwer, B., Leveau, A., Brett, P., Goss, R. J. M., Goosens, A., O'Connell, M. A., and Osbourn, A. (2017) A translational synthetic biology platform for rapid access to gram-scale quantities of novel drug-like molecules. *Metab. Eng.* **42**, 185**–**193.

Reeves, C. D., Murli, S., Ashley, G. W., Piagentini, M., Hutchinson, C. R., and McDaniel, R. (2001) Alteration of the substrate specificity of a modular polykeide synthase acyltransferase domain through site-specific mutagenesis. Biochemistry ***40****, 15464–15470.*

Reid, R., Piagentini, M., Rodriguez, E., Ashley, G., Viswanathan, N., Carney, J., Santi, D. V., Hutchinson, C. R., McDaniel, R., (2003) A model of structure and catalysis for ketoreductase domains in modular polyketide synthases. *Biochemistry* **42**, 72–79.

Repka, L. M., Chekan, J. R., Nair, S.K., and van der Donk, W. A. (2017) Mechanistic understanding of lanthipeptide biosynthetic enzymes. *Chem. Rev.* **117**, 5457–5520.

Rhoads, A., and Au, K, F. (2015) PacBio sequencing and its applications. *Genomics Proteomics Bioinformatics* **13**, 278-289.

Riclea, R., Citron, C. A., Rinkel, J., and Dickschat, J. S. (2014) Identification of isoafricanol and its terpene cyclase in *Streptomyces violaceusniger* using CLSA- NMR. *Chem. Commun.* **50**, 4228-4230.

Rigali, S., Nothaft, H., Noens, E. E., Schlicht, M., Colson, S., Müller, M., Joris, B., Koerten, H. K., Hopwood, D. A., Titgemeyer, F. and van Wezel, G. P. (2006) The sugar phosphotransferase system of *Streptomyces coelicolor* is regulated by the GntR-family regulator DasR and links N-acetylglucosamine metabolism to the control of development. *Molec. Microbiol*. **61**, 1237–1251.

Rohmer, M. (1999) The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. *Nat. Prod. Rep.* **16**, 565-574.

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

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., and

Barrell, B. (2000). Artemis: sequence visualization and annotation. Bioinformatics, **16**, 944–945.

Sahl, H. G. (1994) Gene-encoded antibiotics made in bacteria. *Ciba Found. Symp.* 186, 27-42.

Sato, Y., Horii, T., Maruyama, T., Honjo, M. (1989) Synthesis of the carbocyclic oxetanocin A1 and its related compounds. *Nucl. Acids Symp. Ser.* **21**, 125–126.

Sauter, H., Steglich, W., and Anke, T. (1999) Strobilurins: evolution of a new class of active substances. *Angew. Chem. Int. Ed.* **38**, 1328–1349.

Schneider, K., Chen, X. H., Vater, J., Franke, P., Nicholson, G., Borriss, R., and Süssmuth, R. D. (2007) Macrolactin is the polyketide biosynthesis product of the pks2 cluster of *Bacillus amyloliquefaciens* FZB42. *J. Nat. Prod.* **70**, 1417–1423.

Shao, Z. and Zhao, H. (2009). Manipulating Natural Product Biosynthetic Pathways via DNA Assembler. In: Current Protocols in Chemical Biology. John Wiley & Sons, Inc.

Shao, Z., Luo, Y. and Zhao, H. (2011). Rapid characterization and engineering of natural product biosynthetic pathways via DNA assembler. Mol. BioSyst. **7**, 1056–1059.

Sheehan, L.S., *et al.* (2006) Engineering of the spinosyn PKS: directing starter unit incorporation. *J. Nat. Prod.* **69**, 1702-1710.

Shen, B. (2015) Structural and evolutionary relationships of "AT-less" type I polyketide synthase ketosynthases. *Proc. Natl. Acad. Sci. USA* 112, 12693–12698.

Shen, B., Du, L., Sanchez, C., Edwards, D. J., Chen, M., Murrell, J. M. (2002) Cloning and characterization of the bleomycin biosynthetic gene cluster from *Streptomyces verticillus* ATCC15003. *J. Nat. Prod.* **65**, 422–431.

Sherman, D. H., Malpartida, F., Bibb, M. J., Kieser, H. M., Bibb, M. J., and Hopwood, D. A. (1989) Structure and deduced function of the granaticin-producing polyketide synthase gene cluster of *Streptomyces violaceoruber* Tü22. *EMBO J.* **8**, 2717–2725.

Shimada, N., Hasegawa, S., Harada, T., Tomisawa, T., Fujii, A., and Takita, T.

Oxetanocin, a novel nucleoside from bacteria. *J. Antibiot*. **39**, 1623–1625.

Sidebottom, A. M., Johnson, A. R., Karty, J. A., Trader, D. J., and Carlson, E. E. (2013) Integrated metabolomics approach facilitates discovery of an unpredicted natural product suite from *Streptomyces coelicolor* M145. *ACS Chem. Biol.* **8**, 2009–2016.

Smanski, M. J., Peterson, R. M., Huang, S. X., and Shen, B. (2012) Bacterial diterpene synthases: new opportunities for mechanistic enzymology and engineered biosynthesis. *Curr. Opin. Chem. Biol.* **16**, 132-141.

Sosio, M., Giusino, F., Cappellano, C., Bossi, E., Puglia, A.M. and Donadio, S. (2000). Artificial chromosomes for antibiotic-producing actinomycetes. *Nat. Biotechnol.* **18**, 343–345.

Sparks, T. C., Crouse, G. D., Dripps, J. E., Anzeveno, P., Martynow. J., Deamicis, C. V., and Gifford, J. (2008) Neural network-based QSAR and insecticide discovery: spinetoram. *J. Comput. Aided. Mol. Des.* **22**, 393–401.

Spencer, J. B., and Jordan, P. M. (1992) Purification and properties of 6-methylsalicylic acid synthase from *Penicillium patulum*. *Biochem. J.* **288**, 839–846.

Stachelhaus, T., Mootz M. D., and Marahiel, M. A. (1999) The specificity conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol*. **6**, 493–505.

Starcevic, A., Zucko, J., Simunkovic, J., Long, P. F., Cullum, J., and Hranueli, D. (2008) ClustScan: an integrated program package for the semi-automatic annotation of modular biosynthetic gene clusters and in silico prediction of novel chemical structures. *Nucl. Acids Res.* **36**, 6882–6892.

Stegmann, E., Frasch. H. J., Kilian, R., and Pozzi, R.. (2015) Self-resistance mechanisms of actinomycetes producing lipid II-targeting antibiotics. *Int. J. Med. Microbiol.* **305**, 190–195.

Stewart, Jr,C., Vickery, C. R., Burkart, M. D., and Noel, J. P (2013) Confluence of structural and chemical biology: plant polyketide synthases as biocatalysts for a bio-based future. *Curr. Opin. Plant Biol.* 16, 365–372.

Stierle, A., Strobel, G., Stierle, D., (1992) Taxol and taxane production by Taxomyces andreanae, an endophytic fungus of Pacific yew. *Science* **26**0, 214–216.

Stone, M. and Williams, D. (1992) On the evolution of functional secondary metabolites (natural products). *Molec. Microbiol.* **6**, 29–34.

Strieker, M., Tanović, A., and Marahiel, M. A. (2010) Nonribosomal peptide synthetases: structures and dynamics. *Curr. Opin. Struct. Biol.* **20**, 234–240

Strobel Jr, R. J., and Nakatsukasa, W. M. (1993). Response surface methods for optimizing Saccharop*olyspora spinosa*, a novel macrolide producer. J*. Ind. Microbiol*. **11**, 121–127.

Sugiyama, M., Sakamoto, M., Mochizuki, H., Nimi, O., and Nomi, R. (1983) Purification and characterization of streptomycin 6-kinase, an enzyme implicated in self-protection of a streptomycin-producing micro-organism. *J. Gen. Microbiol.* **129**, 1683–1687.

Sugiyama, M. (2015) Structural biological study of self-resistance determinants in antibiotic-producing actinomycetes. *J. Antibiot.* **68**, 543–550.

Sun, Y., Hong, H., Samborskyy, M., Mironenko, T., Leadlay, P. F., and Haydock, S. F. (2006) Organization of the biosynthetic gene cluster in *Streptomyces* sp. DSM 4137 for the novel neuroprotectant polyketide meridamycin. *Microbiology* **152**, 3507–3515.

Sun, Y., Hahn, F., Demydchuk, Y., Chettle, J., Tosin, M., Osada, H., Leadlay, P. F. (2010) In vitro reconstruction of tetronate RK-682 biosynthesis. *Nat. Chem. Biol.* **6**, 99–101.

Takatsu, T., Nakayama, H., Shimazu, A., Furihata, K., Ikeda, K., Furihata, K., Seto, H., and Otake, N. (1985) Rustmicin, a new macrolide antibiotic active against wheat stem rust fungus. *J. Antibiot*. **38**, 1806–1809.

Tan, G.-Y., Deng, K., Liu, X., Tao, H., Chang, Y., Chen, J., Chen, K., Sheng, Z., Deng, Z. and Liu, T. (2017). Heterologous biosynthesis of spinosad: an omics-guided large polyketide synthase gene cluster reconstitution in *Streptomyces*. *ACS Synth. Biol.* **6**, 995–1005.

Tang, L., Shah, S., Chung, L., Carney, J., Katz, L., Khosla, C. and Julien, B. (2000). Cloning and heterologous expression of the epothilone gene cluster. *Science* **287**, 640–642.

Tang, X., Li, J., Millán-Aguiñaga, N., Zhang, J. J., O'Neill, E. C., Ugalde, J. A., Jensen, P. R., Mantovani, S. M., and Moore, B. S. (2015) Identification of thiotetronic acid antibiotic biosynthetic pathways by target-directed genome mining. *ACS Chem. Biol.* **10**, 2841–2849.

*Tang, Y., Kim, C. Y., Mathews, I. I., Cane, D. E., and Khosla, C. (2006) The 2.7-Angstrom crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. Proc. Natl. Acad. Sci. USA **103**, 11124–11129.*

Tang, Y., Xia, L., Ding, X., Luo, Y., Huang, F. and Jiang, Y. (2011). Duplication of partial spinosyn biosynthetic gene cluster in *Saccharopolyspora spinosa* enhances spinosyn production. *FEMS Microbiol Lett.* **325**, 22–29.

Tanović, A., Samel, S. A., Essen, L. O., and Marahiel, M. A. (2008) Crystal structure of the termination module of a nonribosomal peptide synthetase. *Science* **321**, 659–663.

Tao, M., Wang, L., Wendt-Pienkowski, E., George, N. P., Galm, U., Zhang, G., Coughlin, J. M., and Shen, B. (2007) The tallysomycin biosynthetic gene cluster from *Streptoalloteichus hindustanus* E465-94 ATCC 31158 unveiling new insights into the biosynthesis of the bleomycin family of antitumor antibiotics. *Mol. Biosyst*. **3**, 60–74.

Tao, W., Yurkovich, M. E., Wen, S., Lebe, K. E., Samborskyy, M., Liu, Y., Yang, A., Liu, Y., Ju, Y., Deng, Z., Tosin, M., Sun, Y., and Leadlay, P. F. (2016) A genomics-led approach to deciphering the mechanism of thiotetronate antibiotic biosynthesis. *Chem. Sci.* **7**, 376–385.

Thanapipatsiri, A., Gomez-Escribano J. P., Song, L., Bibb, M. J., Al-Bassam, M., Chandra, G., Thamchaipenet, A., Challis G. L., and Bibb, M. J. (2016) Discovery of unusual biaryl polyketides by activation of a silent *Streptomyces venezuelae* biosynthetic gene cluster. *ChemBiochem*. **17**, 2189–2198.

Thompson, G. D., Michel, K. H., Yao, R. C., Mynderse, J. S., Mosburg,

C. T., Worden, T. V., Chio, E. H., Sparks, T. C., and Hutchins, S. C. (1997) The discovery of *Saccharopolyspora spinosa* and a new class of insect control products. *Down to Ear*th **52**, 1–5.

Tietz, J. I., Schwalen, C. J., Patel, P. S., Maxson, T., Blair, P. M., Tai, H.-C, Zakai, U. I., and Mitchell, D. A. (2017) A new genome-mining tool redefines the lasso peptide biosynthetic landscape. *Nat. Chem. Biol.* **13**, 470–478.

Tobin, M. B., Kovacevic, S., Madduri, K., Hoskins, J. A., Skatrud, P. L., Vining, L. C., Stuttard, C., and Miller J. R. (1991) Localization of the lysine epsilon-aminotransferase (lat) and delta-(L-alpha-aminoadipyl)-L-cysteinyl-D-valine synthetase (pcbAB) genes from *Streptomyces clavu*ligerus and production of lysine epsilon-aminotransferase activity in *Escherichia coli*. *J. Bact*eriol. **173**, 6223–6229.

Tortora, G. J., Funke, B. R., and Case, C. L. (1993) *Microbiology: an introduction*. Benjamin/Cummings Publlishing Co. Inc., California.

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

Tremblay, N., Hill, P., Conway, K. R. and Boddy, C. N. (2016) The use of clustermine360 for the analysis of polyketide and nonrivbosomal peptide biosynthetic pathways. *Methods Mol. Biol*. **1401**, 233–252.

Trindade, M., Zyl, L.J. van, Navarro-Fernández, J. and Abd Elrazak, A. (2015). Targeted metagenomics as a tool to tap into marine natural product diversity for the discovery and production of drug candidates. *Front. Microbiol*. 6.

Trippe, K., McPhail, K., Armstrong, D., Azevedo, M., and Banowetz, G. (2013) *Pseudomonas fluorescens* SBW25 produces furanomycin, a non-proteinogenic amino acid with selective antimicrobial properties. *BMC Microbiol.* **13**, 111.

Trujillo,M. E., Bacigalupe, R., Pujic, P., Igarashi, Y., Benito, P. Riesco, R., Médigue, C., and Normand, P. (2014) Genome features of the endophytic actinobacterium *Micromonospora lupini* strain Lupac 08: on the process of adaptation to an endophytic life style? PLoS One **9**, e10852.

Udwary, D. W., Zeigler, L., Asolkar, R. N., Singan, V., Lapidus, A., Fenical, W., Jensen, P. R., and Moore, B. S. (2007) Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*. *Proc. Natl. Acad. Sci. USA* 104, 10376–10381.

Vais, H., Williamson, M. S., Devonshire, A. L., and Usherwood, P. N. (1991) The molecular interactions of pyrethroid insecticides with insect and mammalian sodium channels. *Pest Manag. Sci.* **57**, 877–888.

Van der Donk, W. A., and Nair, S. K. (2014) Structure and mechanism of lanthipeptide biosynthetic enzymes. *Curr. Opin. Struct. Biol.* **10**, 58–66.

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

Van Epps, H. L. (2006) René Dubos: unearthing antibiotics. *J. Exp. Med.* **203**, 259.

van Wezel, G. P., and McDowall, K. J. (2011) The regulation of the secondary metabolism of *Streptomyces*: new links and experimental advances. *Nat. Prod. Rep.* **28**, 1311–1333.

Velayati, A. A., Masjedi, M. R., Farnia, P., Tabarsi, P., Ghanavi, J., ZiaZarifi, A. H. and Hoffner, S. E. (2009) Emergence of new forms of totally drug-resistant tuberculosis bacilli: super extensively drug-resistant tuberculosis or totally drug-resistant strains in Iran. *Chest* **136**, 420–425.

Vicente, C. M., Payero, T. D., Santos-Aberturas, J., Barreales, E. G., de Pedro, A., and Aparicio, J. F. (2015) Pathway-specific regulation revisited: cross-regulation of multiple disparate gene clusters by PAS-LuxR transcriptional regulators. *Appl. Microbiol. Biotechnol*. **99**, 5123–5135.

Voelkerding, K. V., Dames, S. A., Durtschi, J. D. (2009) Next-generation sequencing: from basic research to diagnostics. *Clin. Chem*. **55**, 641–658.

Waldron, C., Madduri, K., Crawford, K., Merlo, D.J., Treadway, P., Broughton, M.C. & Baltz, R.H. (2000). A cluster of genes for the biosynthesis of spinosyns, novel macrolide insect control agents produced by Saccharo*polyspora spinosa*. *Antonie Van Leeuwenhoek* **78**, 385–390.

Waldron, C., Matsushima, P., Rosteck, P. R. Jr, Broughton, M. C., Turner, J., Madduri, K., Crawford, K. P., Merlo, D. J., and Baltz, R. H. (2001) Cloning and analysis of the spinosad biosynthetic gene cluster of *Saccharopolyspora spinosa*. *Chem. Biol.* **8**, 487–499.

Walsh, F. M. and Amyes, S. G. (2004) Microbiology and drug resistance mechanisms of fully resistant pathogens. *Curr. Opin. Microbiol.* **7**, 439–444.

Wang, H. X., Chen, Y. Y., Ge, L., Fang, T. T., Meng, J., Liu, Z., Fang, X. Y., Ni, S., Lin, C., Wu, Y., Wang, M. L., Shi, N. N., He, H. G., Hong, K., and Shen, Y. M., (2013) PCR screening reveals considerable unexploited biosynthetic potential of ansamycins and a mysterious family of AHBA-containing natural products in actinomycetes. *J. Appl. Microbiol*. **115**, 77–85.

Wang, H. X., Chen, Y. Y., Ge, L., Fang, T. T., Meng, J., Liu, Z., Fang, X. Y.,

Wang, X., Chen, J., and Quinn, P. (eds) Reprogramming microbial metabolic pathways. Springer Verlag, Berlin. pp. 115–138.

