

University of Strathclyde

Strathclyde Institute of Pharmacy and Biomedical Sciences

Determination of the genetic basis of oxytetracycline productivity for *Streptomyces rimosus*, by the examination of genomic data.

Alison C MacFadyen

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

2015

Declaration

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Signed:

Date:

Acknowledgements

I'd like to thank my supervisors, Dr Paul Herron, Prof Iain Hunter and Dr RuAngelie Edrada-Ebel, for their help and support throughout my PhD. I'd particularly like to thank Paul for everything he has done to help me complete this thesis. I'd also like to thank Dr Paul Hoskisson and Dr Nicholas Tucker for their advice, support and encouragement. I appreciate it immensely. My PhD would not have been the same without the wonderful people within the microbiology group. When things didn't quite go right it was good to have such a supportive team around, so thank you to all. In particular I'd like to thank Kirsty, Danny, John, Charles, Leena, Jana, Lewis, Andy, Maeidh, Laura and our honorary member Walid, for cheering me up, giving me great advice or just listening to me while I rant. I'd also like to thank Craig, for being a brilliant desk neighbour.

For tea and cake I need to thank Lisa and Cindy. Chatting over tea and eating cake is the best way to forget about your worries and I can think of no better people to share tea with. To Harry, life as an undergraduate, let alone a postgraduate, would not have been the same without you around. For that, I will always be grateful. Where would a PhD student be without the pub! The Cartographers thank you! Thank you for making my week ever cheerful and fun. Thank you for letting me shine with Disney and for not asking "How's the writing going?".

Finally I want to thank my family. To Allan and Julie, both of you have always been there to help me and that is something I will never forget. To my wonderful parents, Neil and Carole. You have always said you would be happy no matter what I did, you have always encouraged me, supported me and believed in me. I love you both so very much and thank you eternally for everything you have done for me. To my brother Graham. You have entertained me when I have spoken of my PhD work, been keen to learn and inquisitive. That has given me the greatest boost of all. You mean the world to me, can always make me smile and I love you dearly. I dedicate this thesis to you…"I've just been cutting onions".

Abstract

Streptomyces rimosus, the industrial strain used in the production of the Type-II polyketide antibiotic oxytetracycline (OTC), has undergone extensive strain improvement over the past 50 years. This has resulted in OTC levels increasing from less than 0.5 g per litre in the original soil isolate to over 70 g per litre for contemporary production strains. By analysing the genome sequences of four strains that lie within this lineage, each with a different level of OTC productivity, we have investigated the genetic basis of increased OTC yield. Using comparative genomics, a variety of significant chromosomal rearrangements and deletions in the lineage have been identified. These chromosomal rearrangements have resulted in a loss of genes involved in primary metabolism, the synthesis of other secondary metabolites, transcriptional regulators and sigma factors. Most notable, is the loss of the rimocidin biosynthetic cluster in later strains, which suggests that increased precursor supply is an important determinant for increased OTC yield. To confirm this association and to elucidate the significance of the observed genomic rearrangements, gene knockouts were produced in the original soil isolate. A transcriptional regulator putatively associated with the OTC biosynthetic gene cluster was also identified during this analysis. Further to this, the regulator in the one of the production strains contained a mutation within the DNA-binding region of the protein. It is possible this mutation alters DNA-binding affinity and therefore improves OTC production. Ultimately, we aim to use our genomic data to inform rapid improvement of other strains.

Publications

Pethick, F.E., MacFadyen, A.C., Tang, Z., Sangal, V., Liu, T.T., Chu, J., Kosec, G., Petković, H., Guo, M., Kirby, R., Hoskisson, P.A., Herron, P.R., Hunter, I.S. (2013) Draft Genome Sequence of the Oxytetracycline-Producing Bacterium Streptomyces rimosus ATCC 10970. Genome Announcements. 1(2), e00063–13.

General Abbreviations

ACP	Acyl-carrier protein
ACT	Actinorhodin
ADP	Adenosine diphosphate
AT	Acyltransferase
ATP	Adenosine triphosphate
bp	Base pairs
BTAD	Bacterial Transcriptional Activator Domain
CoL	Carboxylic acid-CoA ligase
DH	Dehydratase
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic
ER	Enoylreductase
EMSA	Electrophoretic mobility shift assay
HIS-tag	Histidine tag
HRMS	High resolution mass spectrometry
HTH	Helix-turn-helix
InDEL	Insertion and/or deletion
sInDEL	Short insertion and/or deletion
IPTG	Isopropyl β-D-1-thiogalactopyranoside
KR	Ketoreductase
KS	Ketosynthase
NADH	Nicotinamide adenine dinucleotide, reduced form
NADPH	Nicotinamide adenine dinucleotide phosphate, reduced form
NRPS	Non-ribosomal peptide synthase
NTP	Nucleotide triphosphate
OTC	Oxytetracycline
PAGE	Polyacrylamide gel electrophoresis
PAS	Per, ARNT and Sim domain
PCR	Polymerase chain reaction

PKS	Polyketide synthase
ppGpp	Guanosine tetraphosphate
PPP	Pentose Phosphate Pathway
pppGpp	Guanosine pentaphosphate
RED	Undecylprodigiosin
RNA	Ribonucleic acid
RR	Response Regulator
rRNA	Ribosomal RNA
SARP	*Streptomyces* antibiotic regulatory proteins
SDS	Sodium dodecyl sulphate
SK	Sensor Kinase
T1PKS	Type I polyketide synthase
T2PKS	Type II polyketide synthase
TCS	Two Component System
TEMED	N,N,N',N'-Tetramethylethylenediamine
Trans-Reg	Transcriptional Regulatory
Tris	Tris(hydroxymethyl)aminomethane
tRNA	Transfer RNA
USR	Upstream Region

Contents

Chapter 1 – Introduction

1	Introduction	1
1.1	Polyketides	2
1.1.1	Precursor supply for polyketides	3
1.1.2	Type I polyketide synthases	7
1.1.3	Type III polyketide synthases	10
1.1.4	Type II polyketide synthases	12
1.2	Oxytetracycline	17
1.2.1	OTC biosynthesis	17
1.2.2	Regulation of the OTC biosynthetic cluster	25
1.3	Antibiotic pathway specific regulators	26
1.3.1	*Streptomyces* antibiotic regulatory proteins	26
1.3.2	The GntR family of regulators	27
1.3.3	The TetR family of regulators	28
1.3.4	The LuxR family of regulators	30
1.4	Global regulation of antibiotic production	32
1.4.1	The role of guanosine tetraphosphate and pentaphosphate	32
1.4.2	The role of γ-butyrolactone	33
1.4.3	The role of carbon	34
1.4.4	The role of nitrogen	35
1.4.5	The role of phosphate	36
1.4.6	*Streptomyces* development in relation to antibiotic production	37
1.5	Aims of the project	41

2	Material and Methods	45
2.1	Bacterial Strains and Vectors	45
2.2	Suppliers	48
2.3	Media and Solutions	48
2.3.1	Media	48
2.3.1.1	Media for *Streptomyces rimosus*	48
2.3.1.2	Media for *Escherichia coli*	50
2.3.1.3	Miscellaneous Media	50
2.3.2	Antibiotics	51
2.3.3	Solutions	51
2.3.3.1	Solutions associated with DNA isolation and manipulation	51
2.3.3.2	Buffers for use with polyacrylamide gels	52
2.3.3.3	Buffers for OtcR purification	54
2.3.3.4	Buffers used for *S. rimosus* crude lysate and biotin interaction experiments	55
2.4	Genomic and plasmid DNA purification	55
2.4.1	Genomic DNA isolation from *S. rimosus*	55
2.4.2	Plasmid DNA isolation from *E. coli*	56
2.5	Methods for the manipulation of DNA	56
2.5.1	Restriction endonuclease digestion	56
2.5.2	Polymerase chain reaction (PCR)	56
2.5.3	DNA clean up	57
2.5.4	Agarose gel electrophoresis	57
2.5.5	DNA ligations by use of T4 DNA ligase	58
2.5.6	Sequencing of *S. rimosus* SR1 genomic DNA	58
2.6	Primers used in this work	59
2.7	Plasmid Construction	60
2.7.1	pGB23	60
2.7.2	pOtcR-SR1, pOtcR-SR55 and pOtcR21-SR1	60
2.7.3	pOtcR-KO	61

2.7.4	Plasmids for generating EMSA DNA targets	61
2.8	Microbiological methods	62
2.8.1	*S. rimosus* associated microbial methods	62
2.8.1.1	Preparation of electrocompetent *S. rimosus* mycelium	62
2.8.1.2	Transformation of electrocompetent *S. rimosus* mycelium	62
2.8.1.3	Introduction of DNA into *S. rimosus* spores by intergeneric conjugation	63
2.8.2	*E. coli* associated microbial methods	63
2.8.2.1	Preparation of chemically competent *E. coli* cells	63
2.8.2.2	Heat shock transformation of chemically competent *E. coli*	64
2.8.2.3	Preparation of electrocompetent *E. coli* cells	64
2.8.2.4	Transformation of electrocompetent *E. coli* cells	65
2.8.3	Extraction of OTC and rimocidin	65
2.8.4	Determination of antimicrobial activity of extracts	65
2.9	Protein methods	66
2.9.1	Production of OtcR	66
2.9.2	SDS Polyacrlyamide gel electrophoresis (SDS-PAGE)	66
2.9.3	Small-scale preparations of OtcR for solubility tests	67
2.9.4	Purification of OtcR	67
2.10	Preparation of crude *S. rimosus* cell lysate	68
2.11	DNA sampling of the upstream region of *otcR*	68
2.12	EMSA protocol for pure and crude protein samples	69
2.13	Mass spectrometry data processing	70
2.14	Computational Methods	71
2.14.1	GET_HOMOLOGUES	71
2.14.2	CGView Comparison Tool	71
2.13.3	Burrows-Wheeler Alignment (BWA) and Sequence Alignment/Map (SAMtools)	72

Chapter 3 – Results – Comparative genomics

3	Comparative genomics of *S. rimosus* SR1 and 19 complete *Streptomyces* genomes	73
3.1	Introduction	73
3.2	Genome sequencing of *S. rimosus* SR1	76
3.3	The core, accessory and unique genome of *Streptomyces*	77
3.3.1	Core and Pan-genome Analysis	77
3.3.2	*Streptomyces* core genome	82
3.3.3	*S. rimosus* SR1 accessory genome	86
3.3.4	*S. rimosus* SR1 unique genes	88
3.4	Core and Pan phylogenomic analysis of *Streptomyces*	92
3.5	Additional features of the accessory genome of *Streptomyces*	100
3.5.1	Examination of the secondary metabolites associated with each Streptomycete	100
3.5.2	*S. rimosus* SR1 has 4 sets of respiratory nitrate reductase genes	104
3.6	Discussion	110

Chapter 4 – Results – Genomic differences between *S. rimosus* strains

4	The genomic differences between *Streptomyces rimosus* SR1 and the production strains	120
4.1	Introduction	120
4.2	Generation of comparison files for the identification of SNCs and DIPs	122
4.3	Overview of genomic differences between SR1 and all strains	122
4.3.1	Deletion profiles of each production strain	122
4.3.2	Overview of single nucleotide polymorphisms for each	128

production strain
4.4 Genomic comparisons between *S. rimosus* SR1 and *S. rimosus* SR20 131
4.4.1 SR20 has lost genes associated with two secondary metabolite biosynthetic clusters 131
4.4.2 SR20 has non-synonymous SNCs within genes associated with (p)ppGpp 137
4.5 Genomic comparisons between *S. rimosus* SR1 and *S. rimosus* SR50S 148
4.5.1 SR50S has lost a gene putatively linked with CutRS 148
4.5.2 SR50S contains a DIP within a homologue of *sigQ* 154
4.6 Genomic comparisons between *S. rimosus* SR1 and *S. rimosus* SR55 161
4.6.1 SR55 has lost genes associated with nitrogen metabolism 161
4.6.2 The presence of non-synonymous SNC within a potential regulator of the OTC biosynthetic gene cluster 165
4.7 Discussion 178
4.7.1 Genetic differences between SR1 and SR20 179
4.7.2 Genetic differences between SR1 and SR50/50S 182
4.7.3 Genetic differences between SR1 and SR55 183

Chapter 5 – Results – Characterisation of rimocidin biosynthetic cluster

5 Characterization of the rimocidin biosynthetic gene cluster from *S. rimosus* 185
5.1 Introduction 185
5.2 Elucidation of the rimocidin biosynthetic cluster 188
5.3 Confirmation of rim deletion in SR20, SR50S and SR55 199
5.4 Generation of a rimocidin non-producing strain of *S. rimosus* SR1 201
5.5 Discussion 208

Chapter 6 – Results – Mass spectrometric analysis

6	Mass spectrometric analysis of rimocidin and related metabolites produced by *S. rimosus*	211
6.1	Introduction	211
6.2	Preparation of samples and HRMS data for analysis	214
6.3	Detection of oxytetracycline in *S. rimosus* extracts by HRMS	214
6.4	Identification of rimocidin and CE-108 by HRMS	217
6.5	Identification of derivative 23 by HRMS	222
6.6	Discussion	224

Chapter 7 – Results – Purification of OtcR

7	Purification and initial characterisation of the putative SARP, OtcR	226
7.1	Introduction	226
7.2	Deletion of *otcR* from *S. rimosus* SR1 and SR55	230
7.3	Solubility optimisation and purification of OtcR via nickel affinity chromatography	233
7.4	Investigation of the DNA-binding properties of OtcR	242
7.5	Discussion	251

Chapter 8 – Discussion

8	General discussion and future work	260
8.1	General Discussion	266

8.2 Future Work 268

Chapter 9 – References

9 References 269

11 Appendix

11 Appendix – Publications 306

Index to Figures and Tables

Chapter 1 – Introduction

Figure 1.1	Primary metabolic pathways associated with secondary metabolism precursor supply	6
Figure 1.2	Biosynthetic pathway of 6-deoxyerythronolide-B	9
Figure 1.3	Chalcone formation by the chalcone synthase (CHS)	11
Figure 1.4	Representation of the KSα, KSβ, and ACP minimal PKS complex.	13
Figure 1.5	The biosynthetic pathway for actinorhodin	16
Figure 1.6	Organisation of the OTC biosynthetic gene cluster	19
Table 1.1	OTC biosynthetic genes and predicted function	20
Figure 1.7	Predicted pathway for oxytetracycline biosynthesis	21
Figure 1.8	Conversion of 6-MPT to 4-keto-6-MPT	23
Figure 1.9	*Streptomyces* developmental life cycle	39
Figure 1.10	Schematic for the origin and parentage of *S. rimosus* strains	44

Chapter 2 – Material and Methods

Table 2.1	*S. rimosus* strains used in this work	45
Table 2.2	*E. coli* strains used in this work	46
Table 2.3	Plasmids used in this work	47
Table 2.4	Antibiotic concentrations used for the growth and maintenance of *S. rimosus* and *E. coli*	51
Table 2.5	General PCR conditions used	57
Table 2.6	Primers used in this work	59

Chapter 3 – Results – Comparative genomics

Table 3.1	Candidate streptomycetes subjected to comparative genomic analysis	75
Figure 3.1	CCT map of *S. rimosus* SR1 compared to 19 complete *Streptomyces* genomes	79
Figure 3.2	Random sampling plots of the core and pan-genome	81
Table 3.2	Functional SEED categories of the core genome	84
Table 3.3	Core regulators of *Streptomyces* genomes	85
Table 3.4	Functional SEED categories of the *S. rimosus* SR1 accessory genome	87
Table 3.5	Functional SEED categories of the *S. rimosus* SR1 unique genes	90
Figure 3.6	CCT map location of phages RP2 and RP3	91
Figure 3.7	Phylogeny of the *Streptomyces* species	98
Figure 3.8	Putative whole proteome similarity	99
Figure 3.9	The location of *S. rimosus* SR1 biosynthetic clusters	102
Figure 3.10	Overview of biosynthetic clusters for *Streptomyces* species	103
Figure 3.11	Phylogenetic Trees of Nar Proteins	109

Chapter 4 – Results – Genomic differences between *S. rimosus* strains

Figure 4.1	CCT Map of nBLAST of SR20, SR50S and SR55 versus SR1	124
Figure 4.2	Overview of gene loss for each deletion profile	125
Table 4.1	Total number of genes within each deletion profile and as a percentage of each SEED subsystem	127
Figure 4.3	Categorisation of the single nucleotide polymorphisms for strains SR20, SR50S and SR55	129
Table 4.2	Total number of non-synonymous SNCs and as a	130

	percentage of each SEED subsystem for SR20, SR50S and SR55	
Figure 4.4	Zoomed CCT nBLAST map of the 277 kbp deletion region of SR20	133
Table 4.3	Genes within the SR20 deletion contributing to the main SEED subsystems	134
Table 4.4	CDS in which frame-shift, stop or non-stop condon changes were predicted as well as SNCs contributing to the main SEED subsystems	140
Table 4.5	Top pBLAST hits for proteins with frame-shift, stop or non-stop condon changes within the main SEED subsystem	143
Figure 4.5	CLC amino acid alignment of the N-terminal catalytic domain of RelA from various Gram positive and Gram negative organisms	146
Figure 4.6	Chromosomal synteny of *eshA* (*SRIM7514*) between selected *Streptomyces* species	147
Figure 4.7	SR20 and SR50S sequences mapped to SR1 at the 69 kbp deletion point	150
Figure 4.8	Zoomed CCT nBLAST map of the 142 kbp deletion region of SR50, before the start of the 69 kbp deletion region shared with SR20	151
Table 4.6	Genes within the SR50 deletion profile contributing to the main SEED subsystems	152
Table 4.7	CDS in which frame-shift, stop or non-stop condon changes were predicted as well as SNCs contributing to the main SEED subsystems	156
Table 4.8	Top pBLAST hits for proteins with frame-shift, stop or non-stop condon changes within the main SEED subsystem	158
Table 4.9	Percentage of transition/transversion mutations of all SNCs for SR20, SR50S and SR55	160

Figure 4.9	Zoomed CCT nBLAST map of the 41 kbp deletion region of SR55	163
Table 4.10	Genes within the SR55 deletion profile contributing to the main SEED subsystems	164
Table 4.11	CDS in which frame-shift, stop or non-stop condon changes were predicted as well as SNCs contributing to the main SEED subsystems	168
Table 4.12	Top pBLAST hits for proteins with frame-shift, stop or non-stop condon changes within the main SEED subsystem	172
Figure 4.10	*S. rimosus* SR1 chromosome as a circular representation with SR20, SR50S and SR55 sequence mapped	181

Chapter 5 – Results – Characterisation of rimocidin biosynthetic cluster

Figure 5.1	Schematic for the generation of double and single crossovers	187
Figure 5.2	The Rimocidin biosynthetic cluster	189
Figure 5.3	Model of rimocidin biosynthesis	191
Figure 5.4	Catalytic domains of Module 12	193
Figure 5.5	Domain architecture of RimRI, RimRII, RimRIII and RimRIV	195
Figure 5.6	Alignment of RimRIII with other regulators of polyene biosynthesis	196
Figure 5.7	Alignment of RimRIV with other regulators of polyene biosynthesis	198
Figure 5.8	Confirmation of *rim* deletion in SR20, SR50S and SR55	200
Figure 5.9	Confirmation of pGB23 by restriction digest	203
Figure 5.10	PCR confirmation of *rim* knockout in SR1 and	205

generation of strain ACM23
Figure 5.12 Loss of rimocidin production and analysis of OTC production by *S. rimosus* ACM23 207

Chapter 6 – Results – Mass spectrometric analysis

Figure 6.1 Module 12 modification of Rimocidin and CE-108 212
Figure 6.2 Molecular structure of oxytetracycline, and corresponding chromatogram 216
Figure 6.3 Molecular structures of rimocidin and CE-108, and corresponding chromatograms 219
Figure 6.4 Molecular structures of rimocidin* and CE-108*, and corresponding chromatograms 221
Figure 6.5 Molecular structure of derivative 23, and corresponding chromatogram 223