Wang, Y., Wang, Y., and Zhang, S. (2008) High frequency transformation of the industrial erythromycin-producing bacterium *Saccharopolyspora erythraea*. *Biotechnol. Lett*. **30**, 357–361.

Weissman, K. J. (2009) Introduction to polyketide biosynthesis. *Methods Enzymol.* **459**, 3-16

Weissman, K. J. (2015) The structural biology of biosynthetic megaenzymes. *Nat. Chem. Biol*., **11**, 660-670.

Weissman, K. J. (2017) Polyketide stereocontrol: a study in chemical biology. *Beilstein J. Org. Chem*. **13**, 348–371.

Weng, J. K., Philippe, R. N., and Noel, J. P. (2012) The rise of chemodiversity in plants. *Science* **336**, 1667–1670.

Wenzel, S. C., Bode, H. B., Kochems, I., and Müller, R. (2008) Biosynthetic pathway for the structurally unique *ansa* compound kendomycin. *ChemBiochem.* **9**, 2711–2721.

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

Wilkinson, C.J., Hughes-Thomas, Z.A., Martin, C.J., Böhm, I., Mironenko, T., Deacon, M., Wheatcroft, M., Wirtz, G., Staunton, J. and Leadlay, P.F. (2002). Increasing the efficiency of heterologous promoters in actinomycetes. *J. Mol. Microbiol. Biotechnol*. **4**, 417–426.

Williams, D. E., Dalisay, D. S., Patrick, B. O., Matainaho, T., Andrusiak, K., Deshpande, R., Myers, C. L., Piotrowski, J. S., Boone, C., Yoshida, M., and Andersen, R. J. (2011) Padanamides A and B, highly modified linear tetrapeptides produced in culture by a *Streptomyces* sp. isolated from a marine sediment. *Org. Lett.* **13**, 3936-3939.

Winn, M., Fyans, J. K., Zhuo, Y., and Micklefield J. (2016) Recent advances in engineering nonribosomal peptide assembly lines. *Nat. Prod. Rep.* **33**, 317-347.

Xu, Z., Wang, Y., Chater, K. F., Ou, H. Y., Xu, H. H., Deng, Z., and Tao, M. (2017) Large-scale transposition mutagenesis of *Streptomyces coelicolor* identifies hundreds of genes influencing antibiotic biosynthesis. *Appl. Environ. Microbiol.* **83**, e02889-16.

Xue, C., Duan, Y., Zhao, F., and Lu, W., (2013). Stepwise increase of spinosad production in Saccharopolyspora spinosa by metabolic engineering. Biochem. Eng. J., **72**, 90–95.

Yim, G., Wang, H. H. and Davies, J. (2006) The truth about antibiotics. *Int. J. Med.Microbiol.* **296**, 163–170.
Yim, G., Wang, H. H. and Davies, J. (2007) Antibiotics as signalling molecules. *Philos. Trans. Roy. Soc. Lond. B Biol. Sci.* **362**, 1195–1200.

Zaburannyi, N., Rabyk, M., Ostash, B., Fedorenko, V., and Luzhetskyy, A. (2014). Insights into naturally minimised *Streptomyces albus* J1074 genome. *BMC Genomics* **15**, 97.

Zalacain, M., and Cundliffe, E. (1990) Methylation of 23S ribosomal RNA due to carB, an antibiotic-resistance determinant from the carbomycin producer, *Streptomyces thermotolerans*. *Eur. J. Biochem.* **189**, 67-72.

Zeng, H., Tan, H., and Li, J. (2002) Cloning and function of sanQ: a gene involved in nikkomycin biosynthesis of *Streptomyces ansochromogenes*. *Curr. Microbiol.* **45**, 175–179.

Zhan J. (2009) Biosynthesis of bacterial aromatic polyketides. *Curr. Top. Med. Chem.* **9**, 1958–1610.

Zhang, L., Hashimoto, T., Qin, B., Hashimoto, J., Kozone, I., Kawahara, T., Okada, M., Awakawa, T., Ito, T., Asakawa, Y., Ueki, M., Takahashi, S., Osada, H., Wakimoto, T., Ikeda, H., Shin-Ya, K., and Abe I. (2017) Characterization of giant modular PKSs provides insight into genetic mechanism for structural diversification of aminopolyol polyketides. Angew. Chem. Int. Ed. ***56****, 1740-1745.*

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

Yurkovich, M. E. (2016) PhD thesis, University of Cambridge.

Zhang, W., Ntai, I., Bolla, M. L., Malcolmson, S. J., Kahne, D., Kelleher, N. L., and Walsh, C. T. (2011) Nine enzymes are required for assembly of the pacidamycin group of peptidyl nucleoside antibiotics. *J. Am. Chem. Soc.* **133**, 5240-5243.

Zhang, W., Ntai, I., Kelleher, N. L., and Walsh, C. T. (2011) tRNA-dependent peptide bond formation by the transferase PacB in biosynthesis of the pacidamycin group of pentapeptidyl nucleoside antibiotics. *Proc. Natl. Acad. Sci. USA* **108**, 12249-12253.

Zhang, Z., Pan, H. X., and Tang, G. L. (2017) New insights into bacterial type II polyketide biosynthesis. *F1000Res*. **6**, 172.

Zhao, C., Coughlin, J. M., Ju, J., Zhu, D., Wendt-Pienkowski, E., Zhou, X., Wang, Z., Shen, B., and Deng, Z. (2010) Oxazolomycin biosynthesis in *Streptomyces albus* JA3453 featuring an "acyltransferase-less" type I polyketide synthase that incorporates two distinct extender units. *J. Biol Chem*. **285**, 20097–20108.

Zhao, G. S., Li, S. R., Wang, Y. Y., Hao, H. L., Shen. Y. M., and Lu, C. H. (2013) 16,17-dihydroxycyclooctatin, a new diterpene from *Streptomyces* sp. LZ35. *Drug Discov. Ther.* 7, 185-188.

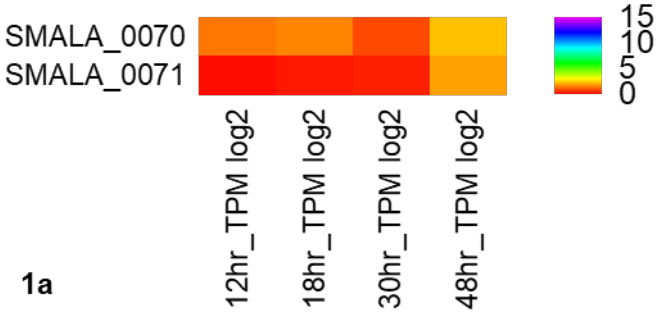
Zhao, J., Xue, Q., Niu, G.-G., Xue, L., Shen, G.-H., and Jun-zhi Du, J.-Z. (2013) Extracellular enzyme production and fungal mycelia degradation by antagonistic *Streptomyces* induced by fungal mycelia preparation of cucurbit plant pathogens'. *Ann. Microbiol.* **63**, 809–812.

Zhao, Q., He, Q., Ding, W., Tang, M., Kang, Q., Yu, Y., Deng, W., Zhang, Q., Fang, J., Tang, G., and Liu, W. (2008) Characterization of the azinomycin B biosynthetic gene cluster revealing a different iterative type I polyketide synthase for naphthoate biosynthesis. *Chem. Biol.* **15**, 693–705.

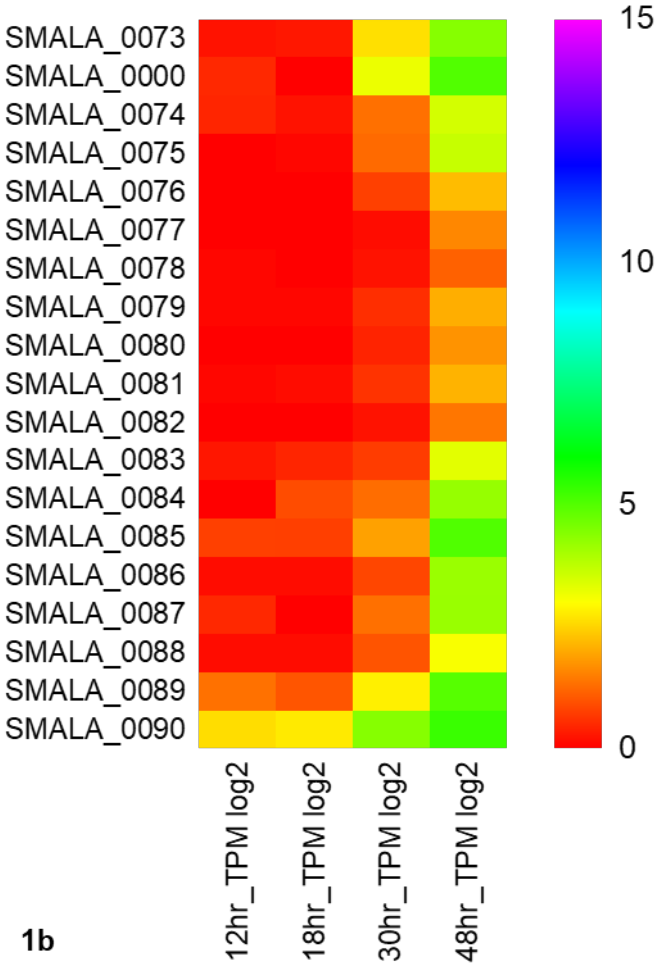
Zhou, Y., Murphy, A.C., Samborskyy, M., Prediger, P., Dias, L.C. and Leadlay, P.F. (2015). Iterative mechanism of macrodiolide formation in the anticancer compound conglobatin. *Chem. Biol.*, **22**, 745–754.

Ziemert, N., Podell, S., Penn, K., Badger, J. H., Allen, E., and Jensen, P. R. (2012) The natural product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLoS One* **7**, e34064

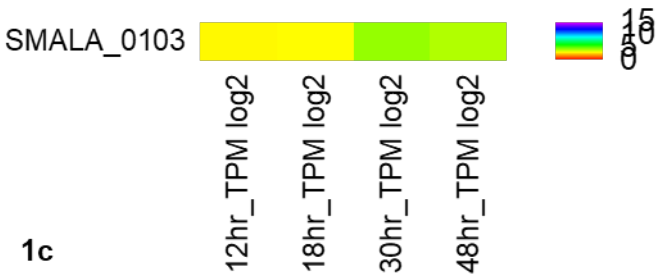
Appendix 1



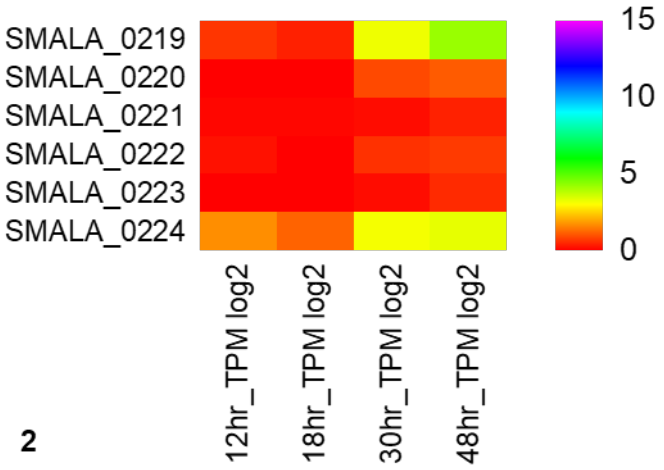
1a



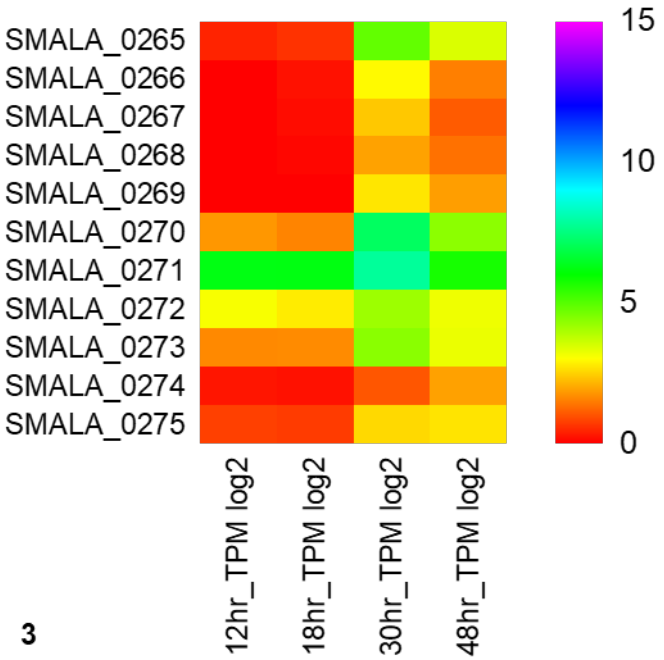
1b


SMALA_0103
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
15
10
0

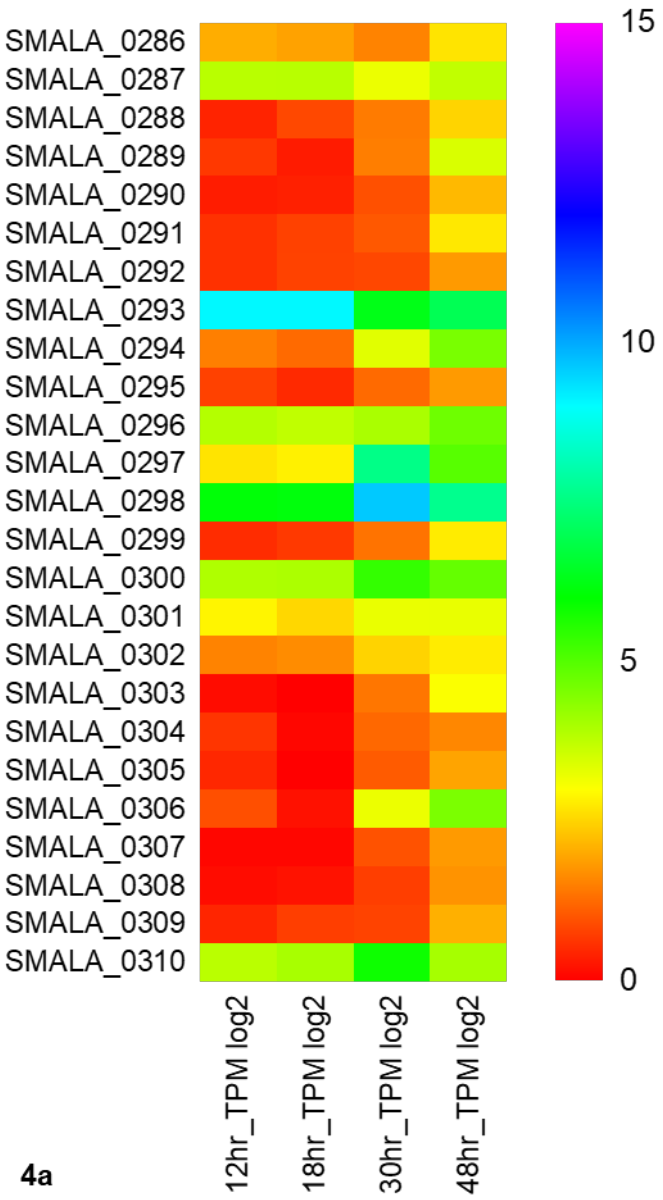
1c


SMALA_0219
SMALA_0220
SMALA_0221
SMALA_0222
SMALA_0223
SMALA_0224
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
15
10
5
0

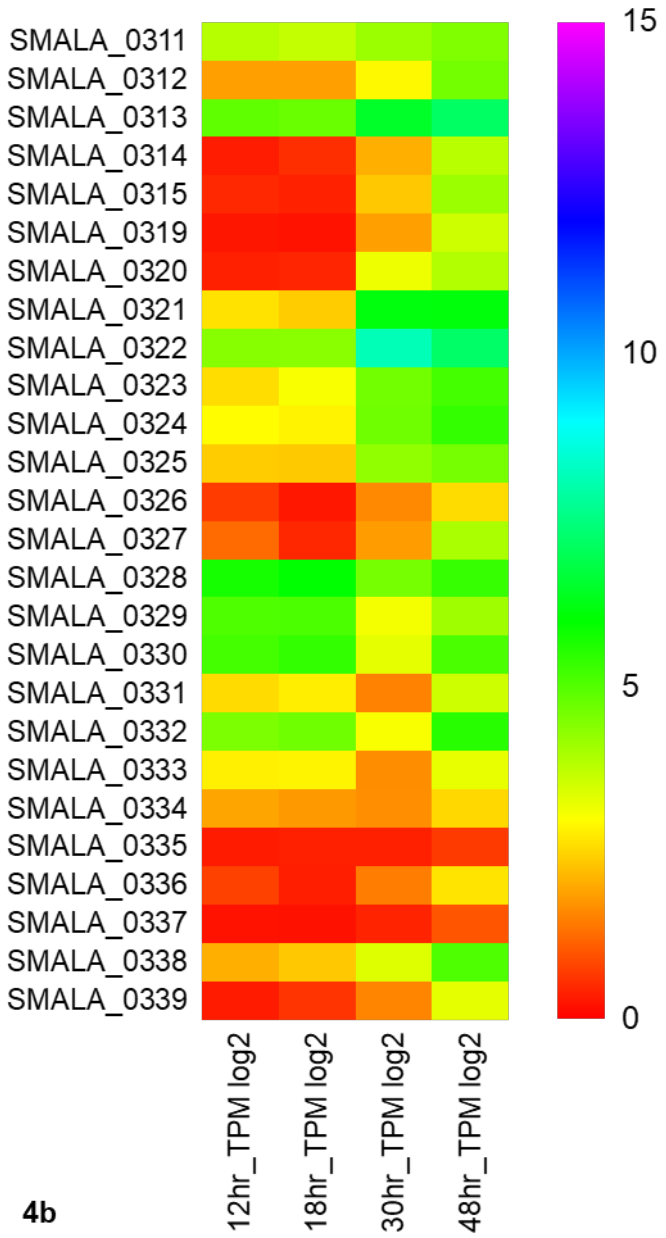
2


SMALA_0265
SMALA_0266
SMALA_0267
SMALA_0268
SMALA_0269
SMALA_0270
SMALA_0271
SMALA_0272
SMALA_0273
SMALA_0274
SMALA_0275
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
15
10
5
0