Chapter 7 – Results – Purification of OtcR

Figure 7.1 Bioinformatic analysis of OtcR 229
Figure 7.2 Isolation of cosmid pGLW207 231
Figure 7.3 SDS-PAGE of OtcR overexpression at 30 °C 236
Figure 7.4 SDS-PAGE gel of OtcR solubility tests 238
Figure 7.5 Binding specificity of OtcR using nickel affinity chromatography 240
Figure 7.6 Nickel affinity chromatography of OtcR 241
Figure 7.7 Analysis of tandem repeats in the USR of *otc* biosynthetic cluster genes 244
Figure 7.8 DNA binding assay with purified OtcR-SR1 245
Figure 7.9 Selected EMSA from crude *S. rimosus* lysate 247
Figure 7.10 Whole cell extract protein profiles of SR1, SR55 and 250

SR50S at 18 and 46 hour growth points
Figure 7.11 Overview of four potential DNA binding motifs in the USR of *otcR* 258
Figure 7.12 Putative model for the regulation of *otcR* 259

Chapter 8 – Discussion

Figure 8.1 Confirmation of *rim* cluster presence and lack of rimocidin biosynthesis by R6-500 264

Chapter 1

1 Introduction

The discovery of antibiotics allowed the effective treatment of bacterial infections that resulted in high mortality and morbidity (Alanis 2005). An example of this is the significant difference to the morality rate of *Neisseria meningitidis* infections. The mortality rate for acute meningococcal meningitis was 70 – 90 % before the use of the antibiotic, sulphanilamide, after which the mortality rate dropped to 10 % (Powers 2004). It is easy to see why antibiotics represent such a significant achievement of modern medicine; even so pharmaceutical companies are reluctant to invest in the development of new antibiotics. Although the antibiotic market is around $66 billion a year (Sánchez *et al.,* 2010); many other therapeutic classes of drugs are more profitable than antibiotics, due to the limited life of a new antibiotic, which in the past has discouraged investment (Powers 2004). However with the increase of antibiotic resistance in recent times, the need for new antibiotics has become ever greater, with more time and energy being invested in the discovery of new sources of antibiotics (Yin *et al.,* 2007; Challis 2008; Corre & Challis 2009; Berdy 2012). Despite this, the majority of cases of infectious disease that require treatment with a new class of antibiotic are rare (Höjgård 2012; So & Shah 2014). Newly-developed antibiotics will consequently be reserved for treatment of these most severe, life-threatening cases, whereas the use of older antibiotics will be for more common, non-life-threatening infections. This represents another reason for the reluctance of pharmaceutical companies to invest in antibiotic discovery, as the potential for profit is limited (So & Shah 2014). In addition, the value of long standing antibiotics is still high, as most non-life-threatening infections are treated with these antibiotics, for example 3 million of the 4 million yearly cases of community-acquired pneumonia in the USA are treated as outpatients, suggesting that the infection is not severe and older antibiotics

can be used to treat this disease (Powers 2004). This, combined with the expense of producing a new antibiotic and bringing it to the market, gives sustenance to the idea that developing new drugs is unneeded and non-profitable. Therefore, not only should an emphasis be placed on developing new antibiotics, but also increasing their efficiency of production and consequently the profitability of both old and new antibiotics. In this way, pharmaceutical companies are more likely to recognize the benefit of investing in a new drug as it could be produced cheaply and in large quantities. In order to explore increased antibiotic production and efficiency the process involved in antibiotic synthesis must be understood. As the productivity of the polyketide antibiotic oxytetracycline is the focus of this thesis, polyketide biosynthesis will be discussed.

1.1 Polyketides

Polyketides are secondary metabolites produced by plants, fungi and bacteria (Jez *et al.,* 2001; Schneider 2005). Many medically important natural products such as antibacterial agents (erythromycin), antifungal agents (nystatin), anticancer drugs (doxorubicidin) and antiparasitic drugs (frenolicin) are polyketides, and are synthesised by polyketide synthases (PKSs) (Rawlings 2001a; Rawlings 2001b; Hertweck *et al.,* 2007). There are three different classes of PKS and they are categorised in relation to the method by which they synthesise their respective polyketide molecule (Robinson 1991; Watanabe *et al.,* 2007; Hertweck *et al.,* 2007). These categories are Type I, which produce modular polyketides, Type II, involved in aromatic polyketide synthesis, and Type III, consisting of single catalytic PKS protein (Moore & Hertweck 2002; Watanabe *et al.,* 2007; Aizawa *et al.,* 2014). Due to similarity in mechanism of action and molecular arrangement of Type I and Type II fatty acid synthases (FAS), Type I and Type II PKS have a similar naming convention; although polyketide synthesis is considered more complex (Robinson 1991; Chan & Vogel 2010). Each of these PKS are functionally similar, but are structurally different. Type I PKSs

consist of multi-module proteins, with each domain within the module performing a particular enzymatic function, Type II PKS have distinct proteins to perform each reaction and Type III consist of a single protein (Gross *et al.,* 2006; Jenke-Kodama & Dittmann 2009).

1.1.1 Precusor supply for polyketide biosynthesis

Polyketide synthesis requires specific precursors, primarily in the form of acetyl-CoA or malonyl-CoA, to form starter and extender units. Precursor supply can therefore influence the productivity of any secondary metabolite, as a lack of specific precursors can inhibit production, if not abolish it. For this reason it is important to understand the metabolic pathways that supply secondary metabolite precursors.

When considering precursor supply the importance of primary metabolism is evident, as many of the precursor molecules required for secondary metabolism are provided via primary metabolism, and primarily central carbon metabolism (Rokem *et al.,* 2007; Hwang *et al.,* 2014). Further to this only 12 precursor metabolites, converted from a variety of carbon sources, that are used in the production of all cellular components, are required for the production of secondary metabolite precursors (Rokem *et al.,* 2007). The 12 precursor metabolites supplied by primary metabolism are; glucose-6-phosphate (P), fructose-6-P, ribose-5-P, erythrose-4-P, glyceraldehyde-3-P, 3-phosphoglycerate, phosphoenolpyruvate, pyruvate, acetyl-CoA, 2-oxoglutarate, succinyl-CoA and oxaloacetate. The majority of these precursors are generated by glycolysis, however the pentose phosphate pathway (PPP) and the tricarboxylic acid (TCA) cycle are also involved (Figure 1.1); (Ryu *et al.,* 2006; Kern *et al.,* 2007; Zha *et al.,* 2009; Pickens *et al.,* 2009). It is from these precursors that the molecules incorporated into the various secondary metabolites are generated. For example, the aromatic amino acids required for glycopeptide biosynthesis are provided through the shikimate pathway. The first step of the shikimate pathway is the generation of deoxy-D-arabino-heptulosonate-7-phosphate, which requires erythrose-4-

P, from the PPP, and phosphoenolpyruvate, from glycolysis (Figure 1.1); (Knaggs *et al.,* 1999). In addition the Type-I polyketides FK506 and rapamycin require dihydroxycyclohexanecarboxylic acid for their biosynthesis, which is derived from shikimic acid, supplied via the shikimate pathway (Wilson *et al.,* 1998).

The majority of Type-I and Type-II PKS use acetyl-CoA or malonyl-CoA for extender or starter units, however Type-I PKS can also use propionyl-CoA, butyryl-CoA, ethylmalonyl-CoA, methylmalonyl-CoA or methoxymalonyl-CoA (Chan *et al.,* 2008; Stirrett *et al.,* 2008). Acetyl-CoA is the main precursor associated with Type-I and II polyketides and is generated either from pyruvate, produced via glycolysis, by the action of pyruvate dehydrogenase or from acetate by acetyl-CoA synthase (Figure 1.1); (Zha *et al.,* 2009). The primary means of generating malonyl-CoA is by carboxylation of acetyl-CoA by acetyl-CoA carboxylases (Ryu *et al.,* 2006; Zha *et al.,* 2009). It has been shown that overexpression of acetyl-CoA carboxylase in *S. coelicolor* can increase the production of ACT, the PKS of which uses malonyl-CoA (Ryu *et al.,* 2006). The alternative means of producing malonyl-CoA is via direct conversion of malonate, using malonyl-CoA synthetase, however this process does not appear to be as commonly utilised (Chan *et al.,* 2008).

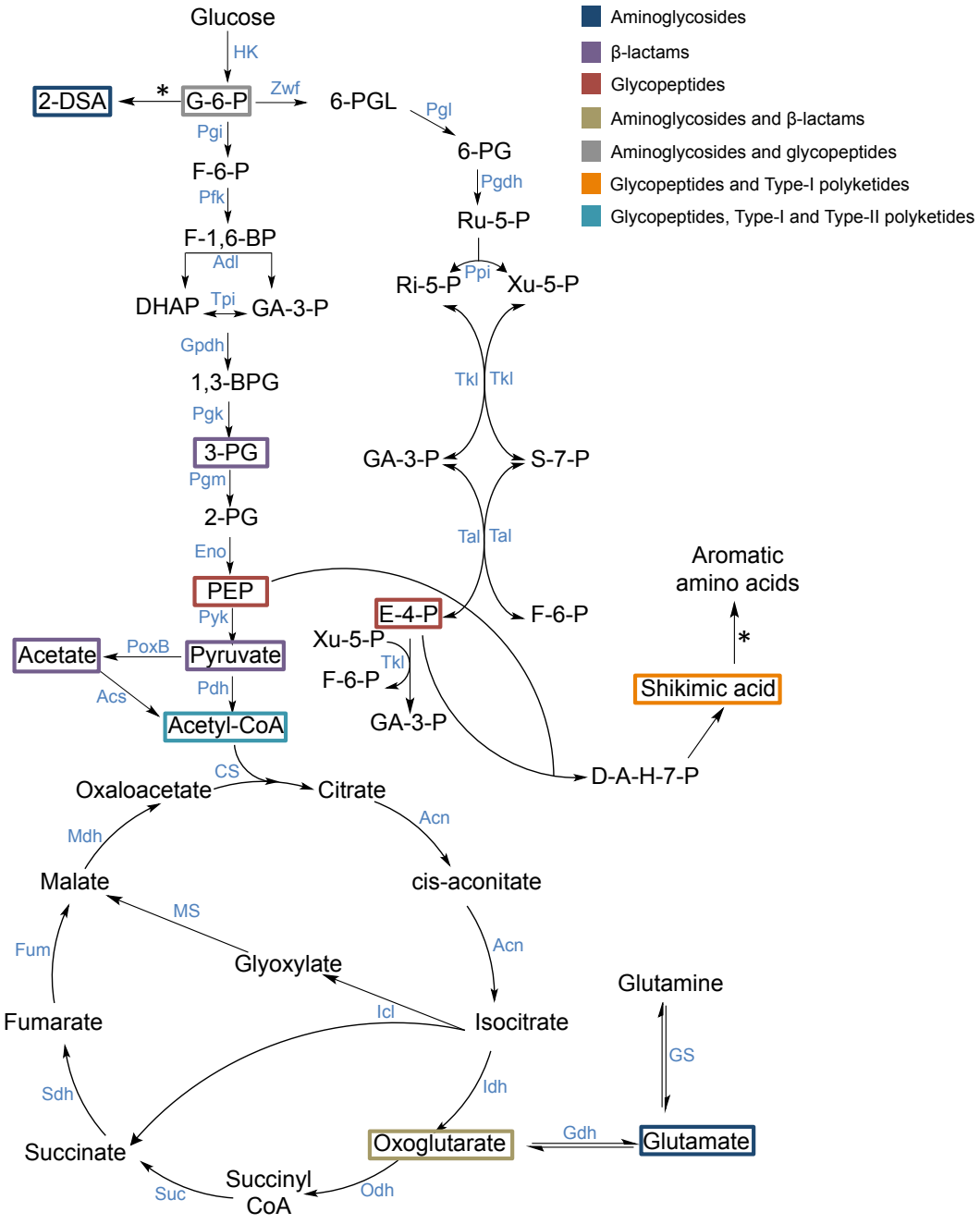

Aminoglycosides
β-lactams
Glycopeptides
Aminoglycosides and β-lactams
Aminoglycosides and glycopeptides
Glycopeptides and Type-I polyketides
Glycopeptides, Type-I and Type-II polyketides
Glucose
HK
2-DSA
*
G-6-P
Zwf
6-PGL
Pgl
6-PG
Pgi
F-6-P
Pfk
Pgdh
Ru-5-P
F-1,6-BP
Adl
Ri-5-P
Ppi
Xu-5-P
DHAP
Tpi
GA-3-P
Gpdh
1,3-BPG
Tkl
Tkl
Pgk
3-PG
GA-3-P
S-7-P
Pgm
2-PG
Tal
Tal
Eno
Aromatic amino acids
PEP
E-4-P
F-6-P
Pyk
Xu-5-P
*
Acetate
PoxB
Pyruvate
Tkl
F-6-P
Shikimic acid
Acs
Pdh
GA-3-P
Acetyl-CoA
CS
D-A-H-7-P
Oxaloacetate
Citrate
Acn
Mdh
Malate
cis-aconitate
MS
Acn
Fum
Glyoxylate
Glutamine
Icl
Fumarate
Isocitrate
GS
Sdh
Idh
Gdh
Succinate
Oxoglutarate
Glutamate
Succinyl CoA
Suc
Odh

Figure 1.1 Primary metabolic pathways associated with secondary metabolism precursor supply. Molecules with rectangular boxes represent a precursor for the associated secondary metabolite, as shown at the top right hand corner of the figure. Molecule abbreviations: **G-6-P**, Glucose-6-phosphate; **F-6-P**, Fructose-6-phosphate; **F-1,6-BP**, Fructose-1,6-biphosphate; **GA-3-P**, Glyceraldehyde-3-phosphate; **DHAP**, Dihydroxy-acetone-phosphate; **1,3-BPG**, 1,3-biphosphoglycerate; **3-PG**, 3-phosphoglycerate; **2-PG**, 2-phosphoglycerate; **PEP**, Phosphenolpyruvate; **6-PGL**, 6-phosphoglucolactone; **6-PG**, 6-phosphogluconate; **Ru-5-P**, Ribulose-5-phosphate; **Ri-5-P**, Ribose-5-phosphate; **Xu-5-P**, Xylulose-5-phosphate; **S-7-P**, Sedoheptulose-7-phosphate; **E-4-P**, Erythrose-4-phosphate; **D-A-H-7-P**, Deoxy-D-arabino-heptulosonate-7-phosphate; **2-DSA**, 2-deoxystreptamine. Enzyme abbreviations: **HK**, Hexokinase; **Pgi**, Phosphoglucose isomerase; **Pfk**, Phosphofructokinase; **Adl**, Adolase; **Tpi**, Triosephosphate isomerase; **Gpdh**, Glyceraldehyde phosphate dehydrogenase; **Pgk**, Phosphoglycerate kinase; **Pgm**, Phosphoglycerate mutase; **Eno**, Enolase; **Pyk**, Pyruvate kinase, **PoxB**, Pyruvate dehydrogenase/oxidase; **Acs**, Acetyl-CoA synthase; **Pdh**, Pyruvate dehydrogenase; **CS**, Citrate synthase; **Acn**, Aconitase; **Idh**, Isocitrate dehydrogenase; **Icl**, Isocitrate lyase; **MS**, Malate synthase; **Odh**, 2-oxoglutarate dehydrogenase; **Suc**, Succinyl-CoA synthetase, **Sdh**, Succinate dehydrogenase; **Fum**, Fumarase, **Mdh**, Malate dehydrogenase; **Zwf**, Glucose-6-phosphate dehydrogenase; **Pgl**, Phosphoglucolactonase; **Pgdh**, Phosphogluconate dehydrogenase; **Ppi**, Phosphopentose isomerase; **Tkl**, Transketolase; **Tal**, Transaldolase. Asterisks represent pathways missing intermediate steps. Figure derived from (Ryu *et al.,* 2006; Kern *et al.,* 2007; Rokem *et al.,* 2007).

1.1.2 Type I polyketide synthases

Type I PKSs are involved in the biosynthesis of such polyketides as the antibiotics erythromycin (Figure 1.2) and rifamycin, the immunosuppressant drugs FK506 and rapamycin, and the antiparasitic drugs avermectin and hitachimycin (Keating & Walsh 1999; Lal *et al.,* 2000; Moore & Hertweck 2002). A Type I PKS module can consist of a ketosynthase (KS), an acyl transferase (AT), dehydratase (DH), enoyl reductase (ER), ketoreductase (KR), thioesterase (TE) and an acyl-carrier protein (ACP), depending on the step of the biosynthetic pathway that is carried out by that module (Figure 1.2); (Rawlings 2001a; Cane 2010; Starcevic *et al.,* 2011). Modular polyketide synthesis begins with a starter unit, generally an acetate or propionate, but there are some other "novel" starter units, which are then condensed in a Claisen-like manner forming carbon to carbon bonds and generate the extender units (Lal *et al.,* 2000; Moore & Hertweck 2002). The condensation of the starter unit and the subsequent extender units is catalysed by the KS, while the AT influences the type of extender unit incorporated, and the ACP secures the growing polyketide chain to the PKS, as well as accepting the next extender unit from the AT. Elongation of the polyketide chain, which will eventually form the backbone of the polyketide itself, is primarily achieved by the AT, ACP and KS and thus these enzymatic sites together are referred to as the minimal PKS (Lal *et al.,* 2000; Rawlings 2001a; Moore & Hertweck 2002). Modification of the polyketide chain is accomplished by the DH, ER and KR domains, with the degree of modification determined by the presence of these domains within a module (Cane 2010; Kong *et al.,* 2013). The TE domain is responsible for the release, and in some circumstances, cyclisation of the polyketide chain (Du & Lou 2010; Xu *et al.,* 2013; Williams 2013). The biosynthesis of 6-deoxyerythronolide-B (6-DEB), which is the precursor for erythromycin, is an example of this process (Figure 1.2).

The DEBS PKS consists of a loading domain and 6 modules encoded by *eryAI, eryAII*, and *eryAIII* (Keatinge-Clay 2008; Cane 2010; Kwan & Leadlay 2010). The loading domain initiates the polyketide synthesis by accepting the propionyl-CoA starter unit and transferring it to the KS of module 1. Subsequently, the growing chain is passed along the PKS via the ACP domain of each module (Cane 2010; Valenzano *et al.,* 2010). Polyketide chain extension is related to the number of modules, as each module partakes in a single round of extension and therefore knowing the number of modules can allow for determination of the chain length (Lal *et al.,* 2000; Rawlings 2001a; Cane 2010). By examining the functional domains of the PKS module, it is possible to predict the type of modification carried out on the chain. An example of this is module 3 of the DEBS PKS (Figure 1.2), which has no reductive domains, therefore it produces a ketoester before transferring the chain to the ACP of module 4 (Rawlings 2001a; Chen *et al.,* 2008a; Cane 2010). Module 6 contains a thioesterase domain which catalyses the cyclisation and release of the polyketide chain to form 6-DEB, which then allows for post-PKS modifications to take place forming erythromycin A (Figure 1.2).

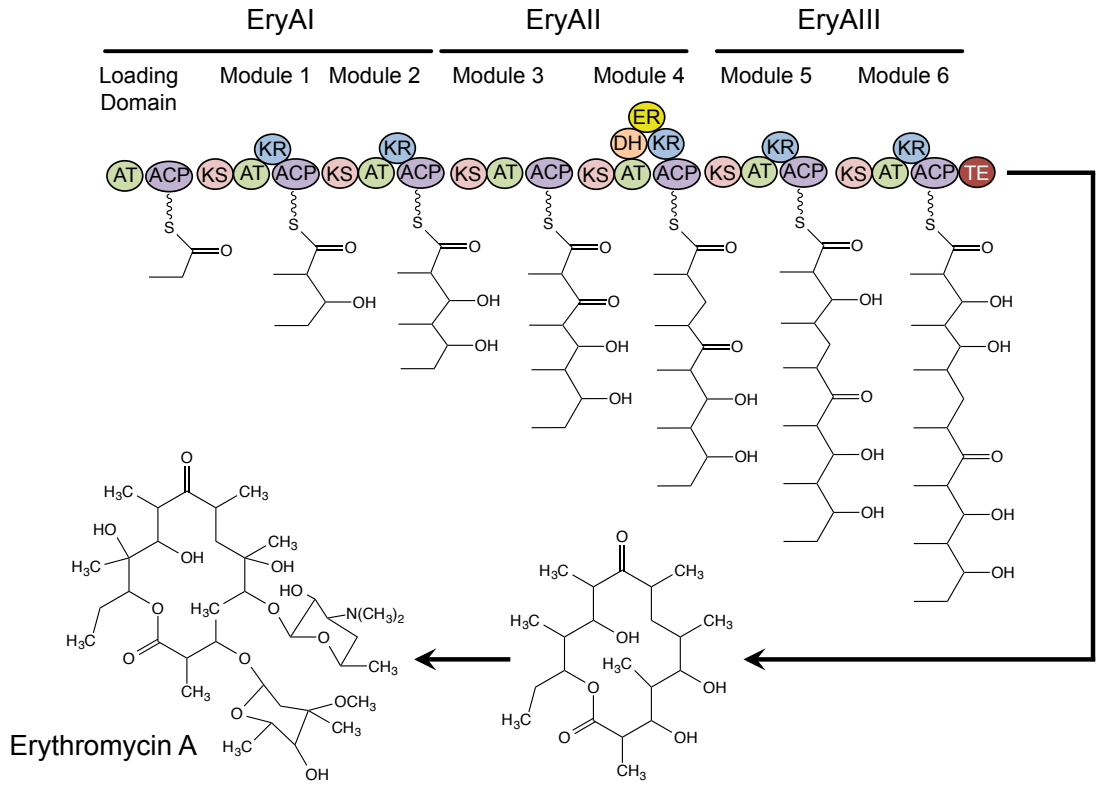


Figure 1.2 Biosynthetic pathway of 6-deoxyerythronolide-B. Enzymatic domains present in each module are marked with the following ACP = acyl carrier protein, KS = ketosynthase, AT = acyl transferase, KR = ketoreductase, DH = dehydratase, ER = enoyl reductase, TE = thioesterase. Derived from Cane, 2010.