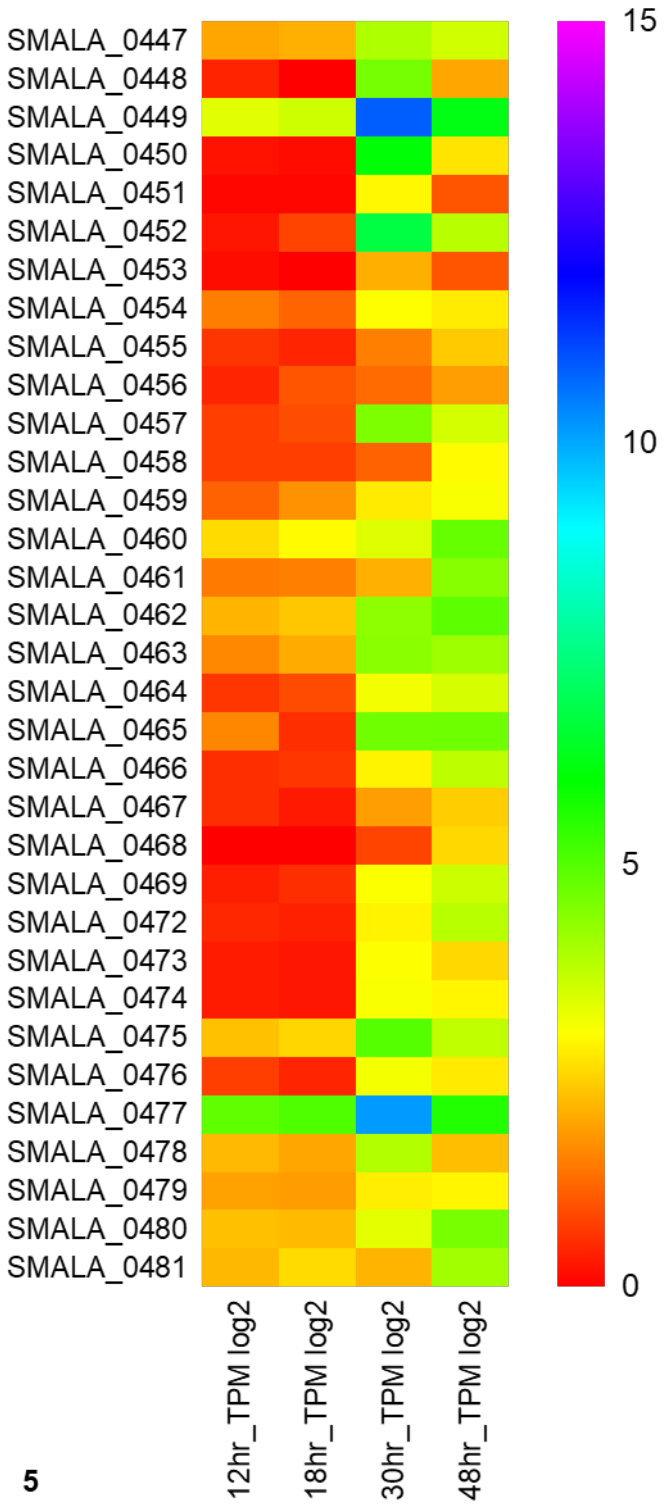
3

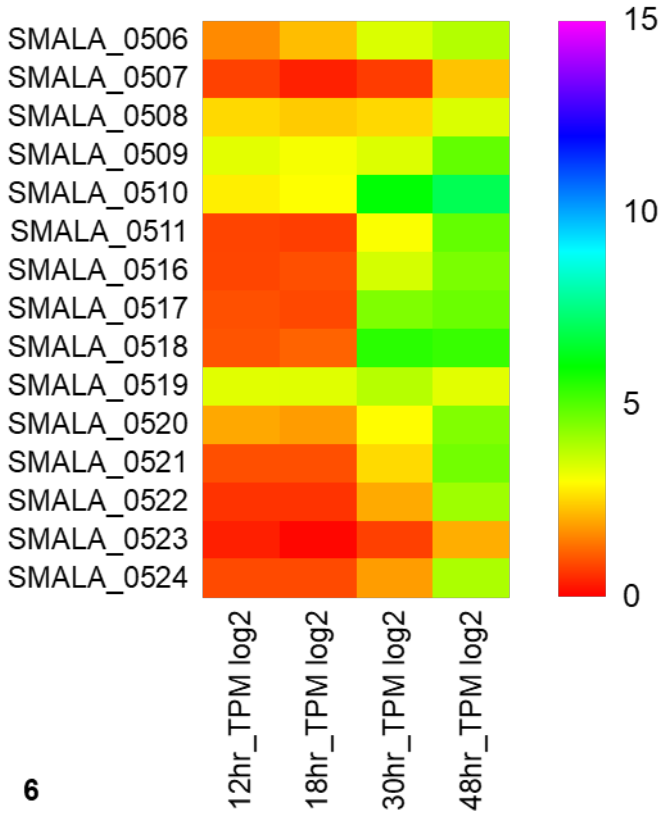

SMALA_0286
SMALA_0287
SMALA_0288
SMALA_0289
SMALA_0290
SMALA_0291
SMALA_0292
SMALA_0293
SMALA_0294
SMALA_0295
SMALA_0296
SMALA_0297
SMALA_0298
SMALA_0299
SMALA_0300
SMALA_0301
SMALA_0302
SMALA_0303
SMALA_0304
SMALA_0305
SMALA_0306
SMALA_0307
SMALA_0308
SMALA_0309
SMALA_0310
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
15
10
5
0

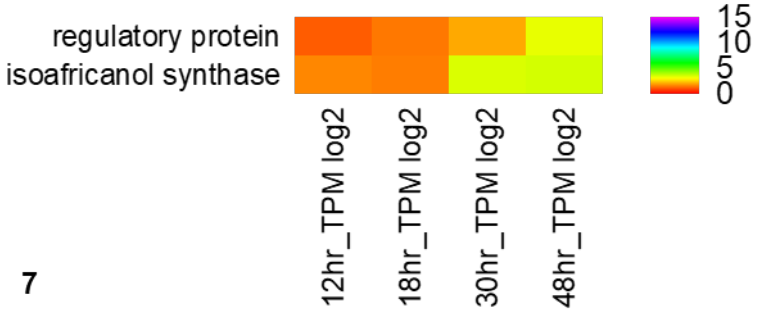
4a

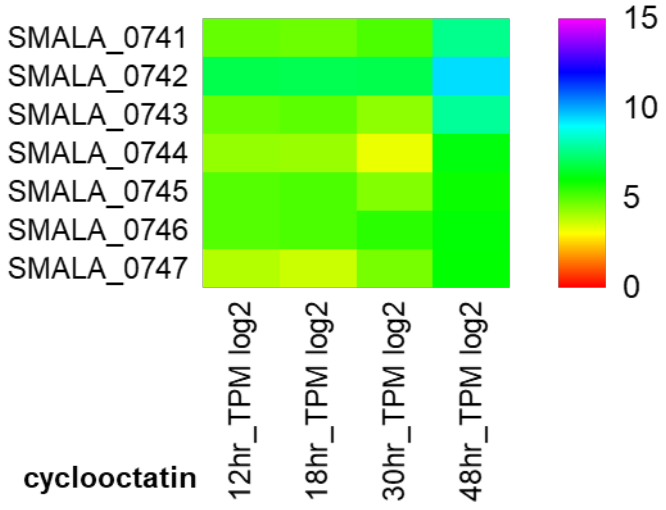

SMALA_0311
SMALA_0312
SMALA_0313
SMALA_0314
SMALA_0315
SMALA_0319
SMALA_0320
SMALA_0321
SMALA_0322
SMALA_0323
SMALA_0324
SMALA_0325
SMALA_0326
SMALA_0327
SMALA_0328
SMALA_0329
SMALA_0330
SMALA_0331
SMALA_0332
SMALA_0333
SMALA_0334
SMALA_0335
SMALA_0336
SMALA_0337
SMALA_0338
SMALA_0339
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
15
10
5
0

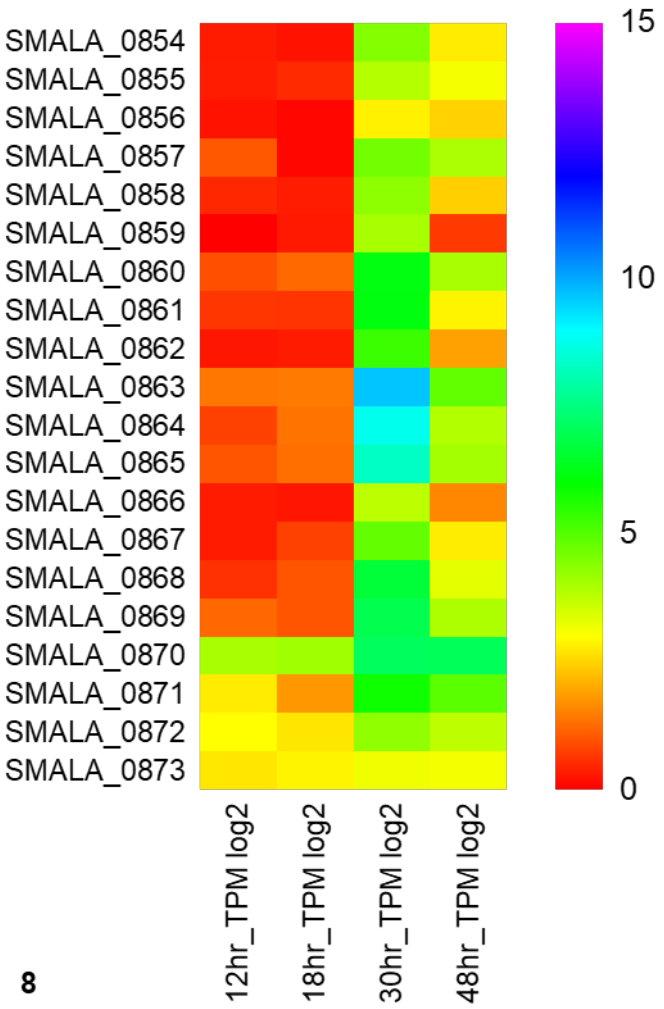
4b


SMALA_0447
SMALA_0448
SMALA_0449
SMALA_0450
SMALA_0451
SMALA_0452
SMALA_0453
SMALA_0454
SMALA_0455
SMALA_0456
SMALA_0457
SMALA_0458
SMALA_0459
SMALA_0460
SMALA_0461
SMALA_0462
SMALA_0463
SMALA_0464
SMALA_0465
SMALA_0466
SMALA_0467
SMALA_0468
SMALA_0469
SMALA_0472
SMALA_0473
SMALA_0474
SMALA_0475
SMALA_0476
SMALA_0477
SMALA_0478
SMALA_0479
SMALA_0480
SMALA_0481
15
10
5
0
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
5

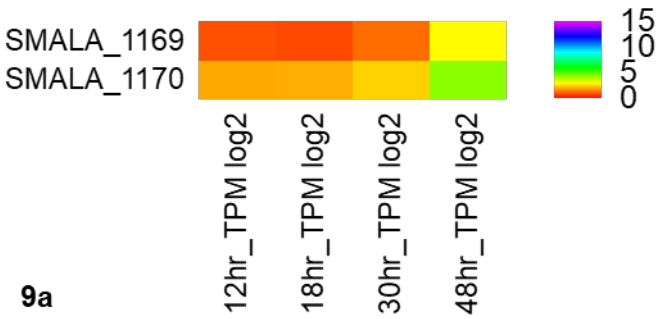

SMALA_0506
SMALA_0507
SMALA_0508
SMALA_0509
SMALA_0510
SMALA_0511
SMALA_0516
SMALA_0517
SMALA_0518
SMALA_0519
SMALA_0520
SMALA_0521
SMALA_0522
SMALA_0523
SMALA_0524
15
10
5
0
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
6


regulatory protein
isoafricanol synthase
15
10
5
0
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
7

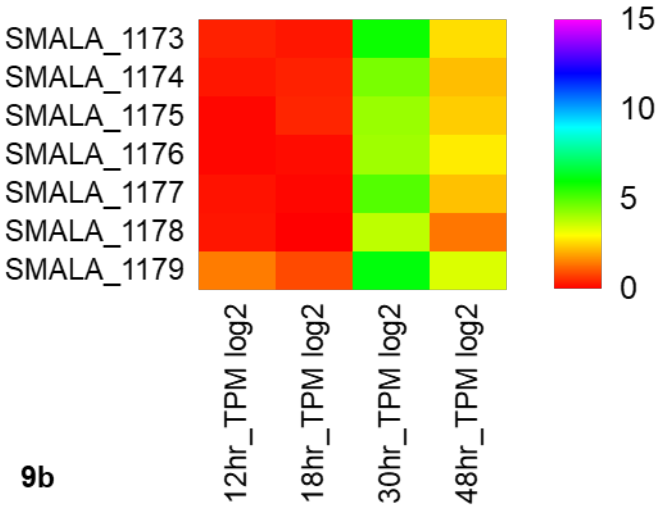

SMALA_0741
SMALA_0742
SMALA_0743
SMALA_0744
SMALA_0745
SMALA_0746
SMALA_0747
15
10
5
0
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
cyclooctatin


SMALA_0854
SMALA_0855
SMALA_0856
SMALA_0857
SMALA_0858
SMALA_0859
SMALA_0860
SMALA_0861
SMALA_0862
SMALA_0863
SMALA_0864
SMALA_0865
SMALA_0866
SMALA_0867
SMALA_0868
SMALA_0869
SMALA_0870
SMALA_0871
SMALA_0872
SMALA_0873
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
15
10
5
0

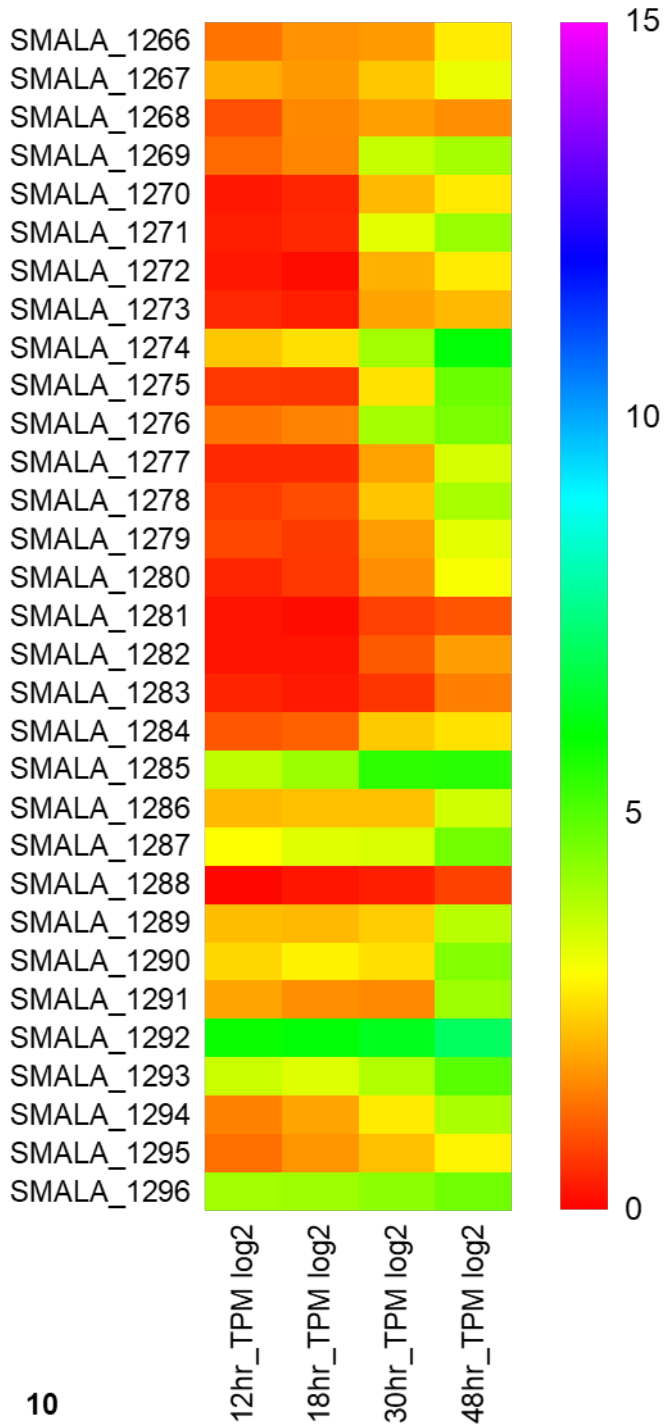
8

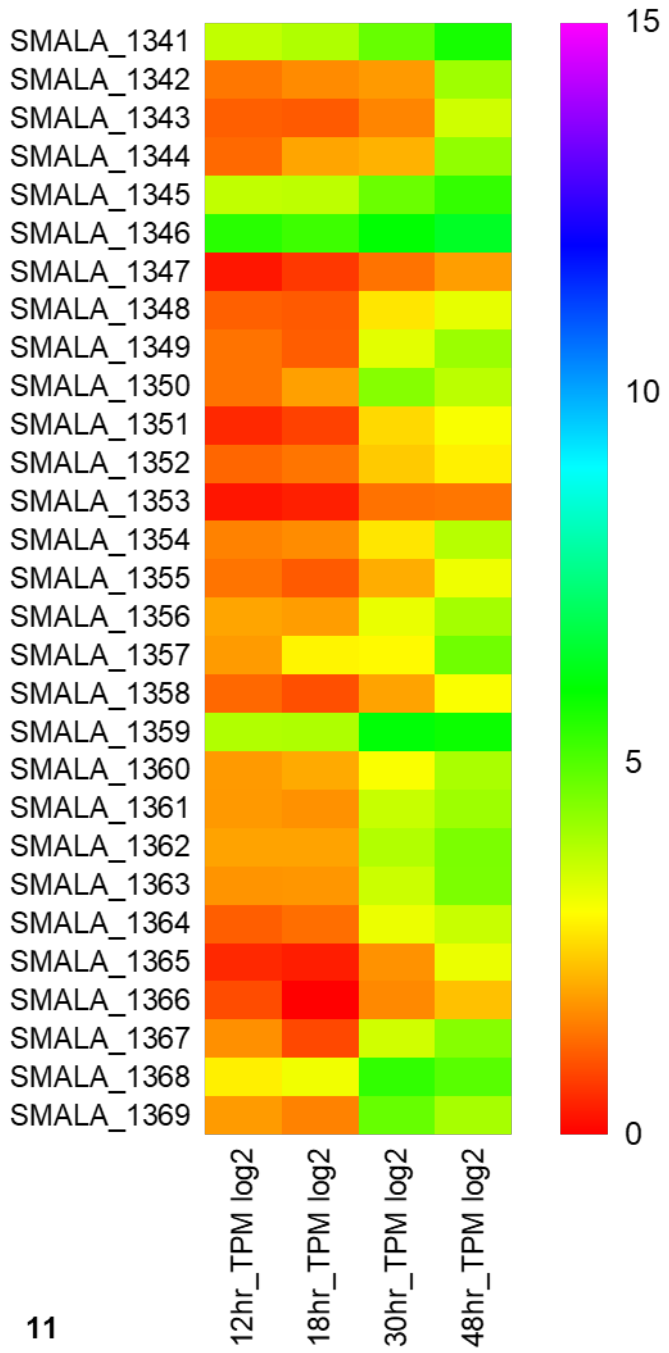

SMALA_1169
SMALA_1170
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
15
10
5
0