1.1.3 Type III polyketide synthases

Type III polyketides are generally associated with secondary metabolism in plants however in recent years these molecules have also been identified in bacteria (Gross *et al.,* 2006). The most common example of a plant type III polyketide is chalcone (Figure 1.3), which is required for the biosynthesis of phytoalexins. These molecules, as well as flavonoids such as anthocyanin, act as a defence mechanism against microbes (Jez *et al.,* 2001). Chalcone and other Type III polyketides are produced via type III PKS which are single proteins containing ketosynthase activity allowing Claisen-like condensation reactions to take place (Austin & Noel 2002).

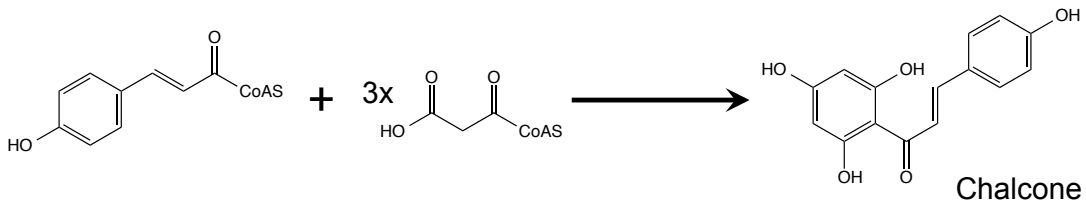


Figure 1.3 Chalcone formation by the chalcone synthase (CHS). p-coumaroyl is transferred to the CHS which initiates the reaction allowing for the subsequent condensation reactions with 3 acetate units and finally leads to the formation of chalcone. Derived from Jez *et al.,* 2001.

1.1.4 Type II polyketide synthase

The antibiotics chlortetracycline and actinorhodin (ACT), as well as the anti-tumour drugs mithramycin and chromomycin, are synthesised by Type II PKS (Hertweck *et al.,* 2007). Like Type I PKS, Type II PKS have components that can be regarded as the minimal PKS that forms the backbone of the polyketide (Bisang *et al.,* 1999; Hertweck *et al.,* 2007; Das & Khosla 2009). The minimal PKS of Type II PKS consists of a ketosynthase (α subunit), which catalyses the Claisen-like condensation, a chain length factor (CLF), also known as the ketosynthase β subunit, and an acyl carrier protein (ACP) acts as the anchor for the chain. The KSβ subunit is thought to control the number of iterations and therefore the final length of the polyketide backbone. This domain is also involved in acceptance of the starter unit and subsequent extender units. (Zhang *et al.,* 2006; Zhang *et al.,* 2008; Das & Khosla 2009). The KSα, KSβ and ACP come together to form a complex (Figure 1.4), which assembles the polyketide chain (Hertweck *et al.,* 2007; Castaldo *et al.,* 2008). It has been proposed that a malonyl-CoA:ACP transferase (MCAT) is required as a fourth component of the minimal PKS in order to prime the minimal PKS with an appropriate starter unit (Hertweck *et al.,* 2007; Beltran-Alvarez *et al.,* 2007; Castaldo *et al.,* 2008). For ACT biosynthesis FabD, from the FAS, is recruited as the MCAT (Carreras & Khosla 1998; Khosla 2009). Type II PKS pathways lack genes for MCAT and there is evidence that FAS MCAT can interact with ACP of other pathways; therefore it is possible that MCAT are recruited from FASs for most, Type II PKS (Arthur *et al.,* 2009). However there is also evidence that such priming via a MCAT is unnecessary as the minimal PKS can self-prime through the ACP. Studies have shown that priming through self-malonylation in vitro with malonyl, methylmalonyl and succinyl CoAs can be achieved in the absence of a MCAT for the ACT, oxytetracycline and griseusin ACPs. This suggests that there may be two routes by which Type II polyketide biosynthesis can be initiated (Hitchman *et al.,* 1998; Arthur *et al.,* 2005).

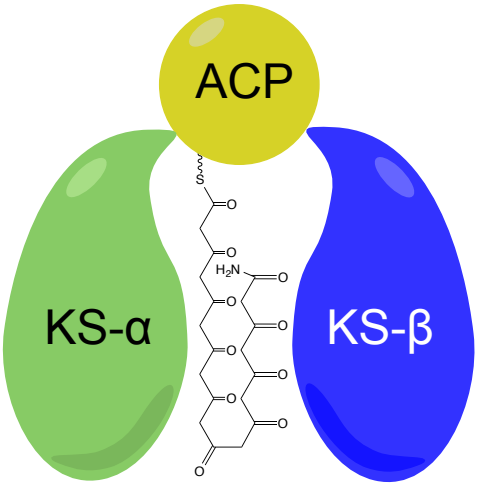


Figure 1.4 Representation of the KSα, KSβ, and ACP minimal PKS complex. Derived from Hertweck *et al.,* 2007.

After production of the polyketide, it is converted into its active natural product by various proteins which function to either fold, cyclise or add chemical groups such as keto- or methyl- groups (Hertweck *et al.,* 2007; Das & Khosla 2009). For this reason the PKS consists of genes encoding modifying proteins such as KR, aromatases (ARO), and cyclases (CYC) (Shen *et al.,* 1999; Pickens & Tang 2009; Pickens & Tang 2010). This differs from Type I polyketides in which the backbone is modified during elongation instead of backbone completion and then modification. The biosynthetic pathway for ACT production provides an example of this process (Figure 1.5).

The ACT biosynthetic cluster consists of 3 regions; an early and intermediate, central and late, and intermediate. The early and intermediate region contains the genes *actI, actIII, actIV, actVB*, and *actVII*, the central region contains the gene *actII*, with the late and intermediate region containing the genes *actVA* and *actVI* (Fernández-Moreno *et al.,* 1991; Fernández-Moreno *et al.,* 1994). ACT biosynthesis begins with an acetate starter unit, which is condensed with a malonate extender unit to initiate chain elongation and the addition of 6 more malonate units (Fernández-Moreno *et al.,* 1994; Okamoto *et al.,* 2009). As previously stated FabD is recruited during ACT biosynthesis to provide the MCAT activity, which primes the minimal PKS with the acetate starter unit (Carreras & Khosla 1998). The minimal PKS consists of the genes *actI-ORF1*, *actI-ORF2* and *actI-ORF3* that encode the KS α subunit, the chain length factor/KS β subunit and the ACP respectively. *actIII* encodes a KR, *actIV* encodes a DH and *actVII* encodes a DH/CYC whose role are to form the bicyclic intermediate from the polyketide backbone (Fernández-Moreno *et al.,* 1994; Das & Khosla 2009). After this the bicyclic intermediate is modified further by ActVI-ORF3 to form 4-dihydro-9-hydroxy-1-methyl-10-oxo-3-H-naphtho-(2,3-c)-pyran-3-acetic acid ((S)-DNAP) which provides DH activity, following this ActVI-ORF2 and ActVI-ORF4 act as oxidoreductases (OR) and lead to the formation of 6-deoxy-dihydrokalafungin (DDHK); (Figure 1.5); (Fernández-Moreno *et al.,* 1994). DDHK is then converted to dihydrokalafungin via the oxygenase

activity of ActVA-ORF6, and then into the ACT intermediate by ActVA-ORF5 (Figure 1.5). Finally ACT is formed when ActVB catalyses the formation of a dimer between two molecules of the ACT intermediate (Figure 1.5); (Fernández-Moreno *et al.,* 1991).

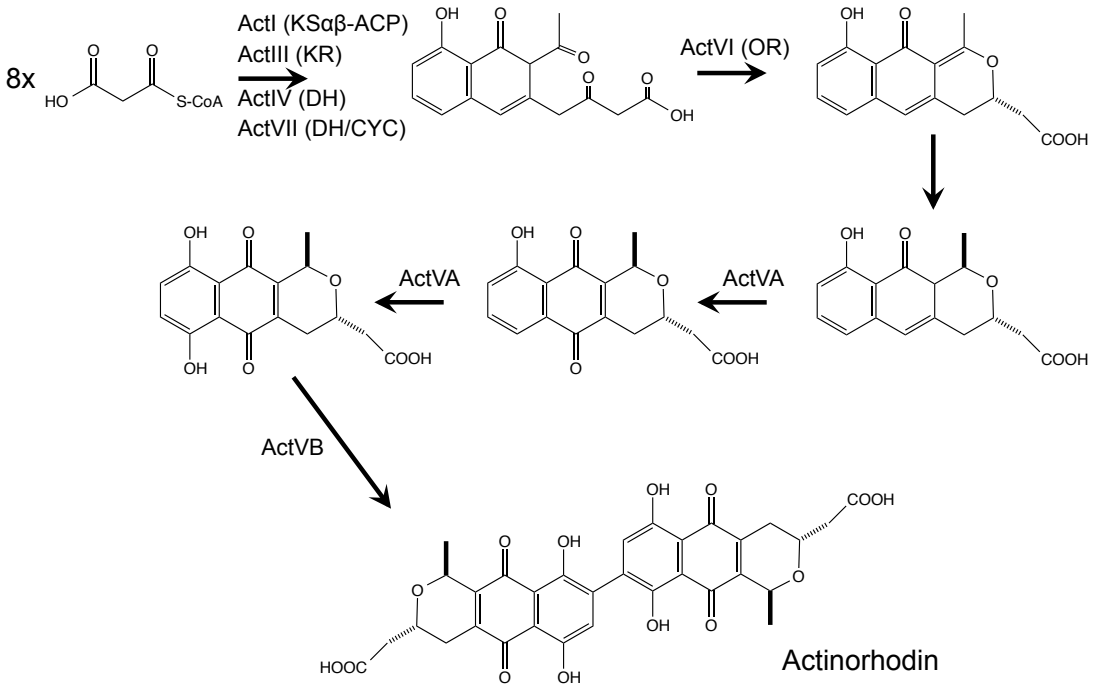


Figure 1.5 The biosynthetic pathway for actinorhodin. The gene *actI* encodes the minimal PKS, and in conjunction with the products of *actIII*, *actVII* and *actIV* produces the bicyclic intermediate from an acetate starter unit and 7 malonate extender units. The *actVI* product catalyses the reductive activity at C-3/C-15 which leads to the formation of 4-dihydro-9-hydroxy-1-methyl-10-oxo-3-H-naphtho-(2,3-c)-pyran-3-acetic acid ((S)-DNPA). The proteins produced by the *actVA* gene are involved in the formation of dihydrokalafungin and the ACT intermediate. Then two ACT intermediate compounds form a dimer via the action of the product of *actVB*. Derived from Fernández-Moreno *et al.,* 1994.

1.2 Oxytetracycline

Oxytetracycline (OTC) is a Type II polyketide antibiotic produced by *Streptomyces rimosus* and is used in aquaculture, agriculture (both plant and animal) and in human medicine (Kerry *et al.,* 1995; McManus *et al.,* 2002). The antimicrobial mechanism of OTC is achieved by binding reversibly to the 30S subunit of the ribosome, conformationally changing the binding site for tRNA and thus preventing the production of proteins (Connell *et al.,* 2003; Wang *et al.,* 2009b). OTC was first described by Finlay *et al.,* in 1950, when *S. rimosus* was first isolated, under the trade name Terramycin (Finlay & Hobby 1950), and since then extensive research has been undertaken to examine and manipulate the biosynthetic pathway of OTC (Petković *et al.,* 2006), allowing for the elucidation of the genes involved in the biosynthesis and a predicted pathway for its biosynthesis (Figure 1.6).

1.2.1 OTC biosynthesis

The OTC gene cluster (Figure 1.6/Table 1.1) is flanked by two genes, *otrB* and *otrA*, which confer resistance to OTC protecting the bacteria from the antibiotic it is producing (Doyle *et al.,* 1991). OtrA provides resistance by interacting with the ribosome and dislodging any bound OTC, while OtrB functions by lowering the concentration of OTC within the cell by efflux, common to other tetracycline resistance mechanisms (Doyle *et al.,* 1991; Petković *et al.,* 2006; Mak *et al.,* 2014). The OTC minimal PKS is composed of OtcY1-1, OtcY1-2 and OtcY1-3 and unlike most type II PKS is primed with an amide unit, rather than acetate, giving rise to speculation about the proteins involved in priming the PKS (Zhang *et al.,* 2006; Pickens & Tang 2010). It has been suggested that OtcX3 would load the PKS as it is a putative acyltransferase, using malonamyl-CoA provided by OtcY2-3, predicted to be a CoA ligase, by attaching a CoA to a malonamate supplied by the amidotransferase OtcY1-4 (Petković *et al.,* 2006). However it has also been suggested that OtcY1-4 adds an amide unit directly to either malonyl-

CoA or malonyl-ACP, which can then be used to prime the PKS *(Zhang et al.,* 2006). The role for OtcY1-4 has recently been confirmed, and was shown to be essential for the biosynthesis of malonamate, by acting as the aminotransferase. Further to this OtcX3 was shown to suppress mispriming with acetate, therefore assisting priming by a malonamyl unit (Wang *et al.,* 2011).

After the 19 carbon backbone is generated from the amide starter unit and 8 acetate extender units (Figure 1.7), further modification occurs initially by OtcY2-1 that acts as the C-9 ketoreductase (Zhang *et al.,* 2006). This is followed by formation of the D-ring catalysed by OtcD1 and subsequent C-ring formation by OtcD4 (Figure 1.7) (Petković *et al.,* 1999). B-ring formation appears to be spontaneous; however A-ring formation, like D and C, is catalysed by the cyclase OtcY2-2 forming pretetramid (Figure 1.7, structure 2). A methyl group is then added to C-6, producing 6-methylpretetramid (6-MPT), by OtcY2-5 predicted to be a methyltransferase (Figure 1.7, structure 3) (Zhang *et al.,* 2006; Petković *et al.,* 2006). The conversion of 6-MPT into 4-hydroxy-6-MPT has not been fully determined however most of the research into this biosynthetic step agrees that either OtcY1-5 or OtcD2 are required. Depending on the hypothesis, the role for OtcY1-5 is thought to differ from a primary role to that of an accessory, with the biosynthetic step in which OtcD2 acts differing as well. Research has shown that deletion of *otcY1-5* does not disrupt OTC production, however the final titre of OTC was lower than that of the wild-type and a shunt product was produced that lacked the hydroxyl group at C-4 (Wang *et al.,* 2009b). Given these results it was suggested that OtcY1-5 ensures hydroxylation at C-4, as the shunt product is not detected for wild-type strains, but is not essential for this hydrolysis.

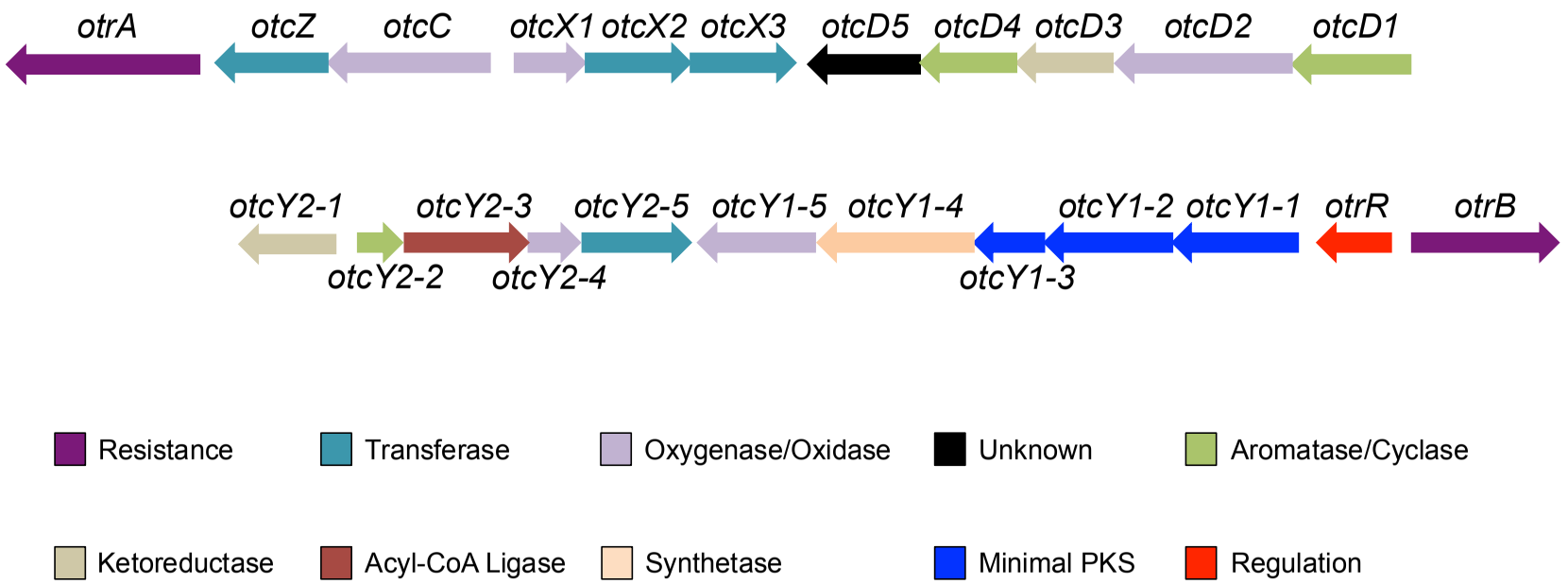


Figure 1.6 Organisation of the OTC biosynthetic gene cluster. A generalized definition of the role of each protein has been highlighted. Derived from Petković *et al.,* 2006.

Gene[α]	Predicted Function	Gene Position	
		Start	End
otrR (oxyTA1)	Transcription Regulation	134	625
otcY1-1 (oxyA)	Ketosynthase	770	2047
otcY1-2 (oxyB)	Chain Length Factor	2044	3312
otcY1-3 (oxyC)	Acyl Carrier Protein	3388	3675
otcY1-4 (oxyD)	Asparagine Synthetase	3689	5524
otcY1-5 (oxyE)	Oxygenase	5521	6777
otcY2-5 (oxyF)	Methyltransferase	6809	7846
otcY2-4 (oxyG)	Oxygenase	7899	8186
otcY2-3 (oxyH)	Acyl CoA Ligase	8241	9833
otcY2-2 (oxyI)	Cyclase	9872	10324
otcY2-1 (oxyJ)	C-9 Ketoreductase	10593	11384
otcD1 (oxyK)	Aromatase	11443	12396
otcD2 (oxyL)	Oxygenase	12386	14059
otcD3 (oxyM)	Ketoreductase	14077	14835
otcD4 (oxyN)	Cyclase	14855	15628
otcD5 (oxyO)	Unknown	15734	16795
otcX3 (oxyP)	Acyltransferase	16813	17835
otcX2 (oxyQ)	Aminotransferase	17832	18911
otcX1 (oxyR)	PNP-oxidase	18908	19336
otcC (oxyS)	Oxygenase	19492	21003
otcZ (oxyT)	*N*-Methyltransferase	21064	22098

Table 1.1 OTC biosynthetic genes and predicted function. α Genes present within parentheses are the designations assigned on Genbank (Accession No. DQ143963). The genes outwith the parentheses are the original ID given to the *otc* PKS loci, and the designation used in this thesis.

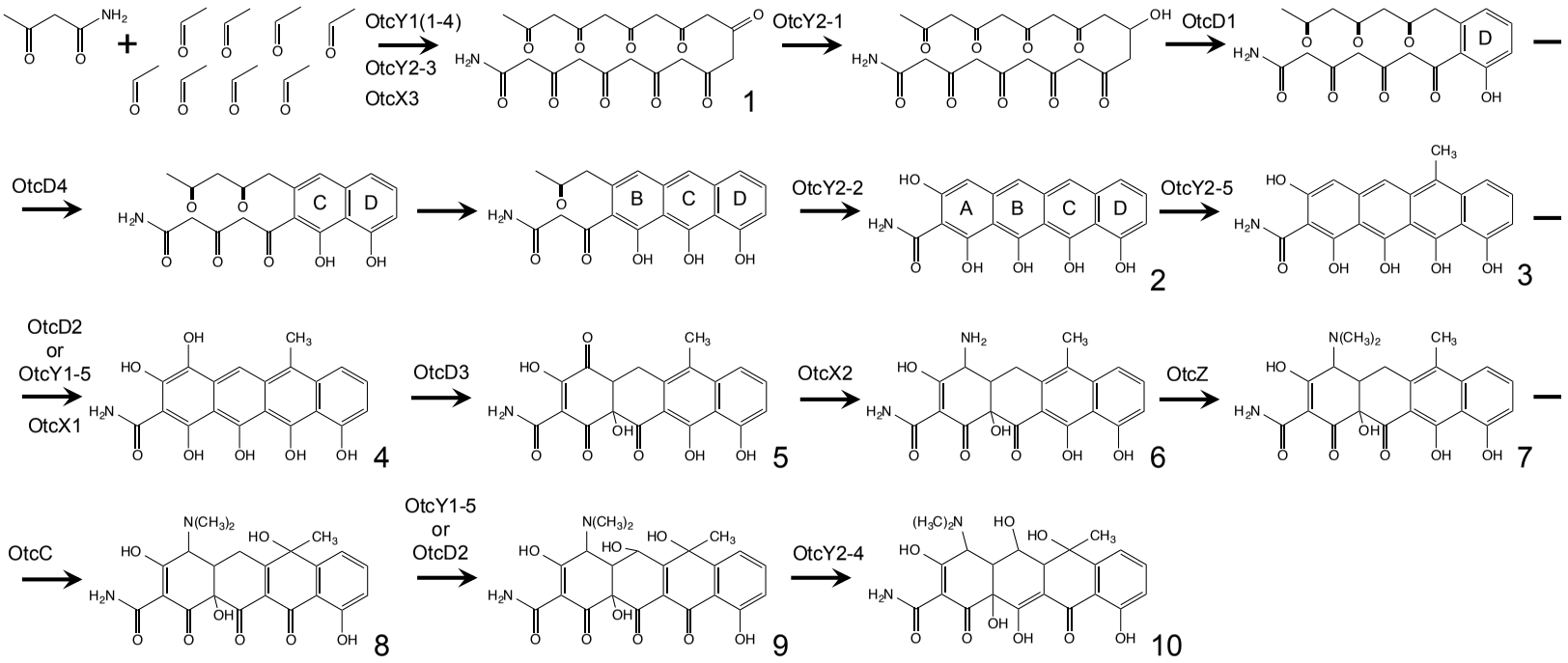


Figure 1.7 Predicted pathway for oxytetracycline biosynthesis. The genes *otcY1-1, otcY1-2, otcY1-3* encode the minimal PKS, and in conjunction with the products of *otcX-3, otcY1-4* and *otcY2-3* are able to produce the initial polyketide chain (structure 1). Further modification is achieved by other proteins that make up the PKS. Specific structures are labeled with numbers and are as follows: 1 = Nonaketamide, 2 = Pretetramid, 3 = 6-methylpretetramid (6-MPT), 4 = 4-hydroxy-6-MPT, 5 = 4-keto-6-MPT, 6 = 4-aminodedimethylaminoanhydrotetracycline (4-amino-ATC) 7 = Anhydrotetracycline (ATC), 8 = 5a,11a-dehydrotetracycline (5-DHTC), 9 = 5a,11a-dehydrooxytetracycline (DHOTC) and 10 = Oxytetracycline. The cyclised rings are labeled, A, B, C and D. Derived from Petković *et al.,* 2006.

Deletion of *otcD2* resulted in the abolishment of OTC production and produced a shunt product in which C-4 hydrolysis had occurred. This suggests that OtcD2 is required for progression past 6-MPT (Zhang *et al.,* 2008; Wang *et al.,* 2009b). It has been proposed that OtcD2 is capable of bypassing production of 4-hydroxy-6-MPT (Figure 1.8), converting 6-MPT into 4-keto-6-MPT, but this process is less efficient resulting in the lower titre of OTC, as seen with the *otcY1-5* mutants (Wang *et al.,* 2009b). Previous work examining the role of OtcD2 in the production of anhydrotetracycline (ATC), as well as the role of OtcY1-(1-4), OtcY2-1, OtcY2-5, OtcD1, OtcD4, OtcX2, and OtcZ, demonstrated that together these proteins could produce ATC without the requirement of OtcY1-5 (Zhang *et al.,* 2008). These results support the hypothesis that OtcD2 can bypass the 4-hydroxy-6-MPT step and convert 6-MPT into 4-keto-6-MPT directly (Figure 1.8). It has been suggested that OtcX1 may provide an ancillary role to either OtcY1-5 or OtcD2, but given the results just discussed, it seems likely that if OtcX1 does it would not be essential for OtcD2. Further to this, it is hypothesised that OtcY1-5 acts as an accessory protein enhancing and ensuring the hydrolysis of C-4, with OtcX1 being the actual catalyst of the reaction (Wang *et al.,* 2009b). This would explain the hydrolysis of C-4 even in the absence of OtcY1-5. The results gained from examining the production of ATC suggest that lack of OtcX1 does not influence the progression past 6-MPT. However this does not explain whether this is due to OtcD2 bypassing the step, and therefore no need for C-4 hydrolysis, or hydrolysis of 6-MPT occurring but at lower efficiency. If the latter, then OtcX1 and OtcY1-5 may work together in synergy to add the hydroxyl group to C-4, enhancing the efficiency, which would account for the occurance of hydrolysis even in the absence of either OtcX1 or OtcY1-5. If the former then neither OtcX1 or OtcY1-5 are required for hydrolysis of 6-MPT.

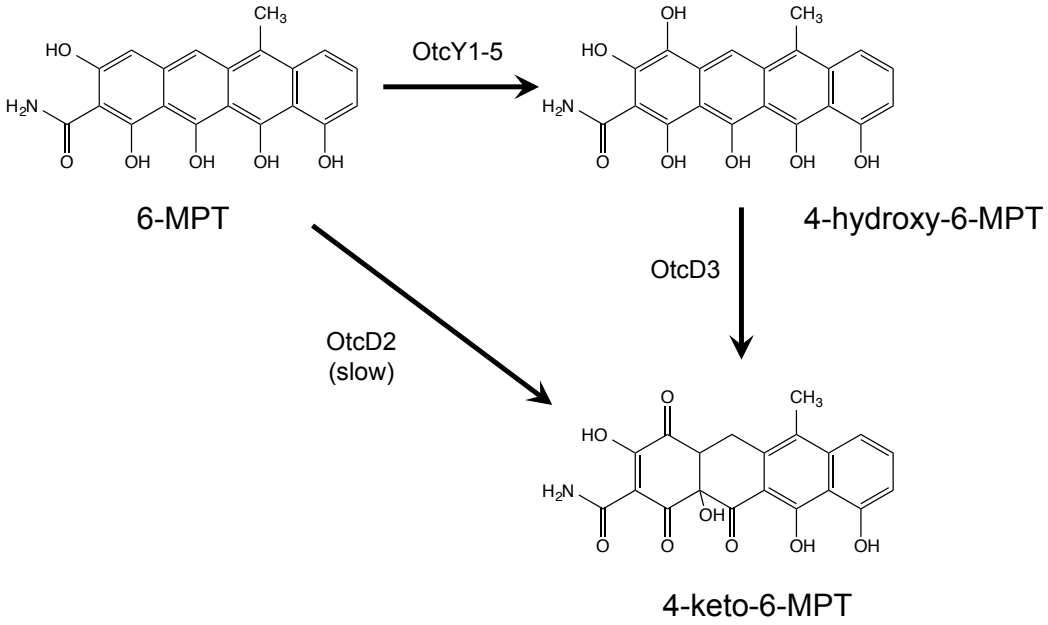


Figure 1.8 Conversion of 6-MPT to 4-keto-6-MPT. Conversion of 6-MPT to 4-keto-6-MPT has been predicted to occur either via the action of OtcY1-5, with the 4-hydroxy-6-MPT as an intermediate molecule converted to 4-keto-6MPT by the action of OtcD3. An alternative pathway for the conversion of 6-MPT to 4-keto-6-MPT is a direct conversion via the action of OtcD2; however, this is predicted to occur at a slower rate. Derived from Petković *et al.,* 2006.

Regardless of the mechanism, 6-MPT, is converted to 4-hydroxy-6-MPT, which is then converted to 4-keto-6-MPT (Figure 1.8). This is achieved by reduction of the hydroxyl group at C-4 which is achieved by the ketoreductase OtcD3; although, it has also been suggested that OtcD2 is responsible (Petković *et al.,* 2006; Pickens & Tang 2010). Following this, the aminotransferase OtcX2 introduces an amino group to the C-4 position of 4-keto-6-MPT forming 4-amino-ATC (Figure 1.7, structure 6), followed by the addition of two methyl groups to the same position. This is most likely carried out by OtcZ, forming ATC (Figure 1.7, structure 7) (Petković *et al.,* 2006; Zhang *et al.,* 2008). Hydroxylation of C-6 results in the formation of 5a,11a-dehydrotetracycline (5-DHTC) (Figure 1.7, structure 8) and this is accomplished by OtcC when in the presence of NADPH and oxygen (Peric-Concha *et al.,* 2005; Pickens & Tang 2010). 5-DHTC is then converted to 5a,11a-dehydrooxytetracycline (DHOTC) (Figure 1.7, structure 9) by addition of a hydroxyl group to C-5 however the exact means for this process has not be determined. It was originally proposed that either OtcY1-5 or OtcD2 catalyse the hydrolysis at C-5; however, given the results of recent studies, this may not be the case (Pickens & Tang 2010). Although OtcD2 is less efficient at converting 6-MPT to 4-hydroxy-6-MPT, it seems likely that the primary role for this protein is at the conversion of 5-DHTC to DHOTC. The final product of OTC is then formed from DHOTC by a dehydrogenase. OtcY2-4 had been putatively assigned to this step as OtcY2-4 shares amino-terminal sequence homology with the protein that converts 5-DHTC to tetracycline (TC) in *S. aureofaciens* (Petković *et al.,* 2006). It is most likely one of the proteins encoded by the genes (table 1.1) within the OTC biosynthetic cluster catalyses this reaction, as production of OTC by *S. lividans, S. albus* and the deltaproteobacterium *Myxococcus xanthus* occurs when the OTC biosynthetic cluster is introduced (Binnie *et al.,* 1989; Stevens *et al.,* 2010; Wang *et al.,* 2012a).

1.2.2 Regulation of the OTC biosynthetic cluster

OTC production can be induced by phosphate starvation; however, the molecular mechanisms of this regulation have yet to be determined (McDowall *et al.,* 1999). Genes within the OTC biosynthetic cluster have promoters that are overlapped by tandem repeats and the spacing of these repeats are reminiscent of the activation sites for the OmpR class of transcription factors (McDowall *et al.,* 1999). *Streptomyces* antibiotic regulatory proteins (SARP) are associated with the OmpR protein family and it is likely that a SARP-like protein is responsible for the initiation of transcription in correlation with decreased phosphate levels (McDowall *et al.,* 1999; Bibb 2005). So far only proteins have been identified that play a role in the regulation of OTC are, OtrR and OtcG. OtrR negatively regulates the expression of the OTC efflux protein OtrB, so as to prevent expression when OTC is not produced (MacGregor-Pryde 1995; Lenik 2009; Mak *et al.,* 2014). OtrR may regulate the expression of other genes within the cluster though this is uncertain, as there is no evidence to support or rule out this possibility. OtcG is reminiscent of the NarL family of proteins (Baikalov *et al.,* 1996) as it contains a amino-terminal (N-terminal) receiver (REC) response domain and a carboxy-terminal (C-terminal) LuxR-type HTH domain; this has recently been identified as a positive regulator of OTC biosynthesis (Lenik 2009). Following the discovery of OtcG it was described as a member of the LAL (Large ATP-binding regulators of the LuxR family) family of regulators, however this categorisation seems unlikely as it is missing the characteristic large ATP-binding domain associated with these regulators (De Schrijver & De Mot 1999; Santos *et al.,* 2012; Guerra *et al.,* 2012). The *otcG* gene is located adjacent to the resistance gene *otrA*, outside the proposed OTC gene cluster, and is transcribed in the opposite direction (Lenik 2009). Gene inactivation and overexpression studies of *otcG,* found that it encodes a positive regulator of OTC production, but is not essential (Lenik 2009). Inactivation of *otcG* resulted in a significant reduction (40%) in OTC production. However, overexpression of *otcG* did not increase OTC

production and so it would appear that OtcG is a conditionally positive regulator (Lenik 2009).

1.3 Antibiotic pathway specific regulators

Secondary metabolites, such as antibiotics, are produced by microorganisms during the stationary phase of growth when secondary metabolism begins (Ruiz *et al.,* 2010; Berdy 2012). Understanding the mechanisms behind the initiation and regulation of secondary metabolism is of great importance and provides the knowledge necessary to manipulate these pathways with the aim of more efficient production. Thus, increasing the productivity and profitability of an antibiotic. As the genus *Streptomyces* has the greatest number of antibiotic producing species, producing around 70% of all commercially available antibiotics (Watve *et al.,* 2001), it is logical to examine activation of secondary metabolism for *Streptomyces sp.,* to better understand this process. Antibiotic-producing species of *Streptomyces* generally synthesise antibiotics during stationary growth when nutrients become limiting. Thus, physiological differentiation coincides with morphological differentiation and the formation of aerial hyphae on solid media. This is seen as a survival strategy to protect dwindling resources and nutrients provided by autolysis of vegetative mycelium, which is required for the erection of aerial hyphae, from invading microorganisms (Bibb 2005; Chater 2006; Rigali *et al.,* 2008; Liu *et al.,* 2013). The genes required for the biosynthesis of an antibiotic are generally clustered together and regulated in a pathway-specific fashion.

1.3.1 *Streptomyces* antibiotic regulatory proteins

SARPs are transcriptional activators and many regulators involved in secondary metabolism belong to this family (Sheldon *et al.,* 2002; Bibb 2005; Novakova *et al.,* 2011). SARPs contain a winged helix-turn-helix (HTH) motif at the N-terminal end which is reminiscent of the C-terminal DNA-binding

domain of the OmpR family of regulators (Wietzorrek & Bibb 1997; Martínez-Hackert & Stock 1997; Barbieri *et al.,* 2013). Proteins that are part of the OmpR family generally recognise tandem repeats at the promoters of genes. These tandem repeats are used as recognition and binding sites (Martínez-Hackert & Stock 1997). SARPs are required for the activation of the ACT, daunorubicin and nogalamycin biosynthetic clusters, in the form of ActII-ORF4, DnrI and SnoA respectively (Wietzorrek & Bibb 1997; Arias *et al.,* 1999; Sheldon *et al.,* 2002). Analysis of the promoter regions of some of the genes associated with the biosynthesis of these antibiotics revealed tandem repeats suggesting, that like OmpR proteins, these repeats are recognition sites for these antibiotic's particular SARP (Tang *et al.,* 1996; Wietzorrek & Bibb 1997). SARPs are associated with both Type II polyketides and Type I polyketides as well as other secondary metabolites such as ribosomally and non-ribosomally synthesised peptides, β-lactams, and azoxy compounds (Bibb 2005).

1.3.2 The GntR family of regulators

The GntR family of regulators is one of the most poorly studied in *Streptomyces* species, but are one of the main class of regulators found in the *S. coelicolor* A3(2) genome (Bentley *et al.,* 2002). Members of the GntR family modulate gene expression in response to environmental stimuli and the activity of the protein is modified by binding of an effector molecule (Hillerich & Westpheling 2006; Hoskisson & Rigali 2009). A GntR regulator is associated with *Streptomyces*, in relation to morphology and antibiotic production, in the form of the master regulator DasR (Rigali *et al.,* 2006; Rigali *et al.,* 2008). However, there are also examples of GntR regulators functioning as pathway-specific regulators. Some strains of *Streptomyces platensis* can produce two antibiotics, platensimycin (PTM) and platencin (PTN), via a dual biosynthetic pathway for both PTM and PTN; however, some strains are limited to just PTN production (Smanski *et al.,* 2012). While attempting to identify the gene cluster responsible for the biosynthesis of

PTM/PTN, a gene encoding a GntR family regulator was discovered within the cluster (Smanski *et al.,* 2009). The identified gene was named *ptmR1*, and it was found that *S. platensis* overproduced both PTM and PTN when *ptmR1* was knocked out (Smanski *et al.,* 2009). An almost identical copy of *ptmR1*, *ptnR1*, was found within the biosynthetic cluster of PTN and it was necessary to knockout ptnR1 for PTN production to occur in *S. lividans* expressing the *ptn* cluster (Smanski *et al.,* 2012). Supporting the hypothesis that *ptnR1/ptmR1* are pathway-specific negative regulators of the *ptm* and *ptn* clusters.

In addition to PTM and PTN a GntR like pathway-specific regulator was identified for the ebosin pathway in *Streptomyces sp*. 139. Although Ebosin is an exopolysaccharide, it is still encoded by a secondary metabolite biosynthetic cluster, and further to this Ebosin was also found to have antirheumatic arthritis activity, making it a molecule of commercial interest (Bai *et al.,* 2013). When gene knockouts of *ste1*, the GntR-encoding gene, were generated there was an increase in transcription of the *ste5*, *ste8* and *ste17* genes, whose products are involved in the biosynthesis of Ebosin (Bai *et al.,* 2013). Further to this, electrophoretic mobility analysis (EMSA) revealed that Ste1 binds directly to the promoters of the Ebosin gene cluster (Bai *et al.,* 2013). In both cases the GntR pathway-specific regulators were exerting a negative influence on the production of each metabolite.

1.3.3 The TetR family of regulators

Another family of transcriptional regulators that are generally associated with repression is the TetR family (Ramos *et al.,* 2005). Two examples are *SAV576* and *SAV577* from *S. avermitilis*, whose products are involved with the downregulation of avermectin production (Guo *et al.,* 2013; Guo *et al.,* 2014). Gene disruption of *SAV576* results in a marked increase in the transcription of *aveR*, the pathway-specific activator of the avermectin biosynthetic genes (Kitani *et al.,* 2009; Guo *et al.,* 2013). However, neither SAV576 or SAV577 are known to bind to the promoter of *aveR*, and

therefore the increase in *aveR* expression is thought to be related to an unknown regulatory gene that is possibly under the control of SAV576 (Guo *et al.,* 2014). SAV576 and SAV577 bind the promoter of *SAV575* and repress its expression. As SAV575 is responsible for providing the precursors for avermectin biosynthesis, it is likely that the disruption of *SAV576* and *SAV577* leads to increased avermectin production due to increased precursor supply (Guo *et al.,* 2013; Guo *et al.,* 2014).