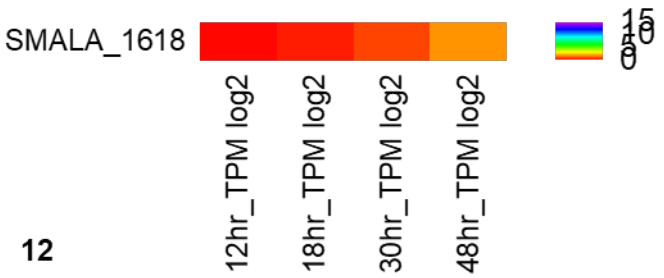
9a


SMALA_1173
SMALA_1174
SMALA_1175
SMALA_1176
SMALA_1177
SMALA_1178
SMALA_1179
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
15
10
5
0

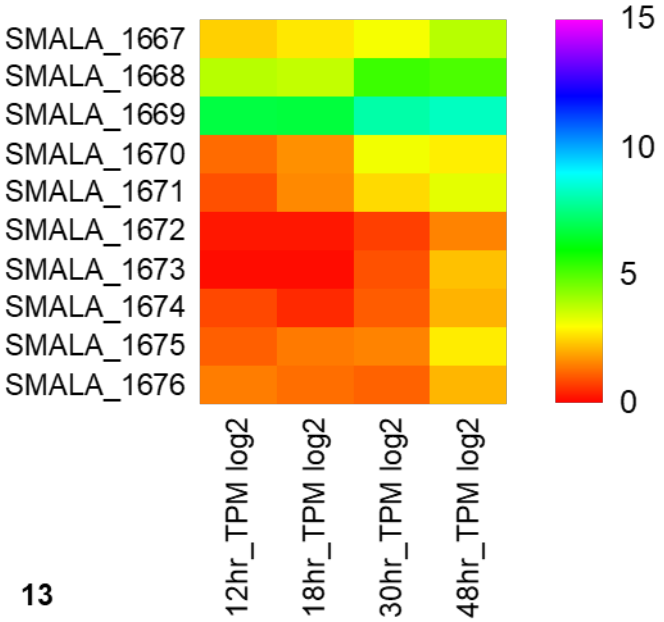
9b


SMALA_1266
SMALA_1267
SMALA_1268
SMALA_1269
SMALA_1270
SMALA_1271
SMALA_1272
SMALA_1273
SMALA_1274
SMALA_1275
SMALA_1276
SMALA_1277
SMALA_1278
SMALA_1279
SMALA_1280
SMALA_1281
SMALA_1282
SMALA_1283
SMALA_1284
SMALA_1285
SMALA_1286
SMALA_1287
SMALA_1288
SMALA_1289
SMALA_1290
SMALA_1291
SMALA_1292
SMALA_1293
SMALA_1294
SMALA_1295
SMALA_1296
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
15
10
5
0
10

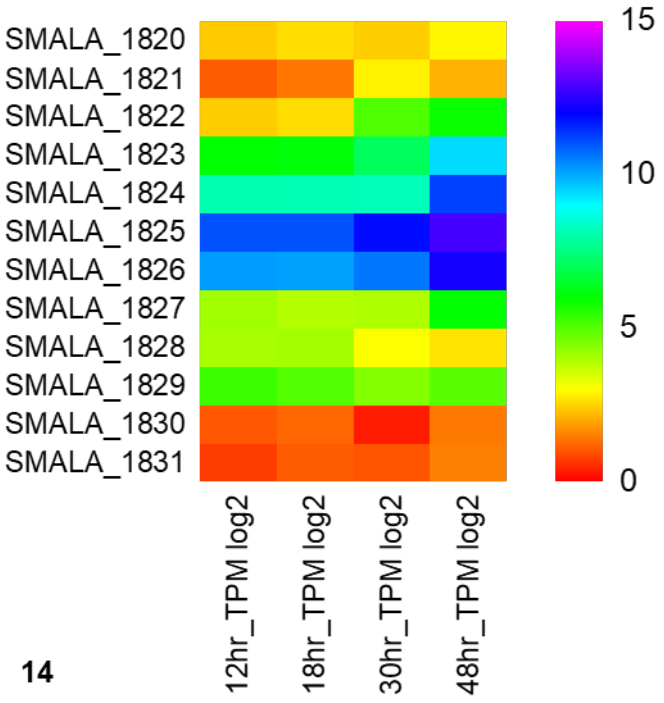

SMALA_1341
SMALA_1342
SMALA_1343
SMALA_1344
SMALA_1345
SMALA_1346
SMALA_1347
SMALA_1348
SMALA_1349
SMALA_1350
SMALA_1351
SMALA_1352
SMALA_1353
SMALA_1354
SMALA_1355
SMALA_1356
SMALA_1357
SMALA_1358
SMALA_1359
SMALA_1360
SMALA_1361
SMALA_1362
SMALA_1363
SMALA_1364
SMALA_1365
SMALA_1366
SMALA_1367
SMALA_1368
SMALA_1369
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
15
10
5
0
11


SMALA_1618
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
15
10
0

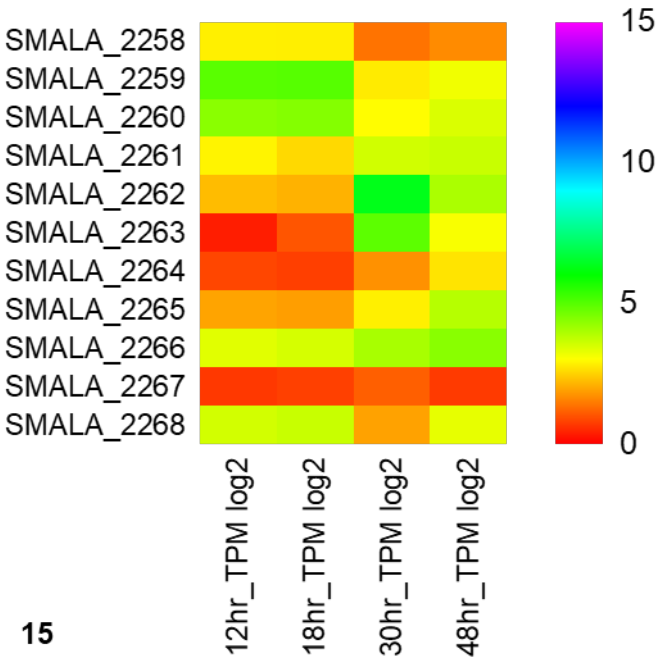
12


SMALA_1667
SMALA_1668
SMALA_1669
SMALA_1670
SMALA_1671
SMALA_1672
SMALA_1673
SMALA_1674
SMALA_1675
SMALA_1676
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
15
10
5
0

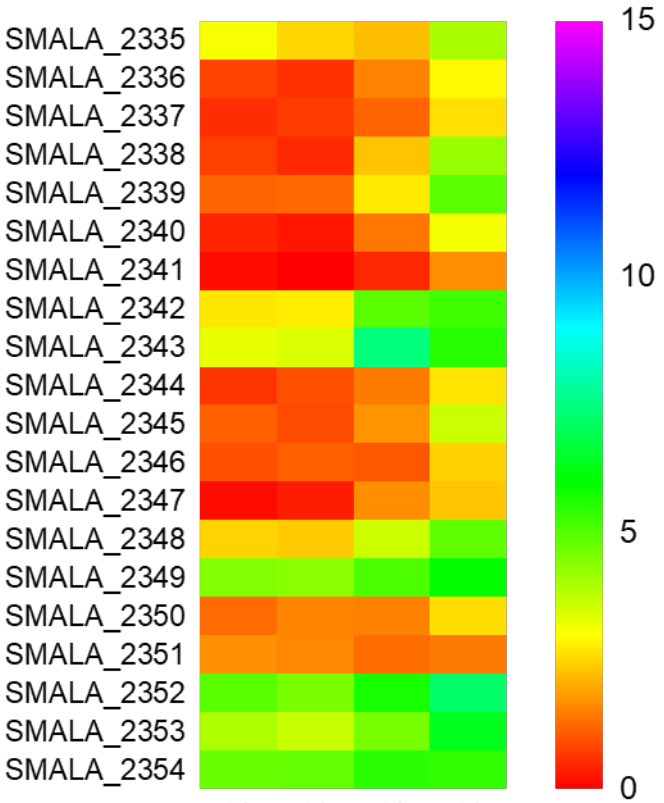
13


SMALA_1820
SMALA_1821
SMALA_1822
SMALA_1823
SMALA_1824
SMALA_1825
SMALA_1826
SMALA_1827
SMALA_1828
SMALA_1829
SMALA_1830
SMALA_1831
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
15
10
5
0

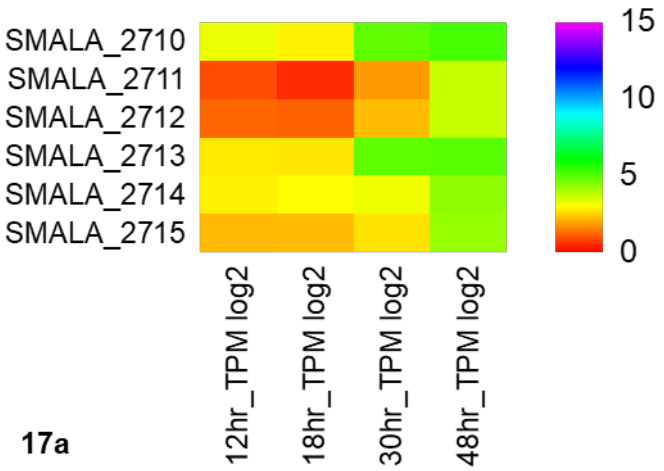
14


SMALA_2258
SMALA_2259
SMALA_2260
SMALA_2261
SMALA_2262
SMALA_2263
SMALA_2264
SMALA_2265
SMALA_2266
SMALA_2267
SMALA_2268
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
15
10
5
0

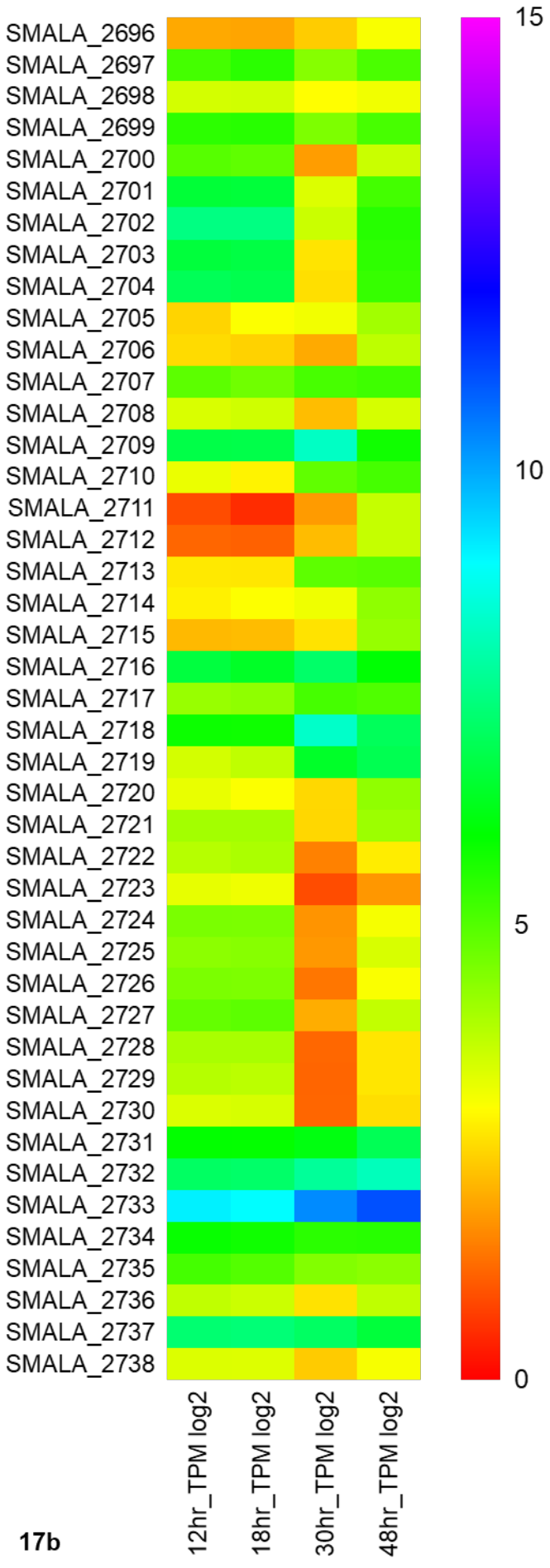
15


SMALA_2335
SMALA_2336
SMALA_2337
SMALA_2338
SMALA_2339
SMALA_2340
SMALA_2341
SMALA_2342
SMALA_2343
SMALA_2344
SMALA_2345
SMALA_2346
SMALA_2347
SMALA_2348
SMALA_2349
SMALA_2350
SMALA_2351
SMALA_2352
SMALA_2353
SMALA_2354
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
15
10
5
0

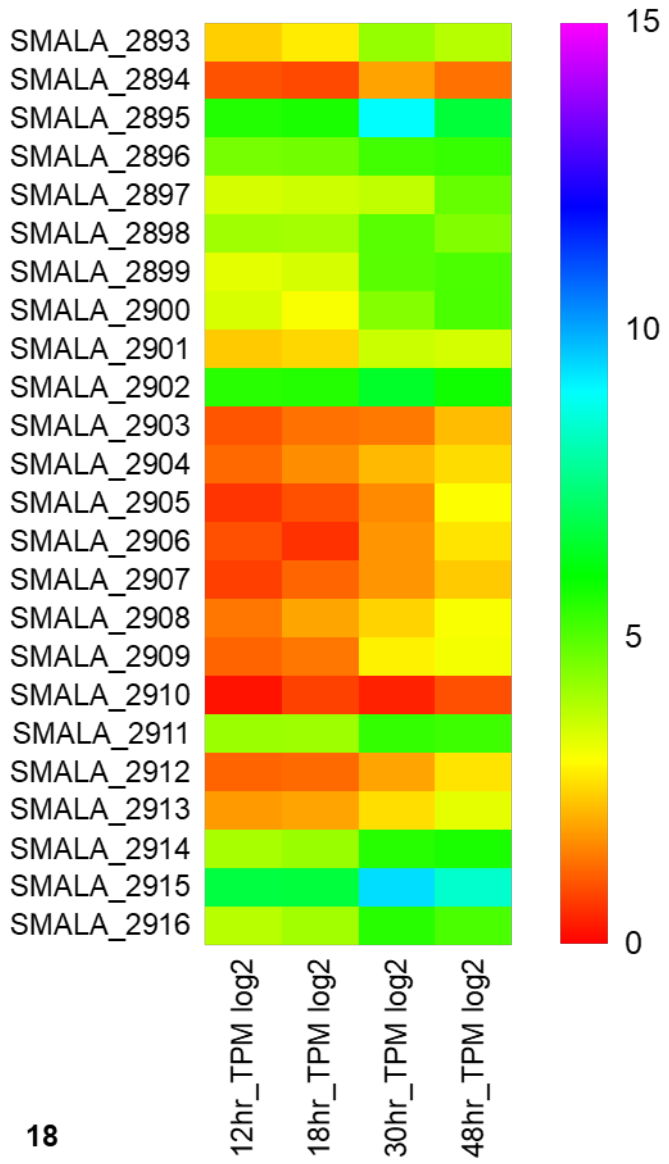
16

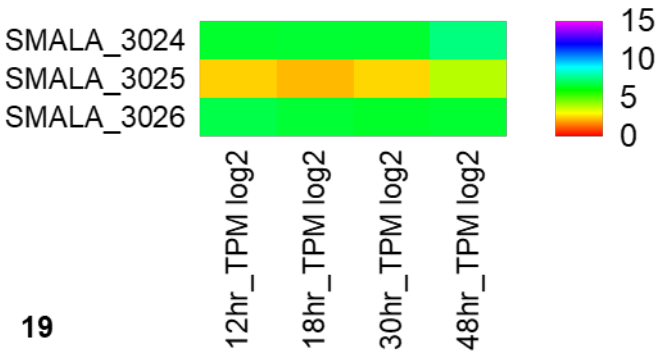

SMALA_2710
SMALA_2711
SMALA_2712
SMALA_2713
SMALA_2714
SMALA_2715
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
15
10
5
0

17a

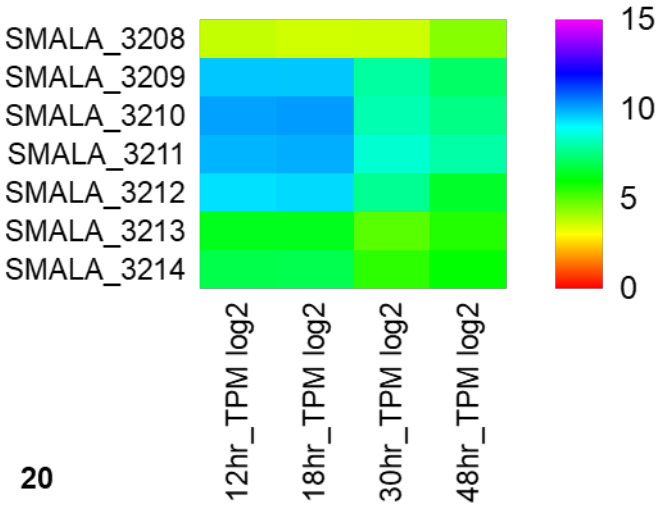

SMALA_2696
SMALA_2697
SMALA_2698
SMALA_2699
SMALA_2700
SMALA_2701
SMALA_2702
SMALA_2703
SMALA_2704
SMALA_2705
SMALA_2706
SMALA_2707
SMALA_2708
SMALA_2709
SMALA_2710
SMALA_2711
SMALA_2712
SMALA_2713
SMALA_2714
SMALA_2715
SMALA_2716
SMALA_2717
SMALA_2718
SMALA_2719
SMALA_2720
SMALA_2721
SMALA_2722
SMALA_2723
SMALA_2724
SMALA_2725
SMALA_2726
SMALA_2727
SMALA_2728
SMALA_2729
SMALA_2730
SMALA_2731
SMALA_2732
SMALA_2733
SMALA_2734
SMALA_2735
SMALA_2736
SMALA_2737
SMALA_2738
15
10
5
0
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2

17b

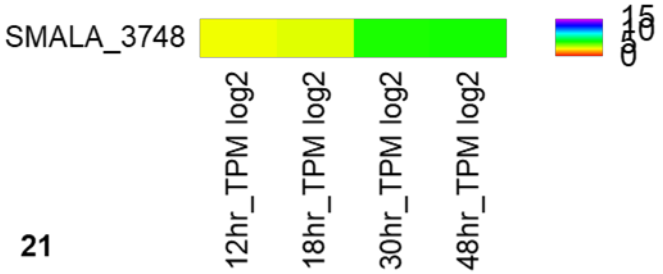

SMALA_2893
SMALA_2894
SMALA_2895
SMALA_2896
SMALA_2897
SMALA_2898
SMALA_2899
SMALA_2900
SMALA_2901
SMALA_2902
SMALA_2903
SMALA_2904
SMALA_2905
SMALA_2906
SMALA_2907
SMALA_2908
SMALA_2909
SMALA_2910
SMALA_2911
SMALA_2912
SMALA_2913
SMALA_2914
SMALA_2915
SMALA_2916
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
15
10
5
0
18


SMALA_3024
SMALA_3025
SMALA_3026
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
15
10
5
0

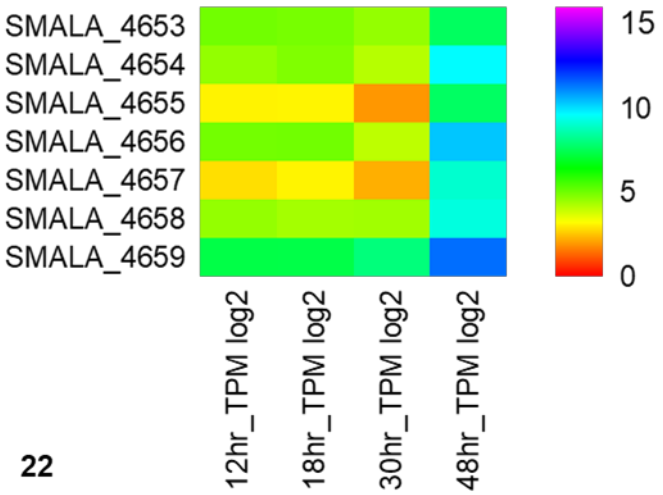
19


SMALA_3208
SMALA_3209
SMALA_3210
SMALA_3211
SMALA_3212
SMALA_3213
SMALA_3214
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
15
10
5
0

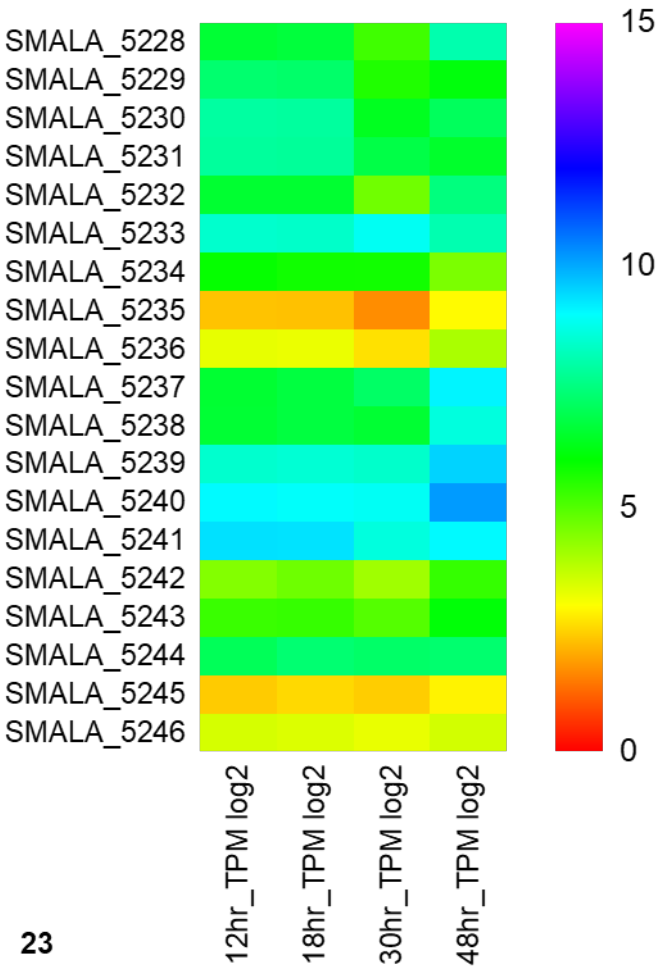
20


SMALA_3748
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2

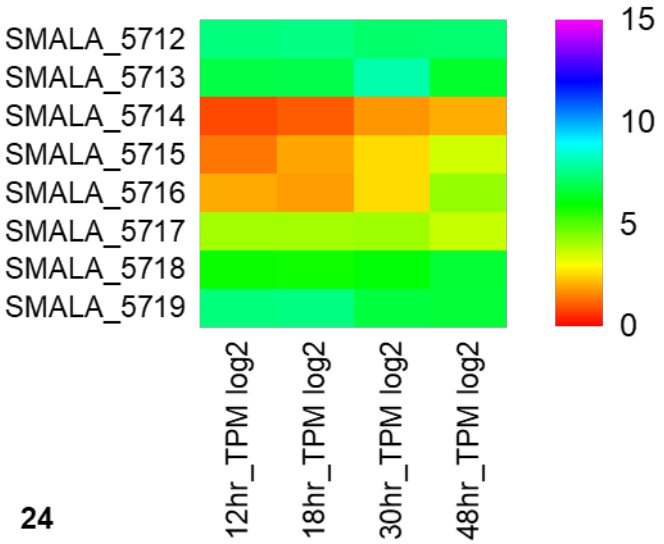
21


SMALA_4653
SMALA_4654
SMALA_4655
SMALA_4656
SMALA_4657
SMALA_4658
SMALA_4659
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
15
10
5
0

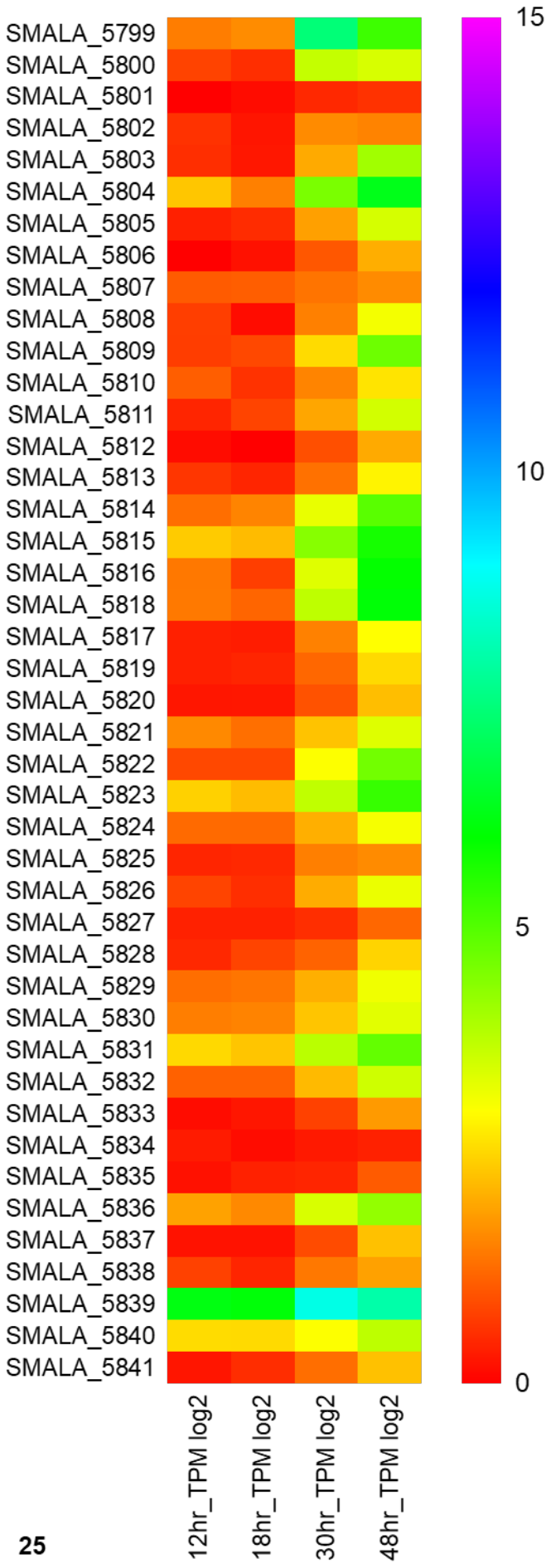
22

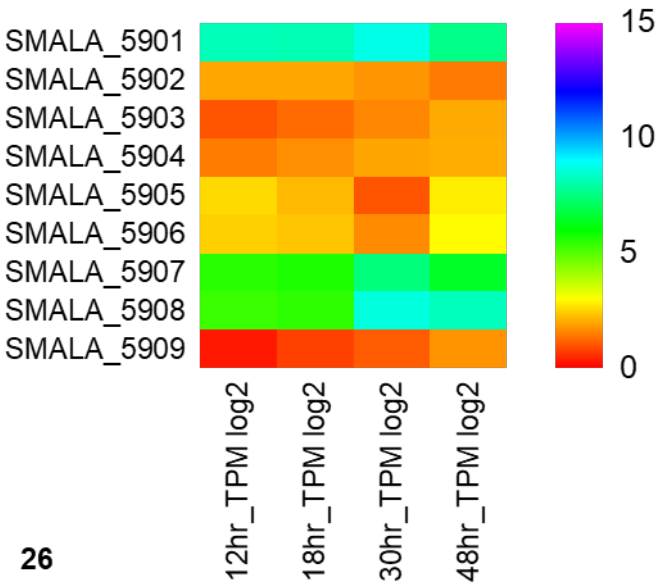

SMALA_5228
SMALA_5229
SMALA_5230
SMALA_5231
SMALA_5232
SMALA_5233
SMALA_5234
SMALA_5235
SMALA_5236
SMALA_5237
SMALA_5238
SMALA_5239
SMALA_5240
SMALA_5241
SMALA_5242
SMALA_5243
SMALA_5244
SMALA_5245
SMALA_5246
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
15
10
5
0

23

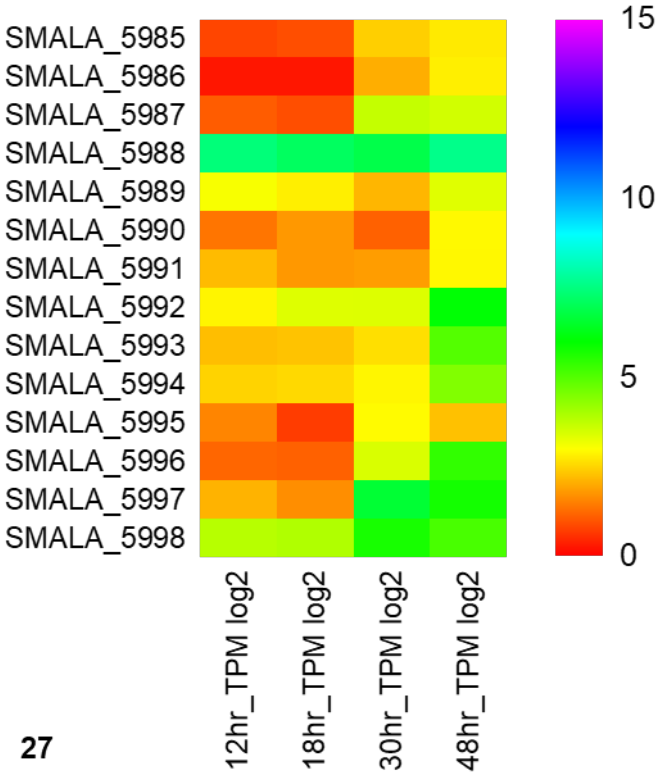

SMALA_5712
SMALA_5713
SMALA_5714
SMALA_5715
SMALA_5716
SMALA_5717
SMALA_5718
SMALA_5719
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
15
10
5
0

24

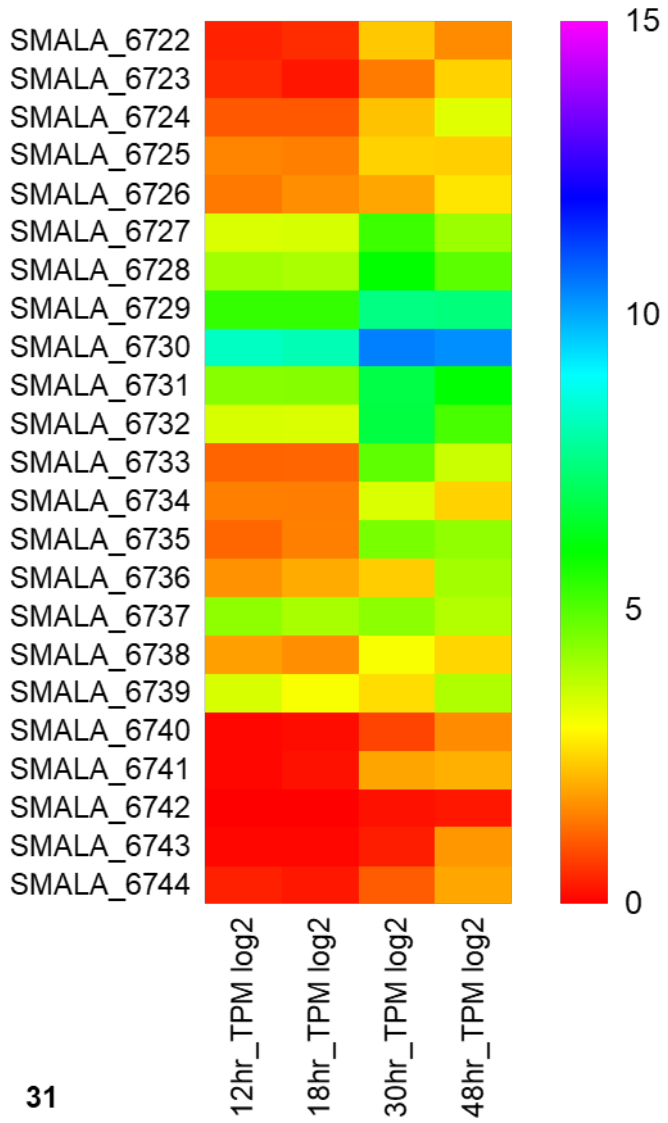

SMALA_5799
SMALA_5800
SMALA_5801
SMALA_5802
SMALA_5803
SMALA_5804
SMALA_5805
SMALA_5806
SMALA_5807
SMALA_5808
SMALA_5809
SMALA_5810
SMALA_5811
SMALA_5812
SMALA_5813
SMALA_5814
SMALA_5815
SMALA_5816
SMALA_5818
SMALA_5817
SMALA_5819
SMALA_5820
SMALA_5821
SMALA_5822
SMALA_5823
SMALA_5824
SMALA_5825
SMALA_5826
SMALA_5827
SMALA_5828
SMALA_5829
SMALA_5830
SMALA_5831
SMALA_5832
SMALA_5833
SMALA_5834
SMALA_5835
SMALA_5836
SMALA_5837
SMALA_5838
SMALA_5839
SMALA_5840
SMALA_5841
15
10
5
0
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
25