S. graminearus produces the antibiotic gougerotin and recently it was shown that the gougerotin biosynthetic cluster is regulated by a TetR transcriptional regulator, GouR (Wei *et al.,* 2014). Unlike the previous example, GouR does not function as a repressor alone but also as an activator. Disruption of *gouR* resulted in a reduction in the amount of gougerotin produced (Wei *et al.,* 2014). When the DNA-binding properties of GouR were tested it showed that GouR could bind to its own upstream region and that of *gouL* and *gouM*. The *gouM* gene has its own promoter, however, the genes *gouL – gouB* form a single operon. GouM is a major facilitator superfamily transporter, thought to be responsible for the efflux of gougerotin, with *gouL – gouF* and *gouB* being structural genes (Niu *et al.,* 2013; Wei *et al.,* 2014). Further examination of the regulatory function of GouR demonstrated that GouR activated transcription of *gouM*, but repressed that of the *gouL* operon, thus ensuring the intracellular concentration of gougerotin does not reach fatal levels (Wei *et al.,* 2014).

A final example of TetR family regulators is ClmR1, which plays a role in the inhibition collismycin A production (Vior *et al.,* 2014). Two transcriptional regulators are associated with the collismycin A biosynthetic cluster, ClmR1, a TetR-like protein, and ClmR2, a LuxR-type regulator (Garcia *et al.,* 2012). The negative role *clmR1* plays in collismycin A production was demonstrated by gene disruption that resulted in a distinct increase in collismycin A (Vior *et al.,* 2014). Disruption of *clmR2* however abolished collismycin A production, which was due to failed expression of structural genes, *clmM* and *clmL*, as well as other genes within the biosynthetic cluster (Garcia *et al.,* 2012; Vior *et al.,* 2014). One of the most

interesting aspects of collismycin A regulation is the role played by iron. Iron can interact with collismycin A; furthermore, high concentrations of iron can prevent the transcription of collismycin A biosynthetic cluster genes. Conversely iron positively regulates the expression of *clmR1*, resulting in ClmR1-associated repression of *clmR2*, thereby providing a mechanism of control that is mediated by iron levels (Vior *et al.,* 2014).

1.3.4 The LuxR family of regulators

The LuxR family of regulators share a common HTH motif DNA-binding domain within the C-terminal and members of this family include regulators of two component sensory transduction systems, acylhomoserine lactone-dependent regulators engaged in quorum sensing and regulators involved in antibiotic synthesis pathways (De Schrijver & De Mot 1999; Santos *et al.,* 2012). The LAL (LuxR) family of regulators are generally larger than most other LuxR proteins, being over 250 amino acids long and contain a nucleotide triphosphate (NTP) binding motif at the N-terminal as well as the usual the C-terminal motif (De Schrijver & De Mot 1999; Guerra *et al.,* 2012; Santos *et al.,* 2012). PikD is a LAL (LuxR) protein and is the sole pathway specific regulator of the pikromycin biosynthesis cluster (Wilson *et al.,* 2001). The expression of the *pikAI* operon, containing the PKS genes, was decreased in a *pikD* mutant, as was the expression of *pikD* itself, however PikD represses the expression of the *pikB* operon. Another bioactive compound that is associated with LAL proteins is rapamycin in which a positive regulator, RapH, has been identified and shares homology to other LAL proteins (Kuščer *et al.,* 2007). Rapamycin is produced by *S. rapamycinicus* (previously classified as *S. hygroscopicus*) and is positively regulated by RapG and RapH (Kuščer *et al.,* 2007; Baranasic *et al.,* 2013). It has been shown that *rapG* is absolutely required for rapamycin synthesis, while *rapH* plays a supportive role (Kuščer *et al.,* 2007). Complementation of a *rapG* and *rapH* mutant with *rapG* alone, only restores rapamycin production by around 10 % compared to the wild type (Kuščer *et al.,* 2007). Further to

this, overexpression of both *rapG* and *rapH* improves rapamycin production significantly, further highlighting the essential role of both RapH and RapG in rapamycin biosynthesis (Kuščer *et al.,* 2007; Park *et al.,* 2010),

As well as the LAL (LuxR) proteins there is also a family of proteins that contain an N-terminal PAS domain with the canonical C-terminal LuxR HTH domain. The majority of PAS domains bind to cofactors or ligands that are thought to be the sensory input required for activating these sensory domains (Hefti *et al.,* 2004). The PAS-LuxR transcriptional regulators are a common feature of Type I PKS biosynthetic clusters of the polyene macrolide family with the most notable example being PimM of the pimaricin cluster (Anton *et al.,* 2007; Santos-Aberturas *et al.,* 2011a). Within the pimaricin biosynthetic gene cluster there are two regulators, *pimM* and *pimR*, with PimR acting as the master regulator of *pimM* (Antón *et al.,* 2004; Santos-Aberturas *et al.,* 2012). PimM controls the expression of the PKS genes (*pimS1 – S4*), modification genes (*pimA, pimB, pimF, pimG, pimI – K*) and pimaricin initiation-signaling gene (*pimE*), by directly binding their promoters (Aparicio *et al.,* 2000; Mendes *et al.,* 2007a; Santos-Aberturas *et al.,* 2011b). It appears that PimM positively regulates all of these genes with the exception of *pimE*, in which PimM most likely acts as a modulator of expression, as deletion of *pimM* does not change *pimE* expression when compared to the wild type strain (Anton *et al.,* 2007). As PimE is an essential signal for pimaricin production, it is unlikely PimM is the activator. However, PimM may ensure either continued expression or repression when production of pimaricin needs to be limited (Mendes *et al.,* 2007a). Regardless of this, it appears that the role played by PimM is not as simple as mere activation.

Finally there is the N-terminal receiver (REC) domain class of LuxR regulators, of which OtcG is a member. A well characterized example is DnrN from *S. peucetius*, which is one of four regulators associated with the daunorubicin/doxorubicin (DNR/DXR) cluster (Vasanthakumar *et al.,* 2013). DnrN activates the master regulator, *dnrI* by binding directly to its promoter and can also work in conjunction with DnrO to further enhance *dnrI*

expression (Furuya & Hutchinson 1996; Jiang & Hutchinson 2006). *dnrN* itself is regulated by DnrO as well as the intracellular concentrations of DNR that results in a feedback loop in which the accumulation of DNR within the cell inhibits DNR/DXR production (Ajithkumar & Prasad 2010). Interestingly there is a REC-LuxR transcriptional regulator associated with undecylprodigiosin (RED) production in *S. coelicolor*, whose role is analogous to DnrN (Guthrie *et al.,* 1998). This regulator, RedZ, is required for the transcription of *redD*, a gene that encodes the transcriptional regulator RedD which is thought to activate the RED biosynthetic gene cluster *(Fujii et al.,* 1996; White & Bibb 1997). Unlike DnrN, RedZ production is negatively autoregulated, and RedZ activity is modulated by the product of the biosynthetic pathway, RED. However, it seems unlikely that this altered activity is due to competition at DNA-binding sites but by interaction between RedZ and Red (Wang *et al.,* 2009a). It is interesting that OtcG is a REC-LuxR protein that positively regulates OTC production and it is possible that it functions in a similar way to RedZ and DnrN although a cognate regulator has yet to be identified.

This brief overview of the different types of pathway-specific regulators gives an idea of how the varied and complex regulation of secondary metabolite clusters operates. However there is an even deeper level of regulation as activation of pathway specific regulators generally occurs by the action of global regulators. The following gives a brief overview of global signals that control these pathway specific regulators.

1.4 Global regulation of antibiotic production

1.4.1 The role of guanosine tetraphosphate and pentaphosphate

Guanosine tetraphosphate (ppGpp) and pentaphosphate (pppGpp) or collectively (p)ppGpp, are highly phosphorylated guanine nucleotides that are accumulated rapidly in bacterial and plant cells during stress conditions such

as nutrient starvation (Chakraburtty & Bibb 1997; Srivatsan & Wang 2008). (p)ppGpp is synthesised in *Escherichia coli* by RelA when the protein is bound to ribosomes and when an uncharged transfer RNA (tRNA) binds to the acceptor site of the ribosome, RelA is activated (Gropp *et al.,* 2001; Srivatsan & Wang 2008; Sajish *et al.,* 2009). As there is no amino acid attached to the tRNA protein synthesis stops, allowing the formation of (p)ppGpp from guanosine triphosphate (GTP) in conjunction with adenosine triphosphate (ATP) to occur. This leads to the inhibition of tRNA and ribosomal RNA (rRNA) synthesis by the action of (p)ppGpp. In response to (p)ppGpp generation growth stops and survival/adaptive mechanisms are initiated and because of this (p)ppGpp's role in secondary metabolite production has been examined (Jones 1994; Jones & Bibb 1996; Hesketh *et al.,* 2007; Srivatsan & Wang 2008). The ppGpp synthetase of *S. coelicolor* A3(2), RelA, is necessary for ACT and RED production in conditions when nitrogen is limiting (Chakraburtty & Bibb 1997). A study performed using a thiostrepton-inducible promoter to activate transcription of N-terminal portions of the *relA* gene in *S. coelicolor* A3(2) demonstrated that ppGpp could initiate the transcription of *actII-ORF4* even in optimal growth conditions *(Hesketh et al.,* 2001). This suggests that ppGpp alone, rather than in combination with the regulatory mechanisms governing the erection of aerial hyphae, growth cessation or other factors, has some role to play in ACT production by activating transcription of *actII-ORF4*. Though whether this is by the direct action of ppGpp with the *actII-ORF4* promoter or by triggering another protein has yet to be determined, despite this it does appear that ppGpp is involved in antibiotic production (Hesketh *et al.,* 2001; Hesketh *et al.,* 2007).

1.4.2 The role of γ-butyrolactone

γ-butyrolactones are signalling molecules produced by *Streptomyces* that are able to regulate antibiotic production and some are also able to influence differentiation (Horinouchi *et al.,* 2001; Takano 2006). γ-butyrolactones bind to highly specific receptor proteins and in doing so

release these proteins from their particular DNA-binding site, located in the promoter of the genes subject to regulation by the receptor protein. Binding of the γ-butyrolactone to the receptor protein removes the repression (Takano 2006). A-factor from *S. griseus* is the best-characterised γ-butyrolactone and is required for activation of streptomycin production and aerial hyphae formation; with its expression being is highest during the middle of the exponential growth phase (Ohnishi *et al.,* 2005; Hong *et al.,* 2007). By expressing *asfA* in *E. coli*, researchers were able to demonstrate that AsfA alone could synthesise A-factor. Initiation of streptomycin production due to the action of A-factor is achieved by release of ArpA from the promoter of *adpA*. This permits expression of AdpA and the consequent activation of *strR*, the pathway-specific regulatory gene of the streptomycin gene cluster (Ohnuki *et al.,* 1985; Hirano *et al.,* 2008). AdpA also activates the expression of other genes within the *adpA* regulon that are involved aerial hyphae formation (Liu *et al.,* 2013).

1.4.3 The role of carbon

Carbon source can have a substantial influence on antibiotic production with glucose, glycerol and mannitol blocking the production of ACT in *S. coelicolor*, actinomycin, in *S. parvulus* and streptomycin, in *S. griseus* (Demain & Inamine 1970; Sánchez *et al.,* 2010; van Wezel & McDowall 2011). It is thought that carbon catabolite repression (CCR), the phenomenon that prevents the utilisation of other carbon sources in the presence of glucose, is linked with this repression (Hodgson 1982; Kwakman & Postma 1994; Saito *et al.,* 1998). ACT production in *S. lividans* is also blocked in the presence of glucose and it is thought to be due to changes in the expression of *afsR2* (*afsS*) (Kim *et al.,* 2001). AfsR2 functions as a global regulator of antibiotic production, specifically ACT and RED in *S. coelicolor*, with disruption of *afsR2* resulting in diminished ACT synthesis (Tanaka *et al.,* 2007; Liu *et al.,* 2013). Growth of *S. lividans* on glucose showed a reduction in the expression of *afsR2* and it seems likely that this

repression by glucose is the cause of failed ACT production for that carbon source (Kim *et al.,* 2001). A similar effect is seen with phenoxazinone synthase (PHS), an enzyme associated with the synthesis of actinomycin production in *S. parvulus*, as its expression is inhibited by the presence of glucose in the culture (Jones 1985).

The carbon and nitrogen source N-acetylglucosamine (GlcNAc) can modulate antibiotic production depending on whether it is present in rich or nutritionally poor media, by blocking or activating production, respectively (Rigali *et al.,* 2008). In the natural environment GlcNAc would become available during the breakdown of substrate mycelium, which could then be imported back into the cell via the phosphotransferase system (PTS) (Nothaft *et al.,* 2010; Liu *et al.,* 2013). During this process, GlcNAc is phosphorylated, before subsequent deacetylation by NagA, which results in the production of glucosamine-6-phosphate (GlcN-6-P). The inhibition of antibiotic production due to GlcNAc has been linked to the GntR transcriptional regulator, DasR. This protein binds directly to the promoters of *actII-ORF4* and *redZ*, thus preventing the activation of these specific antibiotic biosynthetic clusters (Rigali *et al.,* 2008). Relief from this repression is achieved by GlcN-6-P inhibiting the DNA-binding capabilities of DasR. Presumably this due to the activity of DasR acting as an effector molecule, therefore allowing for normal transcription to occur (Rigali *et al.,* 2006; Rigali *et al.,* 2008; Nothaft *et al.,* 2010).

1.4.4 The role of nitrogen

Carbon control over antibiotic production has been well established, however there is an increasing body of evidence to support nitrogen levels influencing antibiotic production (Shu *et al.,* 2008; Yang *et al.,* 2008; Wang *et al.,* 2012b; Yu *et al.,* 2012c). There are at least three proteins found to effect antibiotic production in *Streptomyces* with relation to nitrogen conditions, DraR, AfsQ1 and NdgR, and of these AfsQ1 is the best characterised (Liu *et al.,* 2013). AfsQ1 is part of a two component system (TCS) with AfsQ2 and

stimulates ACT production in *S. lividans* when grown in complex media; however, no such stimulus has been demonstrated for *S. coelicolor* (Ishizuka *et al.,* 1992). Disruption of *afsQ1-Q2* in *S. coelicolor* decreases ACT production when grown on minimal media supplemented with glutamate as the sole nitrogen source (Shu *et al.,* 2008). AfsQ1 can bind and activate the expression of *actII-ORF4, redZ* and *cdaR*, pathway-specific regulators of the antibiotics ACT, RED and calcium-dependent antibiotic (CDA), respectively (Wang *et al.,* 2012b). Further to this, AfsQ1 inhibits genes associated with nitrogen assimilation; thereby, competing with GlnR, an orphan response regulator that is a key regulator of nitrogen assimilation, for the repression and activation of such genes, respectively (Wang *et al.,* 2012b).

DraR, like AfsQ1, forms part of a TCS with DraK, and is linked with both increase and decrease in antibiotic production in relation to nitrogen levels (Yu *et al.,* 2012c). DraR-K functions as an activator of ACT production and as a repressor of yellow-pigmented Type I polyketide (yCPK) production, via direct binding of the pathway-specific regulators (Yu *et al.,* 2012c). RED production is also decreased by the action of DraR-K but this is independent of any pathway-specific regulators. Once again the effect on ACT and RED production is only exerted when *S. coelicolor* mutants of *draR-K* were grown on minimal media supplemented with high concentrations of nitrogen (Yu *et al.,* 2012c). As with DraR-K, NdgR controls ACT production in a way that is dependent on nitrogen source; although, in a negative manner rather than positive (Yang *et al.,* 2008). It is not only antibiotic production in *S. coelicolor* that is affected by different nitrogen sources as validamycin A production in *S. hygroscopicus* is lowered with the introduction of different nitrogen sources to the fermentation media (Wei *et al.,* 2012).

1.4.5 The role of phosphate

The presence of phosphate in the growth environment can repress the synthesis of some secondary metabolites such as streptomycin, tylosin, OTC, clavulanic acid and cephamycin C (Martín 2004). A two component

regulatory system, the PhoR-PhoP system, has been explored to deduce its role in antibiotic production as this system reacts to phosphate levels (Sola-Landa *et al.,* 2003; Mendes *et al.,* 2007b). PhoR is a membrane sensory kinase that is activated by autophosphorylation. The activating phosphate group is then transferred to PhoP, a DNA-binding response regulator that is a member of the OmpR family (Sola-Landa *et al.,* 2003; Allenby *et al.,* 2012). In *S. coelicolor* and *S. lividans,* antibiotic production is mediated by the PhoR-PhoP system and deletion of *phoP* and/or *phoR*, in *S. lividans* results in an overproduction of ACT and RED (Sola-Landa *et al.,* 2003; Martín & Liras 2010). In *S. natalensis* the PhoR-PhoP system influences the production of pimaricin, as at least four genes in the *pim* biosynthetic cluster are repressed when both *phoR* and *phoP* are expressed (Mendes *et al.,* 2007b). The PhoR-PhoP system does not only affect antibiotic biosynthesis, but also the expression of other genes, such as the alkaline phosphatase *phoA*. As a result, phosphorylated PhoP may play an indirect or direct role on antibiotic production (Martín 2004; Bibb 2005). Further to this, it appears that intracellular phosphate may act as an inhibitor of antibiotic synthesis itself as high levels of phosphate could still inhibit antibiotic production in *phoPR* mutants (Liu *et al.,* 2013).

1.4.6 *Streptomyces* development in relation to antibiotic production

Streptomyces have a unique developmental life cycle (Figure 1.9) that is linked closely to secondary metabolism. *Streptomyces* produce spores that germinate to produce substrate mycelium (Figure 1.9, 1), followed by the erection of aerial mycelium (Figure 1.9, 2), aerial mycelium then septates to produce prespores (Figure 1.9, 3) that are released to produce spores (Figure 1.9, 4) and thus allowing the developmental cycle to repeat (Wildermuth 1970; Elliot *et al.,* 2007). For most *Streptomyces* species this life cycle (Figure 1.9) does not occur in liquid cultures (Yagüe *et al.,* 2013) and therefore spores are not produced. Those *Streptomyces* species that don't produce aerial hyphae or spores in liquid culture produce so-called

second multinucleated mycelium, which is the precursor of aerial mycelium on solid cultures (Manteca *et al.,* 2008; Yagüe *et al.,* 2013). Spores are not only robust, and therefore can survive longer in the environment than substrate mycelium, but are also thought to help in the dispersal and spread of *Streptomyces* (Chater 2006; Elliot *et al.,* 2007). Once a spore encounters suitable growth conditions it begins to germinate, with hyphae emerging from the apex of the spore and growing by tip extension (McCormick & Flardh 2011; Ahmed *et al.,* 2012). From the substrate mycelium aerial hyphae is formed, however some of the substrate mycelium undergoes a form of programmed cell death, which releases nutrients which can be utilised by the growing hyphae (Elliot *et al.,* 2007; McCormick & Flardh 2011). Production of aerial hyphae requires the *bld* genes, named for the bald appearance of *Streptomyces* colonies that lack these genes (Claessen *et al.,* 2006). In contrast the *whi* genes are required for aerial hyphae to differentiate into spores (Chater 2001).

Around the time that aerial hyphae are forming, secondary metabolites are produced (Ahmed *et al.,* 2012). In addition, liquid cultures of *Streptomyces* do not produce antibiotics until the second multinucleated mycelium are present, suggesting that it is this mycelium stage that is responsible for antibiotic production (Manteca *et al.,* 2008). Some of the previously mentioned global regulators of antibiotic production also play a role in morphology, such as (p)ppGpp, AdpA and nutrient sources (Sun *et al.,* 2001; Ohnishi *et al.,* 2005; Rigali *et al.,* 2006; Yu *et al.,* 2012), another common link between antibiotic production and morphology. However the relationship between morphology and antibiotic production is further exemplified by the requirement of certain *bld* genes for antibiotic synthesis to occur (Elliot *et al.,* 2003; Elliot *et al.,* 2007). The majority of the *bld* and *whi* genes encode known or putative regulators and sigma factors, as well as a tRNA (Chater 2001; Claessen *et al.,* 2006). Of these 3 *bld* genes will be discussed further in relation to their association with antibiotic production, as the *bld* genes have a greater influence on antibiotic production when compared to the *whi* genes.