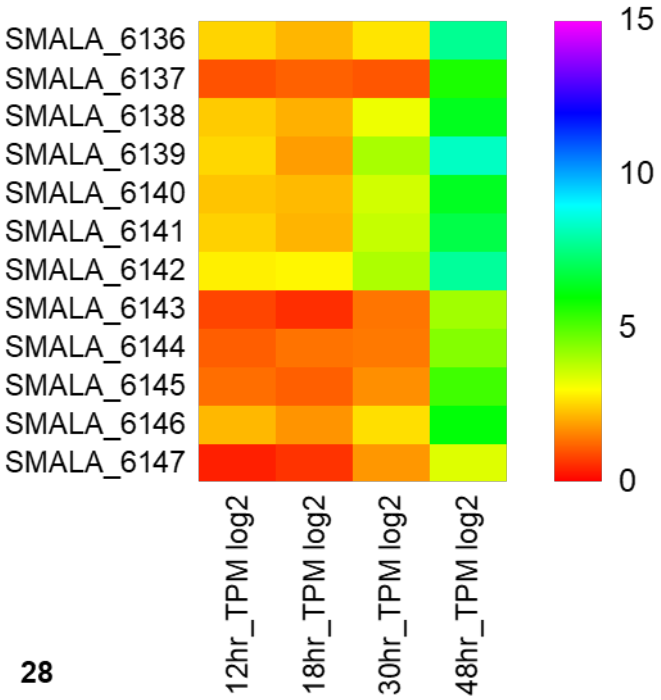

SMALA_5901
SMALA_5902
SMALA_5903
SMALA_5904
SMALA_5905
SMALA_5906
SMALA_5907
SMALA_5908
SMALA_5909
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
15
10
5
0

26

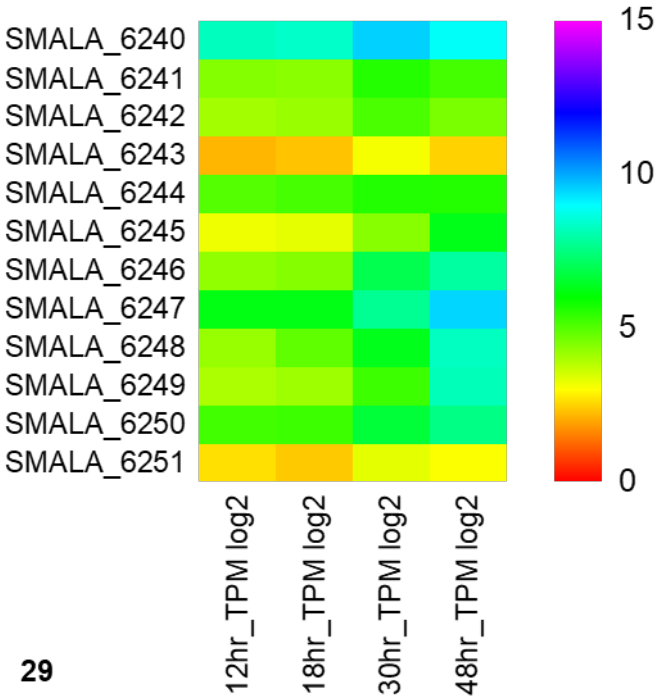

SMALA_5985
SMALA_5986
SMALA_5987
SMALA_5988
SMALA_5989
SMALA_5990
SMALA_5991
SMALA_5992
SMALA_5993
SMALA_5994
SMALA_5995
SMALA_5996
SMALA_5997
SMALA_5998
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
15
10
5
0

27

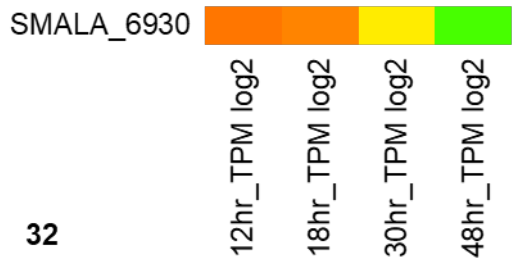

SMALA_6722
SMALA_6723
SMALA_6724
SMALA_6725
SMALA_6726
SMALA_6727
SMALA_6728
SMALA_6729
SMALA_6730
SMALA_6731
SMALA_6732
SMALA_6733
SMALA_6734
SMALA_6735
SMALA_6736
SMALA_6737
SMALA_6738
SMALA_6739
SMALA_6740
SMALA_6741
SMALA_6742
SMALA_6743
SMALA_6744
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
15
10
5
0
31


SMALA_6136
SMALA_6137
SMALA_6138
SMALA_6139
SMALA_6140
SMALA_6141
SMALA_6142
SMALA_6143
SMALA_6144
SMALA_6145
SMALA_6146
SMALA_6147
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
15
10
5
0

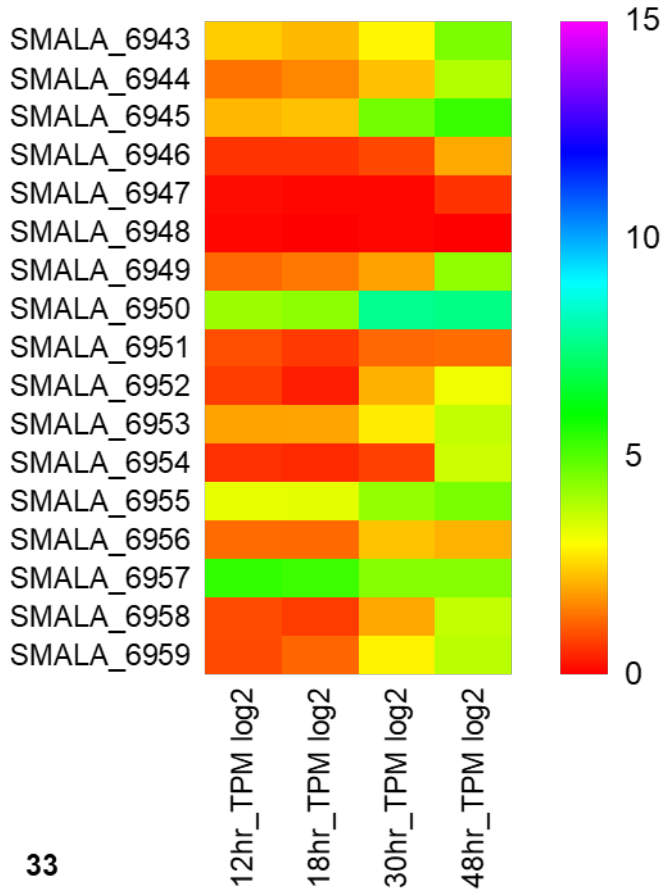
28


SMALA_6240
SMALA_6241
SMALA_6242
SMALA_6243
SMALA_6244
SMALA_6245
SMALA_6246
SMALA_6247
SMALA_6248
SMALA_6249
SMALA_6250
SMALA_6251
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
15
10
5
0

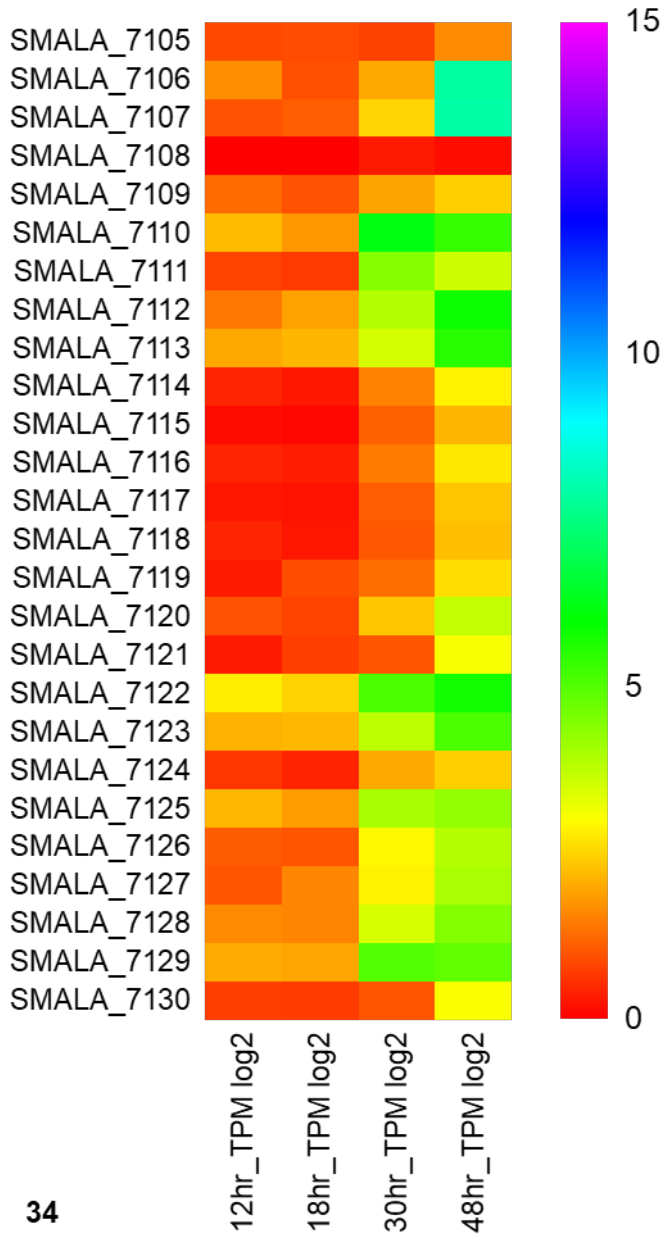
29

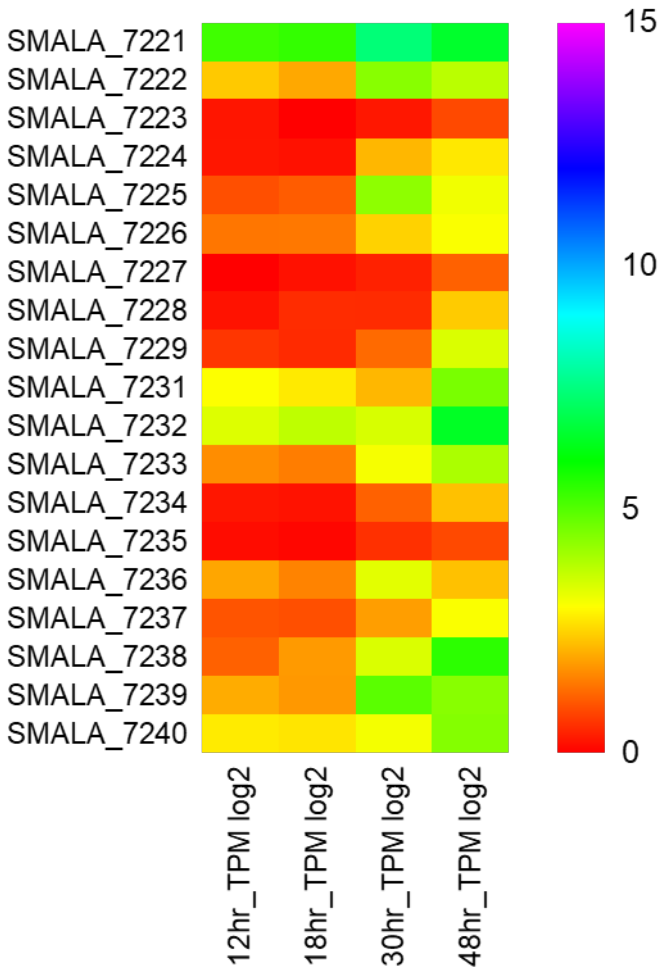

SMALA_6930
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2

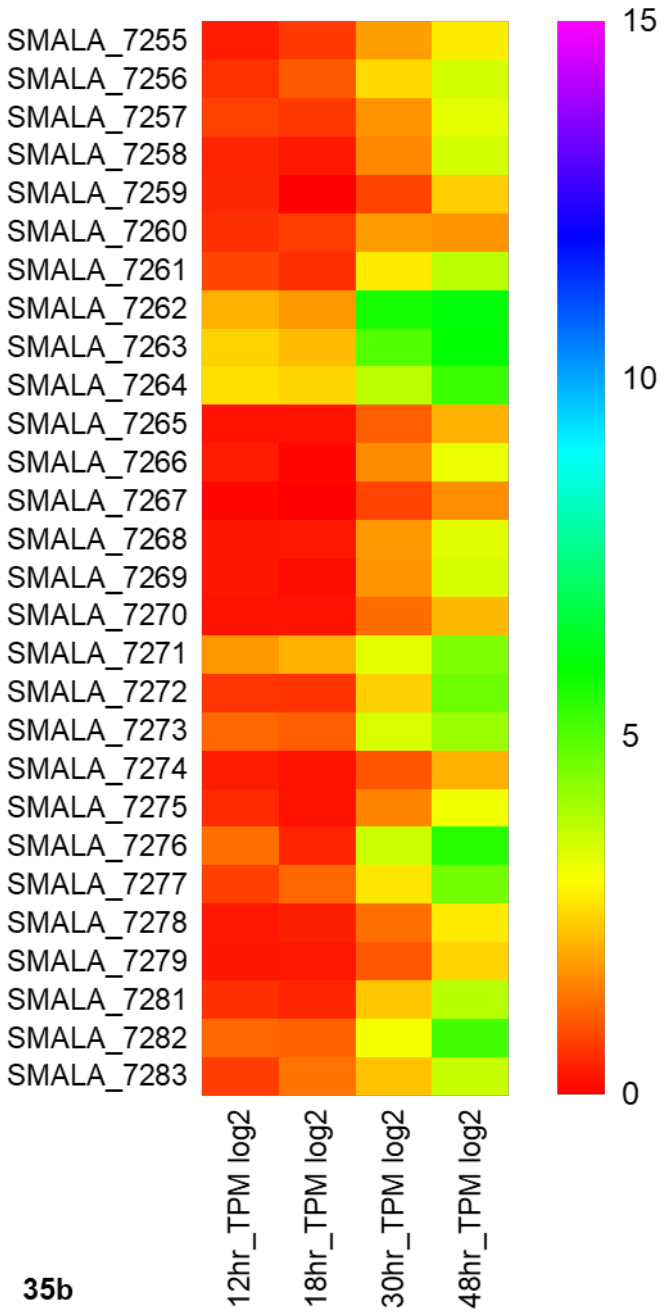
32


SMALA_6943
SMALA_6944
SMALA_6945
SMALA_6946
SMALA_6947
SMALA_6948
SMALA_6949
SMALA_6950
SMALA_6951
SMALA_6952
SMALA_6953
SMALA_6954
SMALA_6955
SMALA_6956
SMALA_6957
SMALA_6958
SMALA_6959
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
15
10
5
0