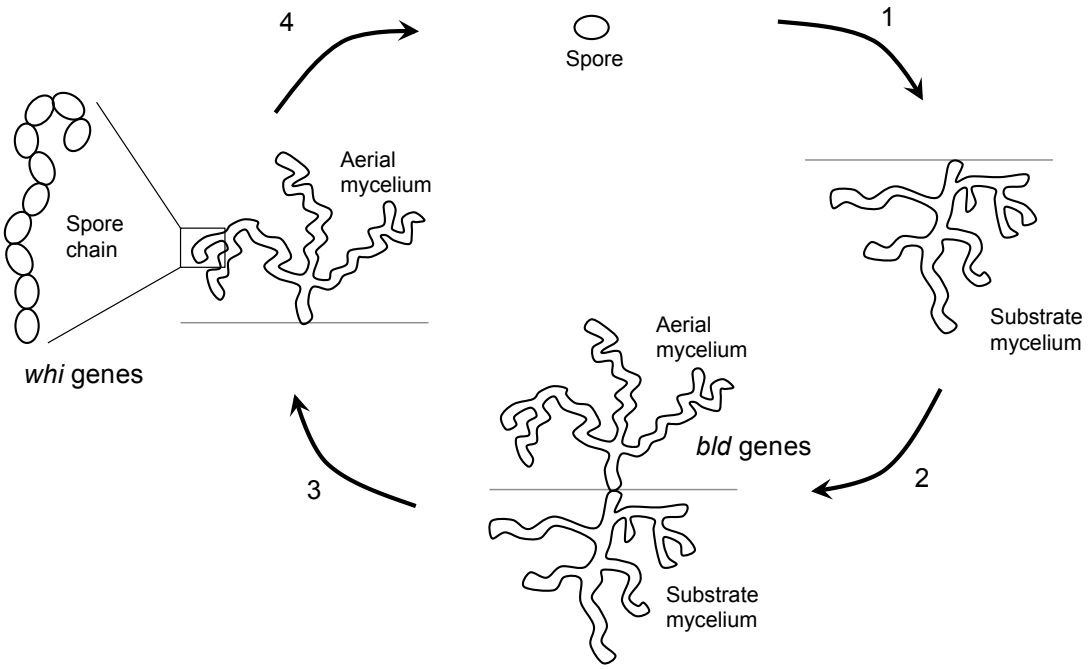


Figure 1.9 *Streptomyces* developmental life cycle. 1. A spore starts to germinate and produces substrate mycelium. 2. Substrate mycelium starts to differentiate into aerial mycelium. 3. Aerial mycelium starts to septate, producing spore chains. 4. Spores are released from the spore chains, allowing the cycle to continue.

bldC encodes a small DNA-binding protein and is required for aerial hyphae formation, which can be restored if grown on minimal media supplemented with either mannitol, galactose or maltose (Merrick 1976; Hunt *et al.,* 2005). *bldC* null mutants were found to be attenuated in ACT production when grown on R2YE media, however RED production was not severely affected. Further analysis demonstrated that *actII-ORF4* transcription was delayed and reduced for *bldC* mutants when compared to the wild type strain, suggesting *actII-ORF4* transcription is dependent on BldC (Hunt *et al.,* 2005). As with BldC, BldD is linked to antibiotic production. BldD is a regulatory protein that negatively regulates the expression of its own gene, *bldD*, but also the expression of other developmental associated genes, *bldN, bldM* and *whiG* (Elliot *et al.,* 1998; Elliot *et al.,* 2001; Hengst *et al.,* 2010). In addition chromatin immunoprecipitation-microarray analysis (ChIP-chip) of the BldD regulon identified *bldC* and *bldA* as targets (Hengst *et al.,* 2010), and it is possible the affect on antibiotic production in *bldD* mutants is due to BldD influencing these genes. However expression of two genes associated with biosynthesis of branched-chain amino acids, *leuB* and *leuC*, was increased in a *bldD* mutant (Hengst *et al.,* 2010). It has been shown that catabolism of branched-chain amino acids supplies 50 % of the acetate required for ACT synthesis (Stirrett *et al.,* 2008). It is possible another means of influencing ACT production by BldD is decreasing the production of branched-chain amino acids therefore allowing acetate to flow directly into ACT production. As *bldA* is a target of BldD, it is possible the role BldD plays in antibiotic production is due to a as yet unknown influence BldD has on *bldA*. The *bldA* gene encodes a leucine tRNA with a TTA (UUA) codon, which is rare in *Streptomyces* genomes due to the high GC content of these organisms (White & Bibb 1997). The expression of genes containing the TTA codon are dependent on the expression of *bldA*, and for this reason no essential genes contain the rare TTA codon (Hengst *et al.,* 2010; Liu *et al.,* 2013). The presence of the TTA codon in genes provides an extra level of regulation due to the requirement of BldA, with *bldA* expression upregulated as the *Streptomyces* bacterium's growth begins to slow and

enter stationary phase (Liu *et al.,* 2013). For this reason *bldA* is associated with antibiotic production and development, as *bldA* mutants are defective in both (Chater & Chandra 2008). In *S. coelicolor* A3(2) antibiotic production is disrupted in *bldA* mutants due to the presence of the TTA codon in the pathway specific regulators *actII-ORF4* and *redZ* (Liu *et al.,* 2013). Further to this, substitution of the TTA codon in *actII-ORF4* with another leucine-associated codon, TTG, could restore actinorhodin (ACT) production in *bldA* mutants (Fernández- Moreno *et al.,* 1991). This highlights the important role *bldA* plays in antibiotic production.

Antibiotics can be regulated by numerous means as seen with ACT which is influenced by (p)ppGpp, SARPs, carbon, nitrogen and phosphate. Therefore it is likely that most antibiotic biosynthetic pathways are under the control of more than one regulator or that the regulator of the pathway has multiple activators.

1.5 Aims of the project

Oxytetracycline (OTC) is produced by *S. rimosus* and there are currently four laboratory strains (Figure 1.10) available to us that underwent strain improvement by Pfizer. These strains include the original soil isolate *S. rimosus* G7, along with three others produced during a strain improvement program, *S. rimosus* M4018, 15883S and 23383 (Figure 1.10). As mentioned previously *S. rimosus* G7 (ATCC 10970) was the original soil isolate. *S. rimosus* has been of interest to various groups around the world, however it was primarily groups in Croatia and the United Kingdom who worked with *S. rimosus* G7, also known as R7 (Petković *et al.,* 2006). The groups in Croatia were working, since the early 1960s, with mutants isolated not only from *S. rimosus* G7 but also another soil isolate, called *S. rimosus* R6, which was found outside the University of Zagreb (Petković *et al.,* 2006). The aim of their work was to generate strains for the commercial production of OTC for the pharmaceutical company PLIVA. It was during the 1970s that

S. rimosus M4018 first appeared, a prototrophic strain used by Pfizer commericially for the production of OTC. Much of this work was carried out by scientists from Pfizer Central Research, in collaboration John Innes Centre and latterly with the University of Glasgow (Friend & Hopwood 1971; Butler *et al.,* 1989). During the strain improvement process *S. rimosus* 15883 was generated, derived from *S. rimosus* M4018 (McDowall *et al.,* 1999), with the ability to hydrolyse sucrose. By the early 1980s a *S. rimosus* strain was produced that was restriction minus and optimized for growth on oil, this strain derived from 15883 was *S. rimosus* 23383 (Kirby *et al.,* 2008). This strain was also 2-deoxyglucose (2-DOG) resistant, as the strain preceding it, *S. rimosus* 17011, had been selected for 2-DOG resistance.

The genomes of the three production strains were sequenced previously and an initial aim of this work was to sequence the genome of *S. rimosus* G7. The aim was to use the G7 sequence as a reference for the three other strains, providing an opportunity to explore the genomic basis behind increased OTC productivity. This was achieved by detecting genetic differences in the form of multiple gene deletions or mutations, such as single nucleotide changes and insertions and deletions. As improved fermentation procedures have not increased G7's OTC productivity to that of 23383, it can be hypothesized that these genomic differences have lead to increased OTC productivity. In addition, not all of the genomic differences between G7 and the production strains will be associated with increased OTC productivity, some will be coincidental. For this reason, it was necessary to select specific genomic differences between G7 and the production strains that would be reproduced in G7. In this way, genomic differences can be linked with increased OTC production. It is also possible to combine genetic manipulation with metabolomics to provide further support for the role played by a genomic difference in improved OTC production.

Therefore the ultimate aim of this project is to use our genomic data to identify those genomic differences that lead to increased OTC productivity, confirm them genetically and use this information to inform rapid improvement of other strains.

Throughout this work the strains of *S. rimosus* will be referred to using a logical nomenclature in reference to the potential OTC production per litre of each strain (Figure 1.10). Therefore *S. rimosus* G7, M4018, 15883S (a OTC non-producer derived from 15883) and 23383 shall be referred to as SR1, SR20, SR50S and SR55 as each are proposed to be able to produce around 1 g/L, 20 g/L, 50 g/L, in reference to 15883, and 55 g/L, respectively.

Standard Strain Name	Strain Name used in Thesis	Oxytetracycline Produced
G7/ATCC 10970	SR1	~1 g/L
M4018	SR20	~20 – 30 g/L
15883	SR50	~50 g/L
15883S	SR50S	Non-producer
23383	SR55	~55 g/L

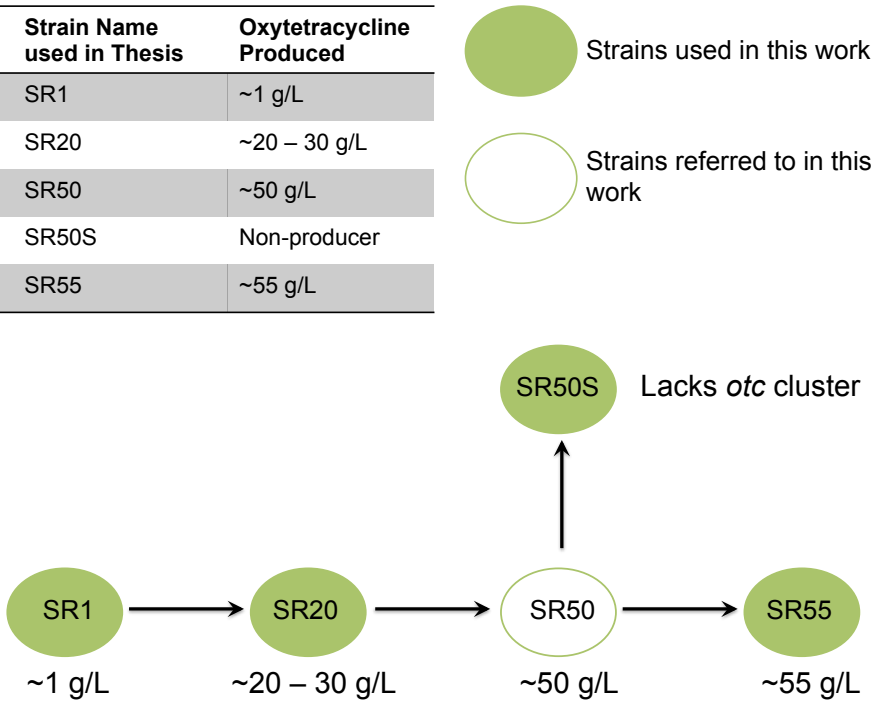


Figure 1.10 Schematic for the origin and parentage of *S. rimosus* strains. The arrows indicate the direction in which the production lineage progresses. The table displays the standard strain name referred to in the literature for each strain within the strain improvement lineage, along with the names referred to in this thesis and the potential OTC produced per litre for each strain.

Chapter 2

2 Material and Methods

2.1 Bacterial Strains and Vectors

Table 2.1 contains a list of all *S. rimosus* strains used in this work. For immediate use of a strain, spore suspensions were streaked onto Emerson's agar for single colonies, to ensure OTC productivity. Liquid cultures were then inoculated. Spore suspensions were produced by inoculating Mannitol-Soy (MS) agar (Keiser *et al.,* 2000) for confluent growth and plates left for 5 – 7 day, to allow for spore formation. Spores were recovered by using 5 mL of sterile 20% (v/v) glycerol applied directly to the plate and a sterile cotton bud. For long term storage these spore suspensions were stored at -20 °C.

Table 2.1 *S. rimosus* strains used in this work

Strain	Description	Reference
G7 (SR1)	OTC^+ RIM^+	ATCC 10970
M4018 (SR20)	OTC^+ RIM^-	(Rhodes *et al.,* 1981)
15883S (SR50S)	OTC^- RIM^-	(McDowall *et al.,* 1999)
23383 (SR55)	OTC^+ RIM^-	(Kirby *et al.,* 2008)
ACM23	OTC^+ RIM^-	This study

A list of all *E. coli* strains used in this work are detailed in Table 2.2. For direct use of a strain, glycerol stocks were streaked onto Luria-Bertani (LB) agar, containing the relevant antibiotics when necessary, for single colonies. From fresh overnight cultures, glycerol stocks were produced by mixing 1.5 mL of the overnight culture with 0.2 mL of sterile 80 % (v/v) glycerol. These glycerol stocks were used for long term storage at -80 °C.

Table 2.2 *E. coli* strains used in this work

Strain	Description	Reference
DH5-α	*sup*E44, Δ*lac*U169, (φ80 *lacZ*Δ M15), *hsd*R17, *rec*A1, *end*A1, *gyr*A96, *thi*-1, *rel*A1	(Sambrook *et al.,* 1989)
BL21(DE3) (pLysS)	*F⁻ ompT hsdS*$_B$($r_B^-m_B^-$) *gal dcm* (DE3), Cm^r	(Studier *et al.,* 1990)
GM2929	*F⁻ dam-13::Tn9, dcm-*6*, hsd*R2*, rec*F143*, galK2, gal*T22*, ara*-14*, lacYI, xyl-5, leu*B6*, thi-I, tonA31, rpsLI36, hisG4, tsx-78, mtl-I, ginV44*	(Binnie *et al.,* 1989)
BW25113	*lacI^q, rrnB$_{T14}$ ΔlacZ$_{WJ16}$, hsdR514 ΔaraBA-D$_{AH33}$, ΔrhaBAD$_{LD78}$*	(Datsenko & Wanner 2000)
ET12567	*dam, dcm, hsdS, Cm^r*	(MacNeil *et al.,* 1992)

Table 2.3 Plasmids used in this work

Plasmid	Description	Reference
pUC19	*bla*, *lacZα*	(Ebright *et al.,* 1985)
pGB1	*aac(C)1, tsr*	(Herron, Unpublished Data)
pGB02	*aac(C)1, tsr,* Rim02 insert between KpnI and EcoRI sites	This study
pGB23	*aac(C)1, tsr,* Rim02 insert, Rim03 insert between HindIII and PstI	This study
pET28a	Expression vector, HIS-tag	Novagen
pOtcR-SR1	pET28a expressing *otcR* from SR1	This study
pOtcR-SR55	pET28a expressing *otcR* from SR55	This study
pGLW207	Right side of *otc* cluster in pSuperCos 1	(Almutairi, 2007)
pIJ773	pKD20 with *aac(3)IV*	(Gust *et al.,* 2003)
pIJ790	λ Red plasmid, Cm^r, temperature sensitive growth (30 °C)	(Gust *et al.,* 2003)
pUZ8002	$OriT^-$ RP4 derivative, used for conjugation	(Kieser *et al.,* 2000)
pOtcR-KO	pGLW207 with *otcR* replaced with *aac(3)IV* cassette	This study
pET21a	Expression vector, no tag	Novagen
pOtcR21-SR1	pET21a expressing *otcR* from SR1	This study

2.2 Suppliers

Suppliers used for all reagents and chemicals were Melford, Sigma-Aldrich or Fisher, unless otherwise stated. Enzymes were purchased from New England Biolabs, Invitrogen, Fisher or Promega.

2.3 Media and Solutions

2.3.1 Media

All media used for growth of bacterial strains were prepared by combining the relevant reagents, as listed below, followed by autoclaving at 121 °C for 15 minutes (unless otherwise stated). For the preparation of agar plates, approximately 25 mL of molten agar was poured into each Petri dish and allowed to set. Plates were poured and set within a laminar flow hood for around 20 minutes, to maintain sterility and allow any excess liquid to evaporate.

2.3.1.1 Media for *Streptomyces rimosus*

Mannitol Soya (MS)

Agar	20 g
Mannitol	20 g
Soya flour	20 g
Tap water	1000 mL

Autoclaved at 121 °C for 30 minutes

CRM

Glucose	10 g
Sucrose	103 g
$MgCl_2x6H_2O$	10.12 g
Tryptone soya broth	15 g
Yeast extract	5 g
Distilled water	1000 mL

Emerson's Agar

Beef extract	4 g
Yeast extract	1 g
Peptone	4 g
Glucose	10 g
NaCl	2.5 g
Agar	10 g
Distilled water	1000 mL

Liquid Emerson's

As Emerson's agar but without the addition of 10 g agar.

Tryptone soya agar (TSA)

Tryptone (Casein Digest USP)	17 g
Soy peptone	3 g
Sodium chloride	5 g
Dipotassium phosphate	2.5 g
Dextrose	2.5 g
Agar	15 g
Distilled water	1000 mL

Tryptone soya broth (TSB)

As tryptone soya agar but without the addition of 15 g agar.

2.3.1.2 Media for *Escherichia coli*

LB Agar

Tryptone	10 g
Yeast extract	5 g
NaCl	10 g
Agar	15 g
Distilled ater	1000 mL

LB Broth

As LB agar but without the addition of 15 g agar.

2x YT Broth

Tryptone	16 g
Yeast extract	10 g
NaCl	5 g
Distilled water	1000 mL

2.3.1.3 Miscellaneous Media

Soft Nutrient Agar (SNA)

Nutrient broth	8 g
Agar	5 g
Distilled water	1000 mL

Yeast Peptone Glucose (YPG) Agar

Yeast extract	10 g
Peptone	20 g
Glucose	20 g
Agar	20 g
Distilled water	1000 mL

YPD Broth

As YPD agar but without the addition of 20 g agar.

2.3.2 Antibiotics

Below is a list of antibiotics and the final concentration used, in either liquid or solid media, when required for *S. rimosus* or *E. coli*.

Table 2.4 Antibiotic concentrations used for the growth and maintenance of *S. rimosus* and *E. coli*

Antibiotic	*S. rimosus*	*E. coli*
Apramycin	400 µg/mL	100 µg/mL
Gentamicin	400 µg/mL	10 µg/mL
Carbenicillin	N/A	100 µg/mL
Kanamycin	600 µg/mL	50 µg/mL
Thiostrepton	50 µg/mL	N/A
Chloramphenicol	N/A	25 µg/mL
Nalidixic acid	N/A	25 µg/mL
Tetracycline	N/A	15 µg/mL

2.3.3 Solutions

2.3.3.1 Solutions associated with DNA isolation and manipulation

TAE buffer (50x)

Tris	242 g
0.5 M EDTA (pH 8)	100 mL
Acetic acid	57.1 mL
Distillied water	842.9 mL

6x bromophenol blue loading dye

Glycerol	375 mL
Bromophenol Blue	0.2 g
Distilled water	1000 mL

TAP buffer (2x)

Tris-acetate (pH 7.9)	597.93 g
Potassium acetate	323.90 g
Magnesium acetate	142.40 g
Ammonium acetate	107.92 g
Dithiothretol (DTT)	30.85 g
PEG 6000	400 g
Distilled water	1000 mL

Phenol:chloroform solution

Chloroform	96 mL
Isoamyl alcohol	4 mL
Phenol	100 mL
8-hydroxy-quinoline	0.1 g

The chloroform was combined with isoamyl alcohol and mixed before being added to the phenol. 8-hydroxy-quinoline was added to phenol to give a final concentration of 10 mg/mL. The phenol:chloroform mixture was overlayed with 5 mL of 50 mM Tris.Cl pH 8.0.