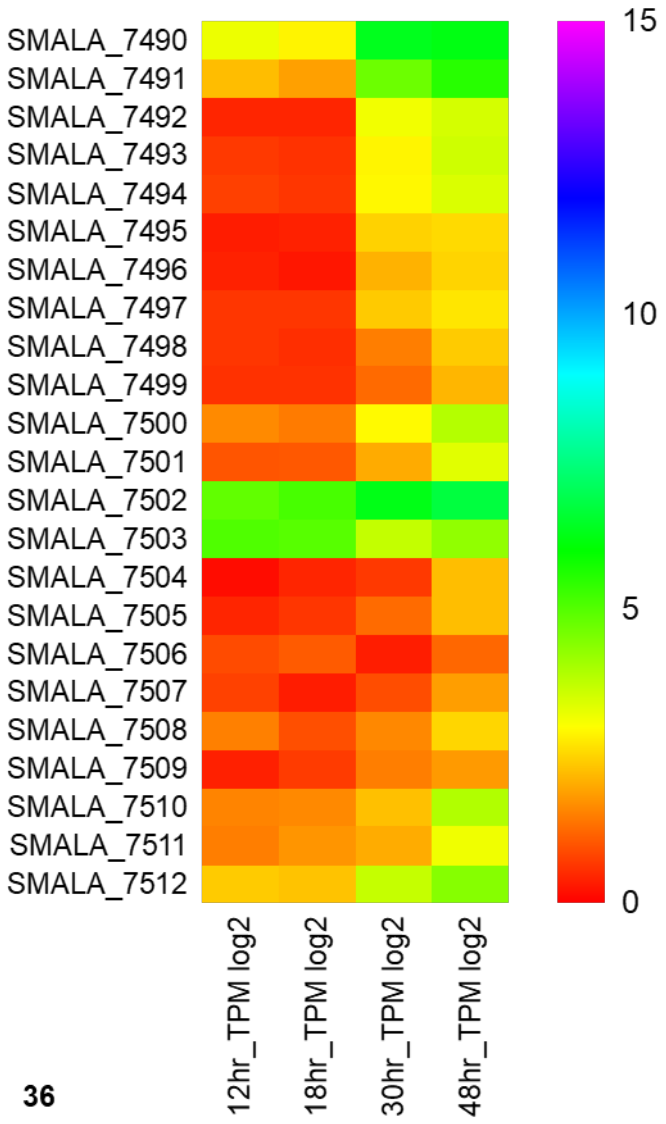
33

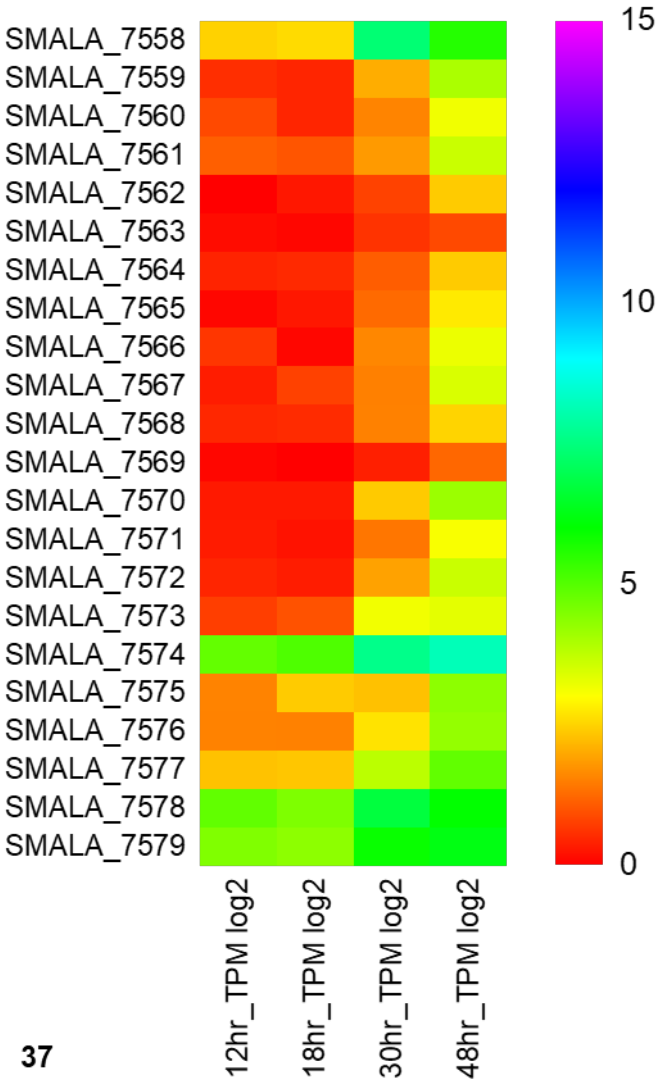

SMALA_7105
SMALA_7106
SMALA_7107
SMALA_7108
SMALA_7109
SMALA_7110
SMALA_7111
SMALA_7112
SMALA_7113
SMALA_7114
SMALA_7115
SMALA_7116
SMALA_7117
SMALA_7118
SMALA_7119
SMALA_7120
SMALA_7121
SMALA_7122
SMALA_7123
SMALA_7124
SMALA_7125
SMALA_7126
SMALA_7127
SMALA_7128
SMALA_7129
SMALA_7130
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
15
10
5
0
34

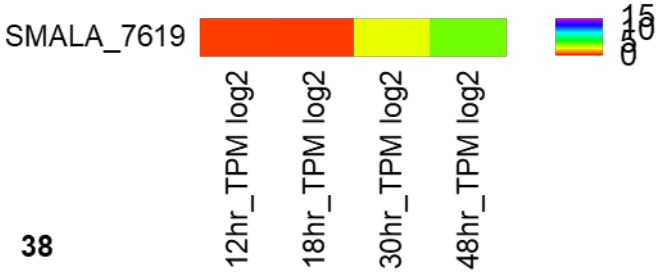

SMALA_7221
SMALA_7222
SMALA_7223
SMALA_7224
SMALA_7225
SMALA_7226
SMALA_7227
SMALA_7228
SMALA_7229
SMALA_7231
SMALA_7232
SMALA_7233
SMALA_7234
SMALA_7235
SMALA_7236
SMALA_7237
SMALA_7238
SMALA_7239
SMALA_7240
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
15
10
5
0

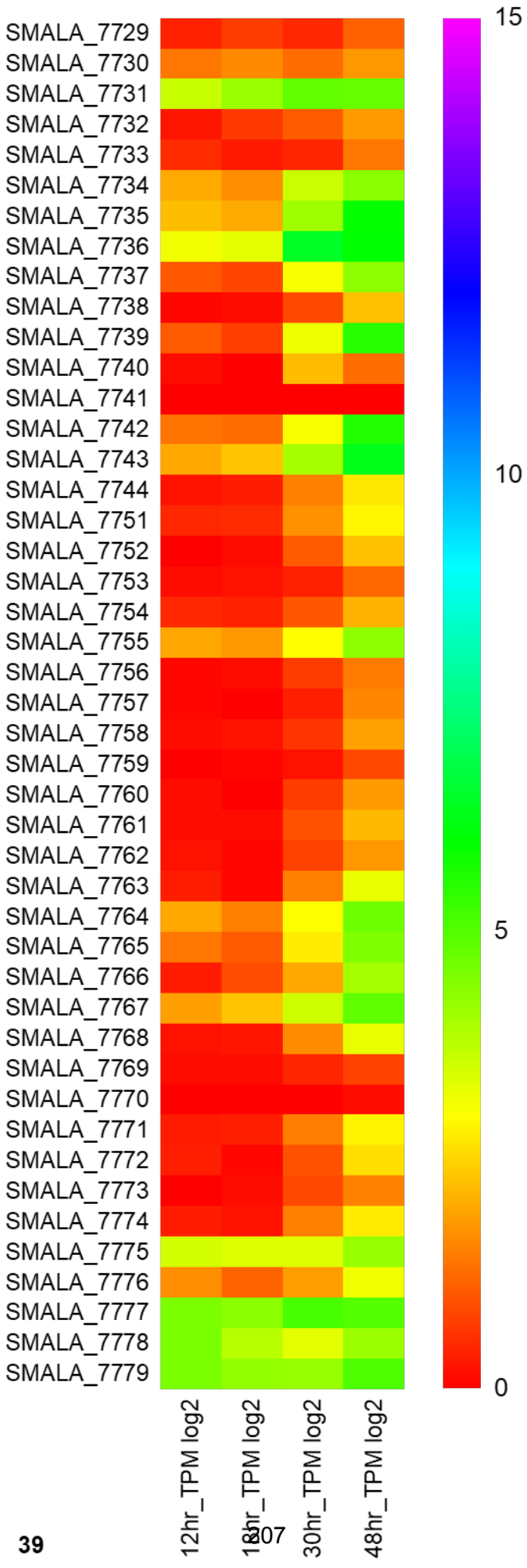

SMALA_7255
SMALA_7256
SMALA_7257
SMALA_7258
SMALA_7259
SMALA_7260
SMALA_7261
SMALA_7262
SMALA_7263
SMALA_7264
SMALA_7265
SMALA_7266
SMALA_7267
SMALA_7268
SMALA_7269
SMALA_7270
SMALA_7271
SMALA_7272
SMALA_7273
SMALA_7274
SMALA_7275
SMALA_7276
SMALA_7277
SMALA_7278
SMALA_7279
SMALA_7281
SMALA_7282
SMALA_7283
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
15
10
5
0

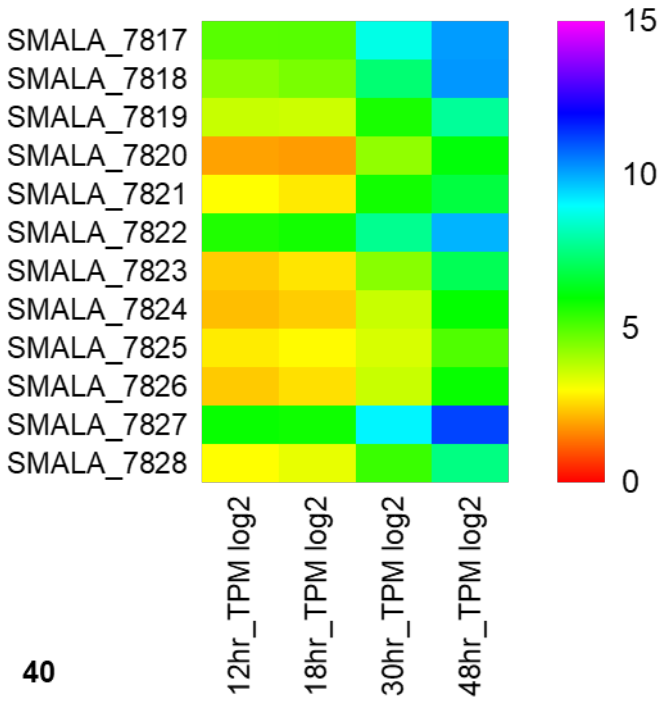
35b

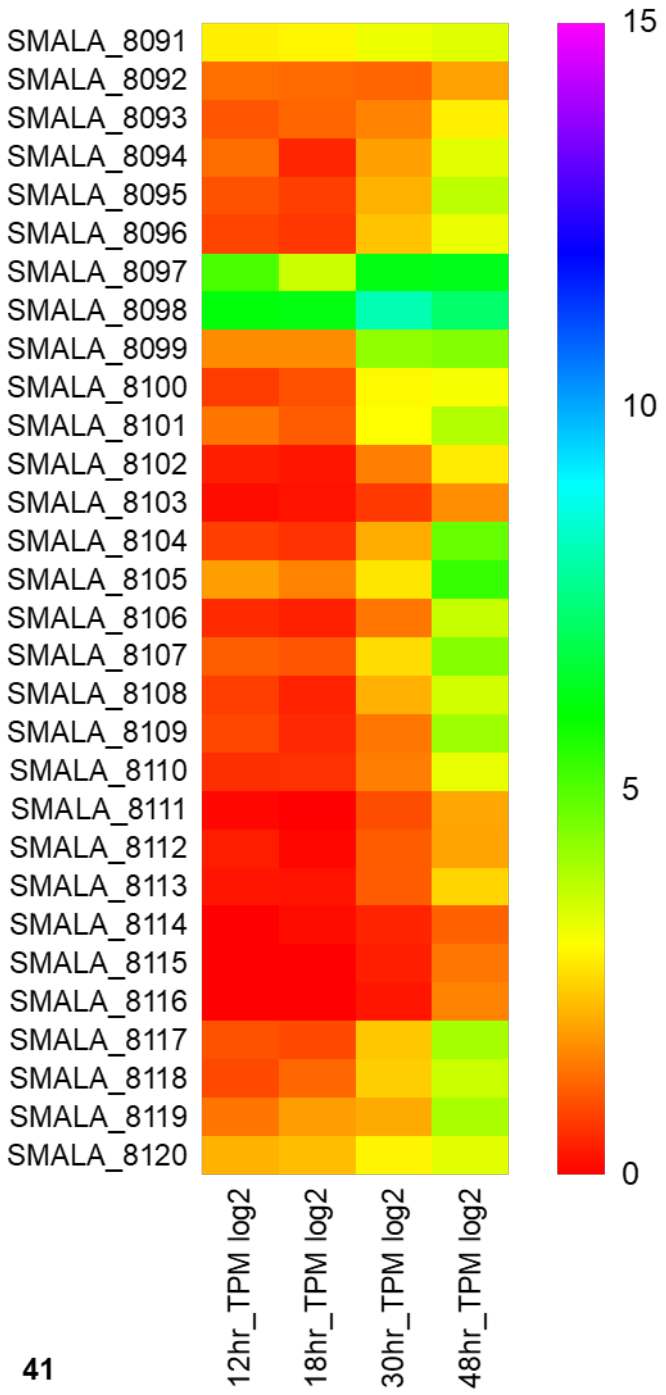

SMALA_7490
SMALA_7491
SMALA_7492
SMALA_7493
SMALA_7494
SMALA_7495
SMALA_7496
SMALA_7497
SMALA_7498
SMALA_7499
SMALA_7500
SMALA_7501
SMALA_7502
SMALA_7503
SMALA_7504
SMALA_7505
SMALA_7506
SMALA_7507
SMALA_7508
SMALA_7509
SMALA_7510
SMALA_7511
SMALA_7512
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
15
10
5
0
36

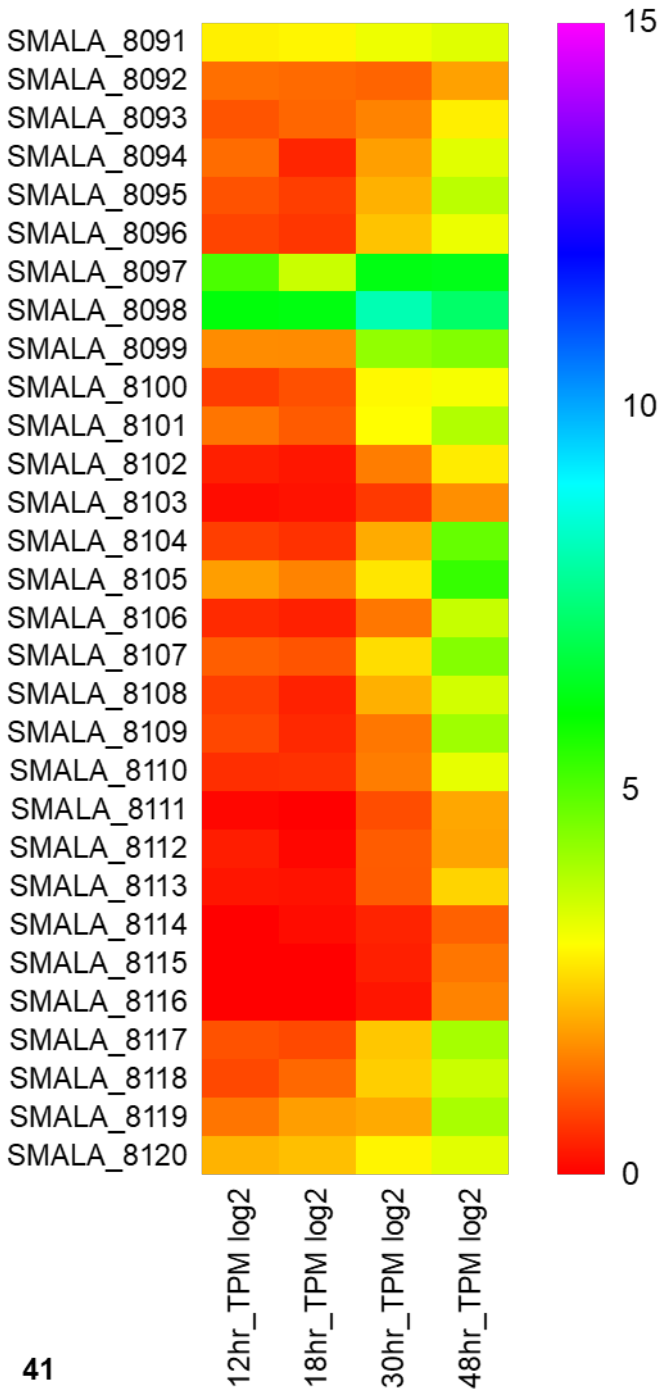

SMALA_7558
SMALA_7559
SMALA_7560
SMALA_7561
SMALA_7562
SMALA_7563
SMALA_7564
SMALA_7565
SMALA_7566
SMALA_7567
SMALA_7568
SMALA_7569
SMALA_7570
SMALA_7571
SMALA_7572
SMALA_7573
SMALA_7574
SMALA_7575
SMALA_7576
SMALA_7577
SMALA_7578
SMALA_7579
15
10
5
0
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
37

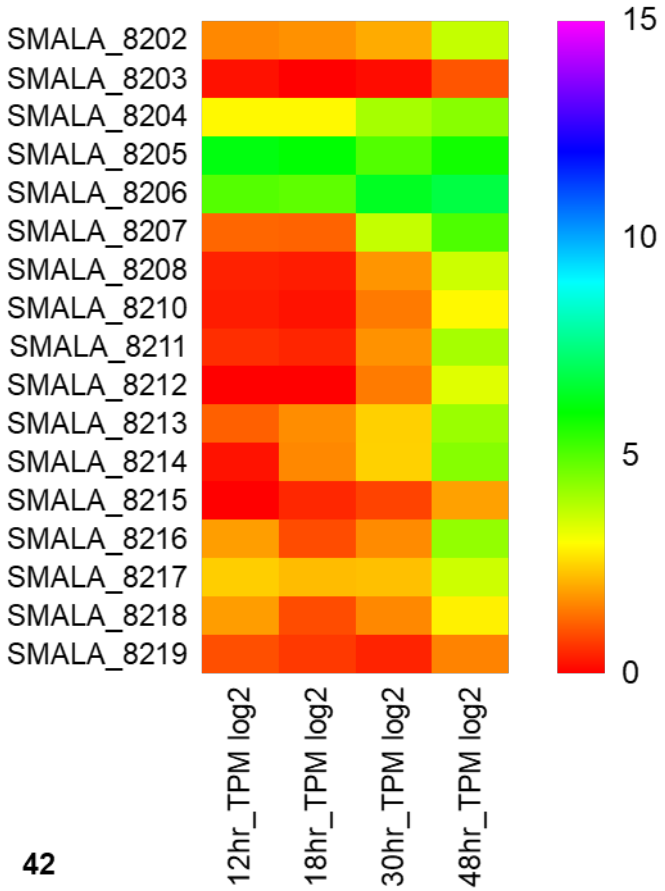

SMALA_7619
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
38


SMALA_7729
SMALA_7730
SMALA_7731
SMALA_7732
SMALA_7733
SMALA_7734
SMALA_7735
SMALA_7736
SMALA_7737
SMALA_7738
SMALA_7739
SMALA_7740
SMALA_7741
SMALA_7742
SMALA_7743
SMALA_7744
SMALA_7751
SMALA_7752
SMALA_7753
SMALA_7754
SMALA_7755
SMALA_7756
SMALA_7757
SMALA_7758
SMALA_7759
SMALA_7760
SMALA_7761
SMALA_7762
SMALA_7763
SMALA_7764
SMALA_7765
SMALA_7766
SMALA_7767
SMALA_7768
SMALA_7769
SMALA_7770
SMALA_7771
SMALA_7772
SMALA_7773
SMALA_7774
SMALA_7775
SMALA_7776
SMALA_7777
SMALA_7778
SMALA_7779
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
15
10
5
0