2.3.3.2 Buffers for use with polyacrylamide gels

Resolving buffer (4x)

Tris-HCl (pH 8.8)	181.71 g
SDS	4 g
Distilled water	1000 mL

Stacking buffer (4x)

Tris-HCl (pH 6.8)	60.57 g
SDS	4 g
Distilled water	1000 mL

Running Buffer

Glycine	14.40 g
Tris	3.03 g
SDS	1 g
Distilled water	1000 mL

SDS-PAGE loading dye

Tris	7.63 g
SDS	20 g
Glycerol	100 mL
β-Mercaptoethanol	50 mL
Bromophenol blue	0.1 g
Distilled water	850 mL

SDS-PAGE stain

Ethanol	415 mL
Acetic acid	165 mL
Coomassie blue	1 g
Distilled water	420 mL

MADAM's (5x)

Tris	3.03 g
Glycine	30.03 g
Distilled water	1000 mL

Adjust to pH 8.6 using acetic acid

2.3.3.3 Buffers for OtcR purification

Resuspension buffer

Tris (pH 8.0)	12.11 g
NaCl	14.60 g
Imidazole	1.36 g
Distilled water	1000 mL

Soluble buffer

Tris (pH 8.0)	12.11 g
NaCl	14.60 g
Imidazole	1.36 g
SDS	10 g
1 mM β-mecaptoethanol	0.75 mL
Distilled water	1000 mL

OtcR buffer A

Tris	12.11 g
NaCl	14.60 g
Imidazole	3.40 g
Distilled water	1000 mL

Adjusted to pH 8.0 with HCl

OtcR buffer B

Tris	12.11 g
NaCl	14.60 g
Imidazole	68.07 g
Distilled water	1000 mL

Adjusted to pH 8.0 with HCl

OtcR storage buffer

Eluted solution with 200 µL of 80 % (v/v) sterile glycerol

2.3.3.4 Buffers used for *S. rimosus* crude lysate and biotin interaction experiments

Binding and washing buffer (2x)

Tris-HCl (pH 7.5)	1.21 g
EDTA	0.29 g
NaCl	116.88 g
Distilled water	1000 mL

TEG buffer

Tris-HCl (pH 7.5)	6.06 g
EDTA	0.15 g
Glycerol	50 mL
NaCl	2.92 g

2.4 Genomic and plasmid DNA purification

2.4.1 Genomic DNA isolation from *S. rimosus*

50 µL of *S. rimosus* spore suspension or a single colony from an Emerson's agar plate was inoculated in 50 mL of TSB and incubated for 30 hours at 30 °C in a shaker incubator (250 rpm). The culture was centrifuged at 4000 rpm (Sorvall SS-34 rotor, 107 mm) for 10 minutes. The pellet was resuspended in 3 mL of 25 mM Tris-HCl pH8 containing lysozyme (2 mg/mL) and RNaseA (4 mg/mL). The tube was then incubated for 30-60 minutes at 37 °C, followed by the addition of 3 mL of 0.3M NaCl with 2 % SDS (w/v), after which 5 mL of phenol solution was added. The solution was mixed by inverting three times. This was followed by centrifugation at 7000 rpm (Sorvall SS-34 rotor, 107 mm) for 5 minutes, the aqueous phase was transferred to a clean Oakridge tube. The process was repeated until the top aqueous layer became transparent. The aqueous phase was then transferred to a sterile 15 mL Falcon tube. 10 mL absolute ethanol was

added. The DNA was spooled out from the solution using a bent glass Pasteur pipette. The DNA was resuspended in 500 µL of sterile distilled water, separated into 50 µL aliquots and stored at -20 °C.

2.4.2 Plasmid DNA isolation from *E. coli*

Plasmid DNA isolation from E. coli was carried out from 5 mL overnight cultures using the alkaline lysis extraction method (Bimboim & Doly 1979) or via Wizard® miniprep plasmid DNA purification kit (Promega) by following the directions of the manufacturer. If plasmid DNA, isolated using the alkaline lysis extraction method, was used in restriction endonuclease digests, then RNAse A was added during the last 30 minutes of incubation to give a final concentration of 1 µg/ µL.

2.5 Methods for the manipulation of DNA

2.5.1 Restriction endonuclease digestion

Around 0.5 - 2 µg of DNA was digested with 10 units of enzyme in the appropriate buffer and 10 µg of bovine serum albumin (BSA), for 1.5 hours at the appropriate temperature.

2.5.2 Polymerase chain reaction (PCR)

PCR reactions were carried out in 10, 25 or 50 µL reactions using either MyTaq (Bioline) or Q5 (NEB) DNA polymerase according to the manufacturer's instructions. General thermocycling conditions are described in Table 2.5.

Table 2.5 General PCR conditions used.

Step	No. of Cycles	Temperature	Time (minutes)
1	1x Denaturation	98 °C	10
2	1x Denaturation	98 °C	1
3	1x Annealing	Ta	1
4	1x Extension	72 °C	1
5	25 x Step 2 – 4		
6	1x Final Extension	72 °C	10

Ta refers to the annealing temperature of oligonucleotide primer set as shown in Table 2.6

2.5.3 DNA clean up

DNA was purified using PCR cleanup kits (Bioline) by following the directions of the manufacturer, or by ethanol precipitation.

2.5.4 Agarose gel electrophoresis

Size determination of DNA was achieved by agarose gel electrophoresis, which was also utilised for purification of DNA. Between 0.7 and 1 % agarose was melted in 1x TAE in the microwave until the agarose dissolved. Once the melted agarose had cooled to around 50 °C, ethidium bromide was added, to allow for visualisation of DNA by UV light, to give a final concentration 1 µg/mL. The solution was then poured into a gel mould. DNA samples to be examined were combined with bromophenol blue loading dye and loaded into the wells of the gel, along with a DNA ladder, and separated on a gel by electrophoresis. Generally a voltage of 50 - 100 V was used dependent on the size of the DNA, and ran for 60 – 120 minutes.

2.5.5 DNA ligations by use of T4 DNA ligase

Ligation reactions were normally carried out overnight at 4 °C for blunt-ended reactions or for 2 hours at room temperature for sticky ended ligations. The appropriate buffer for the ligase was used, as directed by the manufacturer. The total ng of insert DNA to be ligated with vector DNA was generally a molar ratio of 5:1 insert to vector.

2.5.6 Sequencing of *S. rimosus* SR1 genomic DNA

Genome sequencing of *S. rimosus* SR1 was achieved by following the protocol for rapid library preparation for the GS Junior Titanium Series (Roche). 500 ng of genomic DNA was fragmented by nebulization using 30 psi (2.1 bar) of nitrogen for 1 minute. The nebulized DNA was purified using Qiagen MinElute PCR Purification kit, followed by end repair using the rapid library preparation end repair mix. After end repair the rapid library sequencing adaptors were ligated to the fragmented DNA, which was followed by small fragment removal using AMPure beads. At the point the quality of the library could be assessed using an Aglient Bioanalyzer with the high sensitivity DNA chip. The library was deemed to be sufficient quality if the average fragment size was between 600 bp and 900 bp, with less than 10 % below 350 bp. The library was then quantitated and diluted to stock of 1×10^7 molecules/µL before progressing to the emulsion PCR step. After this the prepared library could be sequenced.

2.6 Primers used in this work

The following table contains a list of PCR primers used in this work, including sequence and the annealing temperature (Ta) used for PCR.

Table 2.6 Primers used in this work.

Primer	Sequence (5' – 3')	Ta
Rim02F	CGGCGAATTCTCGTCTACGACTACCCGACA	62 °C
Rim02R	CGGGCCATGGGTGAGAACACTCCTGCCACA	62 °C
Rim03F	GCCGCTGCAGCACCCTCGTCTTCGACTACC	62 °C
Rim03R	GCCCAAGCTTGTGTAGCCGGTCTCGTTCAG	62 °C
RimKO-AF	CTGCGCCGTCTGGCCACCGTCC	68 °C
RimKO-AR	TCGTCGGCGAATTCCGCCGCGAT	68 °C
RimKO-BF	ATCTGCGGGTCAACGGCCTGC	68 °C
RimKO-BR	ACACCACTCCGGTGAACGCGAAC	68 °C
RimKO-CF	TCGTCTACGACTACCCGACA	68 °C
RimKO-CR	CACGCCGGCTCGCGATATCG	68 °C
RimKO-DF	CTGGTCGACAGTGACATCAC	66 °C
RimKO-DR	GTGTAGCCGGTCTCGTTCAG	66 °C
RimKO-EF	CTGCGCCGTCTGGCCACCGTCC	68 °C
RimKO-ER	CACGCCGGCTCGCGATATCG	68 °C
RimKO-FF	CTGGTCGACAGTGACATCAC	66 °C
RimKO-FR	ACACCACTCCGGTGAACGCGAAC	66 °C
otcR-F	CGGGCCATATGGACTTCAAGGCACT	66 °C
otcR-R	CCCGGAAGCTTTCAAGACGCCGACC	66 °C
Apr-F	TATACTTTCTAGAGAATAGGAA	60 °C
Apr-R	TACAGGCCAGGAGCCGGGTCGAC	60 °C
otcR-USR-F	TACGAAGACGGGCTCCGGCACC	63 °C
otcR-USR-R	CCGAGTGCCTTGAAGTCCATGG	63 °C
otrB-USR-F	CCTCGGCGGTCACCTCGATCAG	61 °C
otrB-USR-R	GATGACACGGTTCCTCACTCTC	61 °C
otcY1-USR-F	TCTCCGGAGTGGCCCGGGCGTC	68 °C
otcY1-USR-R	CACGCCGATCCCTGTGATTACG	68 °C
otcY2-USR-F	GGCTGTTCCGGGGCGGTCATA	64 °C
otcY2-USR-R	GTCCTCCGGTGTTGTGGTCT	64 °C
otcX-USR-F	CCGCCAGGCGTTCCAGCACCA	66 °C
otcX-USR-R	GAACGGCATGAAGCGTCCTCG	66 °C
otrA-USR-F	CTCGTCCAGCAGCGTCGAGGA	64 °C
otrA-USR-R	CATGGAGGGGGTTCGGATTCC	64 °C
otcG-USR-F	ATCCGCGCCACGCTCGGGATGCTGA	65 °C
otcG-USR-R	GTCGTCGACGACGAGGACGGTGATG	65 °C
otcR-KO-F	GGTGTGACCACAGATTCAGCGGAGGCGGAA GAAACCATGATTCCGGGGATCCGTCGACC	55 °C
otcR-KO-R	GCATCTCCCAGATCGGCCAGTGAGGCCGCCG GCCCCCTCATGTAGGCTGGAGCTGCTTC	55 °C

Underlined bases represent the location of endonuclease restriction sites.

2.7 Plasmid construction

2.7.1 pGB23

pGB23 is a pGB1-derived plasmid containing PCR-amplified products from the gene *SRIM7671*. These were introduced at a site flanking the thiostrepton resistance gene. Two regions from *SRIM7671* of 1,945 bp and 1,880 bp were designated Rim02 and Rim03 respectively. The start of Rim02 was positioned 5,416 bp from the beginning of *SRIM7671* and with Rim03 starting 9,887 bp in. The two regions from *SRIM7671* were PCR amplified using MyTaq DNA polymerase (Bioline). PCR amplification of Rim02 was achieved using the forward primer Rim02F and the reverse primer Rim02R. EcoRI and KpnI restriction sites were designed into the primers, forward and reverse respectively (Table 2.6). This PCR product was then purified and digested using EcoRI and KpnI and ligated into EcoRI and KpnI digested pGB1, the derived plasmid was named pGB02. After construction of pGB02, Rim03 was amplified using the forward primer Rim03F and the reverse primer Rim03R. Once again restriction sites were designed into the forward and reverse primers, PstI and HindIII respectively (Table 2.6). The Rim03 PCR product was purified, then digested with PstI and HindIII, and ligated with PstI-and-HindIII-digested pGB02, to produce the plasmid pGB23.

2.7.2 pOtcR-SR1, pOtcR-SR55 and pOtcR21-SR1

Both pOtcR-SR1 and pOtcR-SR55 are plasmids derived from pET28a, containing the whole *otcR* gene, either amplified from SR1 or from SR55. Amplification of the *otcR* gene from SR1 or SR55 was achieved using Q5 DNA polymerase (NEB), with the forward primer otcR-F and the reverse primer otcR-R. NdeI and HindIII restriction sites were designed into the primers, in the forward and reverse oligonucleotides respectively (Table 2.6). The PCR product was blunt end ligated into SmaI-digested pUC19 and then

subcloned using NdeI and HindIII restriction sites into NdeI and HindIII digested pET28a. Construction of pOtcR21-SR1 was identical with the exception of pET21a being used as the original vector.

2.7.3 pOtcR-KO

In order to generate pOtcR-KO the REDirect method (Gust *et al.,* 2003) was employed. The antibiotic marker chosen to disrupt the *otcR* gene within the cosmid pGLW207 was apramycin. The apramycin cassette was isolated from pIJ773 by restriction digest with EcoRI and HindIII. Using Q5 DNA polymerase (NEB) and apramycin cassette as the template for the otcR-KO-F and otcR-KO-R primers (Table 2.6), the disruption cassette was amplified. In order to replace the *otcR* gene on the cosmid, the disruption cassette PCR product was electro-transformed into *E. coli* BW25113 containing pGLW207 and pIJ790. After electroporation of the PCR product the transformed cells were left to grow overnight at 37 °C, so as not to maintain the pIJ790 plasmid. After this point pOtcR-KO was isolated from the cells.

2.7.4 Plasmids for generating EMSA DNA targets

Each region of DNA to be used in the electrophoretic mobility shift assay (EMSA); (described below, 2.12) was amplified with Q5 DNA polymerase (NEB) with the associated forward and reverse primers (Table 2.6, labeled -USR, for upstream region). The PCR product generated was blunt-ended and therefore could be blunt end ligated into SmaI digested pUC19. After plasmids containing each fragment of DNA required for the band shift assays were obtained, M13 (-21) Cy5 fluorophore-labeled primers could be used to generate Cy5-labelled products.

2.8 Microbiological methods

2.8.1 *S. rimosus* associated microbial methods

2.8.1.1 Preparation of electrocompetent *S. rimosus* mycelium

This protocol is modified from (Pigac & Schrempf 1995). 100 mL of CRM was inoculated with a single colony selected from an Emerson's agar plate and incubated at 30 °C for 48 hours. Vegetative hyphae was harvested by centrifugation at 4 °C for 5 minutes at 10,000 rpm (Sorvall SS-34 rotor, 107 mm). The pellet was resuspended in a total of 100 mL of ice-cold 10 % (w/v) sucrose and centrifuged again using the same conditions. The pellet was resuspended in 50 mL of ice-cold 15 % glycerol and centrifuged again as before to harvest the mycelium. The pellet was resuspended in 10 mL of 15 % (v/v) glycerol containing lysozyme at a final concentration of 400 µg/mL and incubated at 37 °C for 30 minutes. The mycelium was centrifuged again at 4 °C for 5 minutes at 10,000 rpm and resuspended in ice-cold 15 % glycerol. This was repeated again to ensure there was no residual lysozyme. Finally the pellet was resuspended in 1 to 5 mL of 30 % (w/v) PEG 1000/10 % glycerol/6.5 % sucrose in deionized water. The electrocompetent mycelium was then either used straight away or stored at -80 °C in 50 µL aliquots.

2.8.1.2 Transformation of electrocompetent *S. rimosus* mycelium

Approximately 1 µg of plasmid DNA was combined with 50 µL of electrocompetent mycelium. The DNA and mycelium mixture was then transferred to a 0.4 cm electrocuvette and subjected to a 2-kV electro pulse (10 kV/cm) with a resistance of 800 (800 Ω). The cells were then diluted in 800 µL of ice-cold CRM and placed in a 30 °C incubator for 3 hours, to allow the mycelium to recover. After this incubation period 150 µL of CRM was added to give a total volume of 1 mL. 500 µL of the transformed mycelium

was then pipetted onto two TSB plates containing the selective antibiotic, and spread. The plates were then examined for growth after 30 hours.

2.8.1.3 Introduction of DNA into *S. rimosus* spores by intergeneric conjugation

Around 10^8 *S. rimosus* spores were suspended in 500 µL of 2x YT broth and heat treated at 40 °C for 10 min. The donor strain, *E. coli* ET12567 (pUZ8002) plus desired transferring plasmid (pOtcR-KO), was grown in LB until an optical density at 650 nm (OD_{650}) of 0.4 – 0.6 was reached. The cells were washed twice with LB and resuspended in a final volume of 500 µL of LB. Equal volumes of *E. coli* cells and the *S. rimosus* spore suspension were mixed and spread on TSA, containing 10 mM $MgCl_2$. The plates were then incubated at 30 °C for 16 – 20 hours. The plates were then flooded with 1 mL water containing nalidixic acid (500 µg/mL) and apramycin (10 mg/mL) and incubated further for 3 – 5 days.

2.8.2 *E. coli* associated microbial methods

2.8.2.1 Preparation of chemically competent *E. coli* cells

500 µL of an overnight culture was used to inoculate a 250 mL conical flask containing 50 mL of LB broth and the culture grown at 37 °C until an optical density of approximately 0.4 at OD_{650} was reached. The cells were then harvested by centrifugation at 4 °C for 7 minutes at 4000 rpm (Sorvall SS-34 rotor, 107 mm). The cell pellet was then gently resuspended in 12.5 mL of ice-cold $MgCl_2$ for no longer the 3 – 5 minutes. The cell suspension of washed cells was then harvested by centrifugation at 4 °C for 7 minutes at 4000 rpm (Sorvall SS-34 rotor, 107 mm). The pellet was then resuspended in 3 mL of ice-cold $CaCl_2$, before the addition of 22 mL of $CaCl_2$ and incubated on ice for 20 minutes. The cells were once again harvested by centrifugation at 4 °C for 10 minutes at 4000 rpm (Sorvall SS-34 rotor, 107

mm). The cells were then resuspended in 1.5 mL of ice-cold $CaCl_2$ and 20 % (v/v) glycerol. The compentent cells were then stored in 100 µL aliquots at -80 °C until required.

2.8.2.2 Heat shock transformation of chemically competent *E. coli*

Approximately 500 ng of purified DNA was pipetted into an Eppendorf tube containing 50 µL of competent cells and incubated on ice for 30 minutes. The cells were then subjected to heat shock at 42 °C for 90 seconds, and then transferred to ice for 2 minutes. 450 µL of sterile 2x YT was added to the Eppendorf tube and the cells were allowed to recover for an hour by incubation at 37 °C. 100 µL of the transformed cells were spread onto an LB agar plate containing the relevant antibiotics and incubated overnight at 37 °C.

2.8.2.3 Preparation of electrocompetent *E. coli* cells

500 µL of an overnight culture was used to inoculated a 250 mL conical flask containing 50 mL of LB broth and the culture grown at 37 °C until an optical density of approximately 0.4 at OD_{650} was reached. The cells were then harvested by centrifugation at 4 °C for 7 minutes at 4000 rpm (Sorvall SS-34 rotor, 107 mm). The cell pellet was then gently resuspended in 25 mL of ice-cold 10 % (v/v) glycerol. The cells were once again harvested by centrifugation as before. The pellet was gently resuspended in 12.5 mL of ice cold 10 % glycerol, and centrifuged as before to collect the cells. Once again the cell pellet was resuspended with 5 mL of ice-cold 10 % glycerol and the cells centrifuged as before. The resulting cell pellet was resuspended in 1 mL of ice-cold 10 % glycerol. The electrocompentent cells were then stored in 50 µL aliquots at -80 °C until required.

2.8.2.4 Transformation of electrocompetent *E. coli* cells

Approximately 500 ng of purified DNA was mixed with 50 µL of electrocompentent cells and transferred to a 0.1 cm electrocuvette. The cells were then subjected to 2-kV electro pulse (2.5 kV/cm) at 200 Ω. 450 µL of sterile 2x YT was added to the eppendorf tube and the cells were allowed to recover for an hour by incubation at 37 °C. 100 µL of the transformed cells were spread onto an LB agar plate containing the relevant antibiotics and incubated overnight at 37 °C.

2.8.3 Extraction of OTC and rimocidin

Each strain was grown in 100 mL of liquid Emerson's media in 250 mL flasks. After 10 days the cultures were transferred to a 1 L separating funnel, at which point 300 mL of ethyl acetate was added. The mixture was then shaken to mix and left to stand. The bottom layer was drained and put aside to be extracted again. The upper phase was transferred into a round bottom flask. The solution was then evaporated which was evaporated at 50 °C in a rotary evaporator. The extracts were reconstituted in 1 mL ethanol or HPLC-grade methanol for mass spectrometry.

2.8.4 Determination of antimicrobial activity of extracts

Disc diffusion assays were performed to determine antimicrobial activity of the extracts obtained from each *S. rimosus* strain. Petri dishes of either LB agar or YPD agar were used with an overlay of soft nutrient agar (SNA). The SNA was inoculated with either *B. subtilis* or *S. cerevisiae* as the test organism for either the detection of OTC or rimocidin respectively. 5 µL of 100 mg/mL tetracycline or 5 µL of 10 mg/mL cycloheximide was applied to Whatman discs (125 mm, 11 µm particle retention) as positive controls for OTC or rimocidin production respectively. 5 µL from each extract was spotted onto Whatman discs and laid onto the SNA overlay. Once the discs

were applied to the overlay, the petri dish plates were placed in a 37 °C or 30 °C incubator overnight.