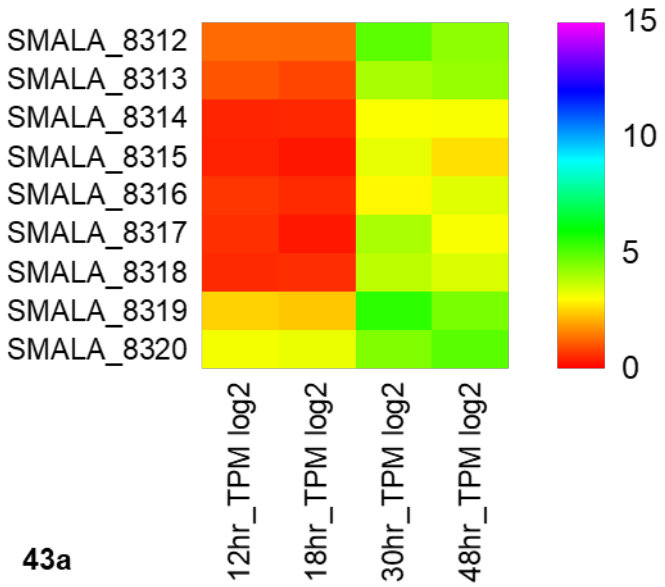

SMALA_7817
SMALA_7818
SMALA_7819
SMALA_7820
SMALA_7821
SMALA_7822
SMALA_7823
SMALA_7824
SMALA_7825
SMALA_7826
SMALA_7827
SMALA_7828
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
15
10
5
0
40


SMALA_8091
SMALA_8092
SMALA_8093
SMALA_8094
SMALA_8095
SMALA_8096
SMALA_8097
SMALA_8098
SMALA_8099
SMALA_8100
SMALA_8101
SMALA_8102
SMALA_8103
SMALA_8104
SMALA_8105
SMALA_8106
SMALA_8107
SMALA_8108
SMALA_8109
SMALA_8110
SMALA_8111
SMALA_8112
SMALA_8113
SMALA_8114
SMALA_8115
SMALA_8116
SMALA_8117
SMALA_8118
SMALA_8119
SMALA_8120
15
10
5
0
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
41

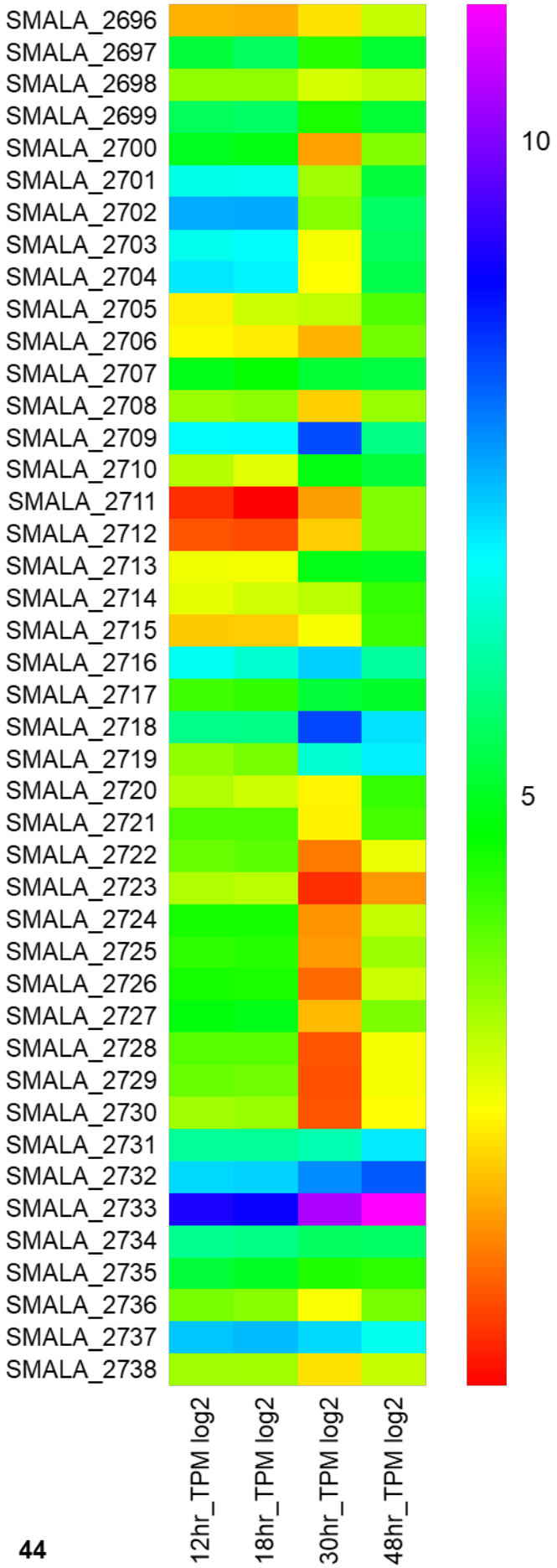

SMALA_8091
SMALA_8092
SMALA_8093
SMALA_8094
SMALA_8095
SMALA_8096
SMALA_8097
SMALA_8098
SMALA_8099
SMALA_8100
SMALA_8101
SMALA_8102
SMALA_8103
SMALA_8104
SMALA_8105
SMALA_8106
SMALA_8107
SMALA_8108
SMALA_8109
SMALA_8110
SMALA_8111
SMALA_8112
SMALA_8113
SMALA_8114
SMALA_8115
SMALA_8116
SMALA_8117
SMALA_8118
SMALA_8119
SMALA_8120
15
10
5
0
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
41


SMALA_8202
SMALA_8203
SMALA_8204
SMALA_8205
SMALA_8206
SMALA_8207
SMALA_8208
SMALA_8210
SMALA_8211
SMALA_8212
SMALA_8213
SMALA_8214
SMALA_8215
SMALA_8216
SMALA_8217
SMALA_8218
SMALA_8219
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
15
10
5
0

42


SMALA_8312
SMALA_8313
SMALA_8314
SMALA_8315
SMALA_8316
SMALA_8317
SMALA_8318
SMALA_8319
SMALA_8320
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2
15
10
5
0

43a


SMALA_2696
SMALA_2697
SMALA_2698
SMALA_2699
SMALA_2700
SMALA_2701
SMALA_2702
SMALA_2703
SMALA_2704
SMALA_2705
SMALA_2706
SMALA_2707
SMALA_2708
SMALA_2709
SMALA_2710
SMALA_2711
SMALA_2712
SMALA_2713
SMALA_2714
SMALA_2715
SMALA_2716
SMALA_2717
SMALA_2718
SMALA_2719
SMALA_2720
SMALA_2721
SMALA_2722
SMALA_2723
SMALA_2724
SMALA_2725
SMALA_2726
SMALA_2727
SMALA_2728
SMALA_2729
SMALA_2730
SMALA_2731
SMALA_2732
SMALA_2733
SMALA_2734
SMALA_2735
SMALA_2736
SMALA_2737
SMALA_2738
10
5
12hr_TPM log2
18hr_TPM log2
30hr_TPM log2
48hr_TPM log2

44

Appendix 2

Primer	Oligonucleotide sequence (5'-3')
Nigericin	
Nig_R_A_pcr	CTACTTCACCATCGGCAACTCC
Nig_R_S_pcr	TGTCCGATGTTGGACTTGATG
Nig_C_A	AAAGTCACTCAGCCGTTCTGCC
Nig_L_S	CGTTTGTGGATTGGTCGCTTAG
Nig_L_A	TAACTCATCCCGGATTTCGTCC
Nig_C_S	TGTTCACACTGGATTGGATTCCG
Nig_C_A	GGCAGAACGGCTGAGTGACTTT
Nig_C_S_2	GTTCGACGAGGTCTGTGAAC
Nig_C_A_2	AAGGTCAGGTCGTCGATAGC
Nig_R_S_2	GTACGCTGCTCAATCTGGTC
Nig_R_A_2	TCCATTGCTTTCACTTCGGC
Nig_L_S_2	AGTCGGTGTCTTCTGAGGTG
Nig_L_A_2	CGAAGACACCTCAAGCAACC
Right_Nig_R_2	GGCGTCATAGGGCCAGTAGAAG
Right_Nig_F_2	ATAAATTTGCGGCCGCTAATACGACTCACTATAGGGAGAG
Left_Nig_R_2	ACGCACCCTTCCTACCATTTATTC
40_bp_Nig_primers	TAGCGGATCCAGGAACGGG
40_bp_Nig_primers	GCTAGGATCCAGGAGCGGAA
Spinosyn	
BI-Spin-T-A	ATAGACGATCGGCTCCAGGTAG
BI-Spin-T-S	AGTACTGGAACAGCCTCTACATC
BsrGI-A	ATGTCCGGTGTGGACAGATAGAC
Primer	Oligonucleotide sequence (5'-3')

BsrGI-S	CGATGAAGCCCATCAGCGAGAT
epi_F	GAACGTCCTGAAATGACCGAG
epi_R	ATGGGAACACGTGTGAAGTGATTTTGCTCGTGTTCCGAATGCA GGCTAGCGGTG...
gdh_kre_F	...GGAACACGAGCAAAATCACTTC...
gdh_kre_R	TCCGACCAGGGCGGCAGGCCTCGGTCATTTCAGGACGTTC
gtt_F	GCAGGCGGTATGTATTCTTTTC
gtt_R	...GGGATTCTCGAACTGCTCAAC...
L1b-pSEAC13-A	GAACTCGACATCGAAATTCTGAAC
L1b-pSEAC13-S	ACTTCCGCTTCCTGTAGTGATG
L2b-pSEAC13-A	GGATCCTCTCCCTATAGTGAGTC
L2b-pSEAC13-S	AGACAGTCATAAGTGCGGCGACGATAG
R1-pSEAC13-A(F)	AGCACCGGAAGACGTGGCACAC
R1-pSEAC13-S(R)	GATCTTCATCATGCGCCAGTTC
R1b-pSEAC13-A	AGCACCGGAAGACGTGGCACAC
R1b-pSEAC13-S	GATCTTCATCATGCGCCAGTTC
R2b-pSEAC13-A	AAGGCTTGAGTCGCTCCTCCTGCCA
R2b-pSEAC13-S	GGATCCTTCTATAGTGTCACC
Rham_pSET152_F	TGCGTTCGGGGTACGCGCCGCTCATAATGGGAACACGTGT
Rham_pSET152_R	GTTACCCAACTTAATCGCCTTG
Rham_pSET152_R	GCATTGCTGGCTTGCGAAGAAAAGAATACATACCGCCTGC
S. spinosa G. DNA-A	ACACCACCACGGCGAATGGCGAC
S. spinosa G. DNA-S	TGCTCCGGGATTCTCCTGCTTG
sp_left-delete-F	AAAATCTACAACTCCCACCGGCCGCA
sp_left-delete-R	TGGTACATACTCAGGCGTCTCGGGCG
sp_right-delete-F	GACCAAGGCCGTATTGCTGCTCCACTAT
sp_right-delete-R	TTCTCGTCGGCTCGGATCACAACGTCTG
SpA-Spin-T-A	CTGACACCGATGAGAACAAGAC
SpA-Spin-T-S	ACTGTCTCAGCTGATGTGAAGAAG
SpE-Spin-T-A	AGGAATACCTGTTCCCAATTCAC
Primer	Oligonucleotide (5’-3’)

SpE-Spin-T-S	AATTGATTACGCCTCGCATTCGC
Spin_L1-pSEAC13-A	GAACTCGACATCGAAATTCTGAAC
Spin_L1-pSEAC13-S	ACTTCCGCTTCCTGTAGTGATG
Spin_L2-pSEAC13-A	GGATCCTCTCCCTATAGTGAGTC
Spin_L2-pSEAC13-A	CATCACTACAGGAAGCGGAAGT
Spin_L2-pSEAC13-S	AGACAGTCATAAGTGCGGCGACGATAG
Spin_R1-pSEAC13-A	AGCACCGGAAGACGTGGCACAC
Spin_R1-pSEAC13-S	GATCTTCATCATGCGCCAGTTC
Spin_R2-pSEAC13-A	AAGGCTTGAGTCGCTCCTCCTGCCA
Spin_R2-pSEAC13-S	GGATCCTTCTATAGTGTCACC
Spin_R2-pSEAC13-S	GTGTGCCACGTCTTCCGGTGCT
SpnA_ermE_pIB139	ATGAGCGAAGCCGGGAACCTGATAGC
SpnA_ermE_pIB139	GCACAATCGTGCCGGTTGGTAGGATCCACA
SpnA_ermE_R	ATGAGCGAAGCCGGGAACCTGATAGC
SpnA_ermE_R	CACAATCGTGCCGGTTGGTAGGATCCACAT
spnA_F	TCGCAGTGCCTCGACAACTTCTTCGTA
spnA_R	CTGGCTGAGCAACTTCTGGAATCGCTG
spnB_F	CATAGCCCTCGTACATCAATCCCGCGA
spnB_R	GGACGAAATTTCCCTGGTAGTGGCCGA
spnC_F	CAGCGTTTCGGCTCATCTCCTGCAAG
spnC_R	ACAACATCGGATTCTGCTGGACCTCGT
spnD_F_2	CACAGTTCGGCCAGCGACACCATCATG
spnD_R	TTGGAAGATGTACGCGTGTCTCCGGAG
spnD_R_2	GTAGTGGAGATGGGTTTGGCGGATTTTG

spnD-F	CGAAGAACCGATAGCCCTGATCTGCCA
spnE_F	GAAAGGACATGATTCGCCAGGCCGATG
spnE_R	GAGCTCGACAGGCTGGATACCACGTTG
spnF/S_F	CTCATTCGGTACTGCCCCTCCGATTCC
spnF/S_R	GGTCAAAACTTGCCGGCATCAGAGACG

Primer	Oligonucleotide (5'-3')
Xb-Spin-T-A	CGTTATTAACCAGCACATCGAG
Xb-Spin-T-S	GAGCCTACCGCTGATTCATAT

Albucidin

RadSamKOseq	CTACCGGGGCGACGACACGATC
RadSamKOseqrev	GGAAGATCTTGCGCGCGATCTC
Biotin_pIB139_new_F	GATTCACCGACGGCTGGGCTCAAA
Biotin_pIB139_new_R	GGCTTCCGATTCACTTGGGCTGTC
Biotin_synth_pCJW93_F	TATGACGCTCTTATCCACCGAGCTTGACG
Biotin_synth_pCJW93_R	TCATGCGGGCAGGCGGGAGAACAACC
R1_pIB139	GTTGACCGGTCATGCTGCCACCTTTCGCTCGGTC
F2_pIB139	GAGCGGCTCTTCCGGGAGCACTCGTTCATGAC
pCJW93_test_F	CCATCATCATCATCATCACAGCAGCGG
pCJW93_test_R	GATCTGGGGAATTCGGATCCAAGCTT
pIB139_Test_F	GTGCGGGGAGGATCTGAC
pIB139_Test_R	TCACTCATTAGGCACCCCAG
RadSam_pIB139_F	ATGTCCCCGGTTGAGGCCGA
RadSam_pIB139_R	TATTCGGCGACGCACTGCTCGG
RadSAM_pIB139_new_F	GGGAATCCATGTCCCCGGTTGAGG
RadSAM_pIB139_new_R	ACTTCCGAGTCCTGTCATGTCCGA
RadSam_rptR_F	ATGACGCTCTTATCCACCGAGCTTGACG

RadSam_rptR_R	TTCAAAGACAGCGGTCGCTGGTGC
RadSam_rtpr_pIB139_F	ATGTCCCCGGTTGAGGCCGA
RadSam_rtpr_pIB139_R	ATGTCCGACCGCTATTCGGCGACGCA
RadSam2_pIB139_F	ATGACGCTCTTATCCACCGAGCTTGACG

Primer	Oligonucleotide sequence (5'-3')
RadSam2_pIB139_R	CTATTCGGCGACGCACTGCTCGG
RadSamKOseqrev	GGAAGATCTTGCGCGCGATCTC
rtpR_pCJW93_F	ATGACAGGACTCGGAAGTAGTGGACCGACGTTCCGGCTCG
rtpR_pCJW93_R	TTCAAAGACAGCGGTCGCTGGTGCAG
rtpR_RadSam_pIB139_F	CGCCGAATAGCGGTCGGACATGACA
rtpR_RadSam_pIB139_R	TTCAAAGACAGCGGTCGCTGGTGC
F1	ATGACGCTCTTATCCACCGAGCTTGACG
F2	TTCAAAGACAGCGGTCGCTGGTGC
R1	ATGTCCCCGGTTGAGGCCGA
R2	ATGTCCGACCGCTATTCGGCGACGCA
R3	ATGACGCTCTTATCCACCGAGCTTGACG