2.9 Protein methods

2.9.1 Production of OtcR

pOtcR-SR1 or pOtcR-SR55 in BL21(DE3)(pLysS) was used for the production of OtcR. 20 mL of overnight culture was used to inoculate 1 L of LB broth containing 50 µg/mL of kanamycin. The culture was grown at 30 °C until an OD of 0.4 at OD_{650} was reached, at this point IPTG was added to give a final concentration of 100 mM to induce expression of OtcR. After around 3 hours the cells were harvested by centrifugation for 10 minutes at 5000 rpm (Beckman, JLA 9.1000, 185 mm), at 4 °C. Cells were stored at -80 °C if not used immediately.

2.9.2 SDS Polyacrlyamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was employed to separate proteins by molecular weight. Glass gel plates (Bio-Rad), gel combs and spacer plates were cleaned using distilled water, followed by 100 % ethanol, and allowed to dry before assembly of the gel mould. 12 % polyacrylamide gels were used throughout this work. The resolving gel was made by combining 2.5 mL resolving buffer (4x), 3.5 mL distilled water and 4 mL 30 % acrylamide (37.5:1). To this mixture, 100 µL of 10 % ammonium persulphate (APS) and 10 µL of TEMED poured into the pre-prepared gel mould. The stacking gel was then prepared by mixing together 2.5 mL stacking buffer (4x), 5.5 mL of distilled water and 2 mL 30 % acrylamide. Once again 100 µL of 10 % APS and 10 µL of TEMED was added, mixed and quickly poured into the gel mould. A comb was added to produce the wells and the gel was left to set. An equal volume of SDS-PAGE loading dye was added to each sample and boiled for 10 minutes before loading into the gel. Proteins were separated by

electrophoresis for 90 minutes at 120 V and 20 mA. Gels were then stained in around 20 mL of SDS-PAGE staining gel for 15 minutes and allowed to destain overnight in distilled water or by microwaving the gel in 500 mL of distilled water.

2.9.3 Small-scale preparations of OtcR for solubility tests

50 mL cultures were used to overexpress OtcR as described previously (2.8.1). The pellet was resuspended in 1 mL of the resuspension buffer being tested and transferred to a 1.5 mL Eppendorf. The cells were disrupted by sonication using a Bioruptor® system, set to high with 30 second bursts, followed by 30 seconds of rest for a total of 10 minutes. The disrupted cells were centrifuged at 4 °C, for 10 minutes at 14,000 rpm (Beckman Coulter, FX241.5P, 66 mm). The cell lysate was transferred to a new Eppendorf tube and the cell pellet resuspended in 1 mL of distilled water. Samples were analysed by SDS-PAGE.

2.9.4 Purification of OtcR

30 mL of OtcR buffer A was added to the cell pellet (See section 2.9.1) and resuspended. The cells were disrupted in a French pressure cell at 1000 psi, and insoluble material was removed by centrifuging at 15,000 rpm (Sorvall, SS-34, 107 mm) for 1 hour at 4 °C. The soluble fraction was purified using a 1 mL HisTrap HP Ni sepharose column (GE Healthcare), that had been equilibrated with OtcR buffer A for at least 20 minutes using an ÄKTA protein purification system. OtcR was eluted using a gradient of OtcR buffer B introduced at a flow rate of 1 mL/min over 20 minutes. Fractions were collected and those containing OtcR were determined by SDS-PAGE.

2.10 Preparation of crude *S. rimosus* cell lysate

Around 10^8 *S. rimosus* spores were suspended in 500 µL of 2x YT and heat shocked at 50 °C for 10 minutes. This suspension was used to inoculate 5 mL of TSB to produce a seed culture, which was grown for a minimum of 18 hours and no greater than 20 hours. 1 mL of this seed culture was used to inoculate 100 mL of liquid Emerson's and incubated at 30 °C. At each appropriate time point 50 mL was removed and the cells harvested via centrifugation at 4000 rpm (Sorvall, SS-34, 107 mm) for 10 minutes. The cells were resuspended in 1 mL of TEG buffer and transferred to an Eppendorf tube. The cells were washed 3 times with 1 mL of TEG buffer. At this point the cells were disrupted by sonication using a Bioruptor® system, set to high with 30 second bursts, followed by 30 seconds of rest for a total of 15 minutes. Throughout this process the cells were kept as close to 4 °C as possible. The disrupted cells were centrifuged at 4 °C, for 10 minutes at 14,000 rpm (Beckman Coulter, FX241.5P, 66 mm). The cell lysate was transferred to a fresh Eppendorf tube and 50 µL removed for EMSA (see section 2.12), with the pellet resuspended in 1 mL of sterile distilled water.

2.11 DNA sampling of the upstream region of *otcR*

To identify any proteins that bind to the upstream region (USR) of the *otcR* gene, biotin-labelled *otcR*-USR was used in conjunction with streptavidin magnetic beads. 150 µL of Dynabeads® M-270 Streptavidin magnetic beads were washed using 1 mL of binding and washing (B&W) buffer (1x), as outlined in the manufacturer's protocol (Invitrogen). After 3 total washes the beads were ready to be used to immobilise the biotin labelled *otcR*-USR. 300 µL of B&W buffer (2x) was used to resuspend the beads, and to this 55.5 µg of biotin-labelled *otcR*-USR, diluted to 300 µL with distilled water, was added. The mixture was incubated for 10 minutes with gentle rotation at room temperature. The beads were then washed 3 times

with B&W buffer (1x). At this point crude *S. rimosus* cell lysate (See section 2.10) was mixed with the Dynabeads® either with or without bound biotin-labelled *otcR*-USR. The lysate-bead mixture was incubated at room temperature with gentle rotation for 10 minutes. The beads were then collected using a Dynamag™-2 magnet, and all non-bound cell lysate was removed and kept. The beads were then washed 5 times using B&W buffer (1x). In order to elute the bound protein from the beads 30 µL of SDS-PAGE loading dye was added, as this would denature the protein and disrupt binding. At this point all samples to be analysed were resolved by SDS-PAGE.

2.12 Electrophoretic mobility shift assay protocol for pure and crude protein samples

Band shift assay gel mould components were cleaned using distilled water followed by 100 % ethanol and then allowed to dry. For band shift assays 30 % acrylamide (80:1) was used, which was produced by mixing together 75 mL of 40 % acrylamide and 18.75 mL of N,N'-methylene-bis-acrylamide. EMSA gels were produced by combining 4 mL of MADAM's buffer (5x), 2.8 mL of 30 % acrylamide, 2.5 mL of 20 % glycerol (v/v) and 10 mL of sterile distilled water, and then mixed. To this solution 200 µL of 10 % APS and 25 µL of TEMED was added and mixed. The mixture was then quickly poured into the gel mould and a comb added to produce the wells. This gel was allowed to set for 2 hours. Protein/DNA binding was achieved by mixing around 4 ng of cyanine 5 (Cy5) fluorophore-labeled DNA with 5 µL of TAP buffer (2x), 1 µL of BSA and 3 µL of either OtcR protein or crude cell lysate, to give a total volume of 10 µL. This mixture was incubated at 30 °C for 5 minutes. 2 µL of 6x bromophenol blue loading dye was added to each sample and loaded onto the gel, which was ran for 45 minutes at 100 V and 16 mA. The gel was then visualised using a Typhoon scanner set with a 670 nm band-pass filter for Cy5 flourescence, with a red (633 nm) laser and a 400 photomultiplier tube.

2.13 Mass spectrometry data processing

High-resolution mass spectrometry (HRMS) was carried out on secondary metabolite extracts from *S. rimosus* by Dr Lynsey MacIntyre at the University of Strathclyde. The raw MS data was processed using MZmine version 2.10 (Pluskal *et al.,* 2010) in order to generate peak lists. The data processing steps required peak detection involving mass detection and chromatogram building. The default parameters were used for mass detection as well as chromatogram building, with the exception of noise level, which was set to 1000. This was followed by chromatogram deconvolution using the local minimum search algorithm. Parameters for chromatogram deconvolution were: chromatographic threshold was 90 %, with a minimum retention time of 0.4 minutes, a 5 % minimum relative height, with a minimum absolute height of 10000, minimum ratio of peak top/edge was 3 and peak duration range was 0.2 – 5 minutes. Deisotoping with a mass to charge ratio (m/z) tolerance of 5 parts per million (ppm), retention time (RT) tolerance of 0.1 minutes, a maximum charge of 2 and representative isotope set to most intense, then followed. Chromatogram alignment and gap filling was then carried out. The m/z tolerance was set to 5 ppm, with a weight for m/z of 20, a RT tolerance 5 relative (%) and weight for RT of 20 for join aligner. Setting the weight of m/z and RT to the same value allows for equal importance to be given. At this point gap filling could be applied to the data to identify any previously undetected peaks, and an m/z tolerance of 5 ppm was used, with a 1 % intensity tolerance and a RT tolerance of 0.5 minutes. All peaks found within the solvent were removed as well as peaks with low intensity; this cut off was 1.0E4. In order to predict molecular formulas and minimise mis-assignment, adducts and complexes were identified. This was achieved by searching for Na, K, NH4 and ACN+H (acetonitrile), with a RT tolerance of 0.2 minutes, m/z tolerance of 5 ppm and max relative adduct peak height of 30 %. Complexes were determined using a positive electrospray ionisation mode, with a RT tolerance of 0.2 minutes, m/z tolerance of 5 ppm and max complex peak height of 50 %.

2.14 Computational Methods

2.14.1 GET_HOMOLOGUES

The GET_HOMOLOGUES script get_homologues.pl was run using the default parameters of a minimum of 75 % coverage and a maximum E value of 1e-05. The script used to generate core orthologous groups, using the OrthMCL algorithm, with genome composition analysis was as follows:

```
get_homologues.pl -d Folder_with_sequences -c -M
```

The script used for the generation of the pan orthologous groups, using the OrthMCL algorithm, with genome composition analysis was as follows:

```
get_homologues.pl -d Folder_with_sequences -c -M -t 0
```

For plotting the core and pan genome the following command was used:

```
plot_pancore_matrix.pl -i path_to_.tab_files -f core_both
plot_pancore_matrix.pl -i path_to_.tab_files -f pan
```

For the generation of a pan genome matrix, for the production of a parsimony tree, the following script was used:

```
compare_clusters.pl -o name_of_output_folder -m -T
```

2.14.2 CGView Comparison Tool

For the generation of maps the following command was used:

```
cgview_comparison_tool.pl -p project_name -t --cct
```

For different comparisons, i.e DNA, amino acid the project.settings.conf file was modified to give the appropriate “query_source” and “database_source”. For the inclusion of specific labels, a text file with the required information was added to the project folder and another text file to the features folder.

2.14.3 Burrows-Wheeler Alignment (BWA) and Sequence Alignment/Map (SAMtools)

The generation of .sam files for the Illumina paired end data was achieved with the following commands:

```
bwa index SR1.fasta
bwa aln SR1.fasta Illumina_data_1.fastq > Illumina_data_1.sai
bwa aln SR1.fasta Illumina_data_2.fastq > Illumina_data_2.sai
bwa sampe SR1.fasta Illumina_data_1.sai Illumina_data_2.sai
Illumina_data_1.fastq Illumina_data_2.fastq > Illumina_data.sam
```

.sam files were produced for the 454 reads with the following commands:

```
bwa index SR1.fasta
bwa mem SR1.fasta 454_data.fastq > 454_data.sam
```

The following commands were used to generate .bam files from all .sam files:

```
samtools faidx SR1.fasta
samtools import SR1.fasta.fai data.sam data.bam
samtools sort data.bam data.sorted
samtools index data.sorted.bam
```

The above commands produce a .bam, .sorted.bam and .sorted.bam.bai.

Chapter 3

3 Comparative genomics of *S. rimosus* SR1 and 19 complete *Streptomyces* genomes

3.1 Introduction

In order to investigate the genetic basis for increased OTC productivity it was necessary to generate a genome sequence of the original soil isolate *S. rimosus* SR1 (G7), as well as the subsequent production strains. *S. rimosus* SR1 was used as the reference sequence and a whole-genome shotgun sequencing approach was performed with a Roche 454 GS Junior sequencer. This approach was used as the coverage and average read length was optimal for assembling a scaffold genome sequence (Loman *et al.,* 2012), which could then be used for genomic comparisons between *S. rimosus* SR1 and the subsequent strains in the lineage. In addition, obtaining the genome sequence of *S. rimosus* SR1 allowed for the opportunity to undertake comparative genomic analysis between this bacterium and other streptomycetes.

The use of comparative genomics provides insight into the mechanism of environmental adaption, variability of virulence factors or identification of epidemiological markers (Li *et al.,* 2013a; Jeukens *et al.,* 2014; Méric *et al.,* 2014). *Streptomyces* are a diverse bacterial genus known for the vast number of distinct antibiotics produced, contributing around two-thirds to the pool of known antibiotics (Lal *et al.,* 1996; Berdy 2012). Further to this *Streptomyces* species have been isolated from a variety of different sources, such as soil, ants, Antarctic lichen, garlic bulbs and the ocean (Seipke *et al.,* 2011; Klassen *et al.,* 2011; Baranasic *et al.,* 2013; Shin *et al.,* 2013; Olano *et al.,* 2014), highlighting the varied ecological niches in which these bacteria can grow. Given the varied habitats and the importance of this genus, as a source of secondary metabolites, the use of comparative genomics provides

an opportunity to better understand the genetic relationship between members and also those elements that make these organisms unique. With the recent progress in bioinformatics an increasing number of programs are available for handling large genomic datasets, such as GET_HOMOLOGUES (Contreras-Moreira & Vinuesa 2013), CGView Comparison Tool (CCT) (Grant *et al.,* 2012) and Mugsy (Angiuoli & Salzberg 2011). Such tools allow for rapid comparison of whole genomes to elucidate core features within a species and define phylogenetic relationships. Other online tools can be used to identify components within individual genome sequences, such as secondary metabolite clusters as seen with antiSMASH (Medema *et al.,* 2011; Blin *et al.,* 2013) or phage discovery, as achieved by PHAST (Zhou *et al.,* 2011). These tools, in combination with the increasing number of whole genome sequences becoming available, allows for entire genome comparison of different *Streptomyces* species. This type of comparison can aid the identification of core genes that can be used to define the *Streptomyces* genus and those that were required for an individual species to adapt to a particular environment.

As the *S. rimosus* SR1 genome had never been available before for this type of analysis, it was selected, along with all complete *Streptomyces* genomes (Table 3.1), with the aim of identifying those core genes, present in all the *Streptomyces* species in Table 3.1. In addition some of the phylogenetic relationships between these species are defined, with *S. rimosus* SR1 and *S. cattleya* being more evolutionarily related to each other compared to the other *Streptomyces* (Anderson & Wellington 2001; Kirby *et al.,* 2008). Therefore selecting these *Streptomyces* species provides a diverse range of organisms, thus providing greater insight into this genus.

Table 3.1 Candidate streptomycetes subjected to comparative genomic analysis.

Genome	Accession Number	Size (Mbp)	GC (%)	Genes	Proteins
S. albus J1074	CP004370	6.84	73.5	5937	5832
S. avermitilis MA-4680	BA000030	9.03	71	7670	7580
S. bingchenggensis BCW-1	CP002047	11.94	71.2	10107	10022
S. cattleya DSM 46488	CP003219	6.28	73	5904	5822
S. cattleya NRRL 8057	FQ859185	6.28	73	5870	5771
S. coelicolor A3(2)	NC_003888	8.66	72	7910	7767
S. collinus Tu 365	CP006259	8.27	72.9	7099	7005
S. davawensis JCM 4913	HE971709	9.46	71	8584	8503
S. flavogriseus ATCC 3331*	CP002475	7.34	71.4	6531	6298
S. fulvissimus DSM 40593	CP005080	7.91	71.8	7027	6925
S. griseus NBRC 13350	AP009493	8.55	72.4	7224	7138
S. hygroscopicus jinggangensis 5008	CP003275	10.14	72.3	8936	8849
S. *hygroscopicus* jinggangensis TL01	CP003720	9.84	72.4	8706	8619
S. sp. PAMC26508	CP003990	7.53	71.3	7054	6969
S. rapamycinicus NRRL 5491	CP006567	12.7	71.4	10144	10002
S. rimosus SR1	ANSJ00000000	9.5	71.9	8760	8651
S. scabiei 87 .22	FN554889	10.15	71.5	8907	8809
S. sp. SirexAA-E	CP002993	7.41	71.9	6648	6357
S. venezuelae ATCC 10712	FR845719	8.23	72.4	7541	7453
S. violaceusniger Tu 4113	CP002994	10.66	70.9	9063	8482

*As of 12/05/14 *S. flavogriseus* ATCC 3331 has been renamed *S. pratensis* ATCC 3331 due to the findings of (Rong *et al.,* 2013).

3.2 Genome sequencing of *S. rimosus* SR1

A whole-genome shotgun sequencing approach was used for sequencing the genome of *S. rimosus* SR1, using the Roche GS-Junior 454 apparatus. A total of 440,253 reads were obtained from three single-ended runs, with a mean read length of 325 bp. The combined reads gave a 17x coverage and were assembled using GS *de novo* Assembler software, Newbler (Roche), producing an assembly of 453 contigs of >500 bp, with an average contig size of 21 Kbp. The draft assembly gave a total genome size of 9,501,372 bp, which lies within the average size range of Streptomyces genomes (Zhou *et al.,* 2012b), and a G+C content of 71.88 %. The Rapid Annotation using Subsystem Technology (RAST) (Aziz *et al.,* 2008) server software was used to obtain automatic functional annotation results. The RAST annotation predicted 8,651 protein-coding genes, with 66 tRNAs, within the draft genome of *S. rimosus* SR1. The resulting annotated draft assembly was used for all subsequent analysis and comparisons.

3.3 The core, accessory and unique genome of *Streptomyces*

3.3.1 Core and Pan-genome Analysis

It was necessary to determining the core and pan-genome of the candidate *Streptomyces* species in order to identify genomic features common to all and to distinguish those that are unique to *S. rimosus* SR1. For this purpose 19 complete genomes (Table 3.1) were obtained from the NCBI ftp genbank site (ftp://ftp.ncbi.nlm.nih.gov/genbank/genomes/Bacteria) to use in the analysis with *S. rimosus* SR1. Initial CCT comparisons of *S. rimosus* SR1 with the 19 complete genomes demonstrated little sequence identity at the nucleotide level (Figure 3.1 A), although amino acid sequence identity was higher (Figure 3.1 B). In order to avoid false negatives and ensure the identification of all potential orthologues, it was decided to compare the predicted proteomes of the 20 *Streptomyces* species when establishing the core and pan-genome, rather than nucleotide genomes. The 155,622 total genes from the 20 *Streptomyces* genomes resulted in a predicted proteome of 152,854 protein-coding sequences. Using the GET_HOMOLOGUES (Contreras-Moreira & Vinuesa 2013) software package and applying the OrthoMCL (Fischer *et al.,* 2002; Li *et al.,* 2003) clustering algorithm (See Chapter 2.14), 17,133 homologous clusters were identified, in addition to this a total of 18,839 unique proteins (proteins present only in one species) were also discovered. From this initial analysis a putative core-genome of 1879 clustered proteins was predicted, with the accessory (proteins present in one or more but not all genomes) consisting of 15,254 proteins, resulting in a pan-genome of 35,972 proteins with the inclusion of the 18,839 unique proteins. In addition, random sampling experiments were carried out to estimate the theoretical core and pan-genome size. For the theoretical core-genome this involved the random addition of each *Streptomyces* genome to the pan-genome pool, to track the proportion of common and unique genes contributed. This can be considered in the following way: genome A is compared to genome B and C,

with genome A sharing more genes in common with genome B than C. The resulting core genome of A and B would be greater than that of A and C, and this is represented by the black circles on the plot in Figure 3.2 A. By the time a total of 20 genomes are considered the variation in core genome size is diminished, as all 20 genomes must contain the same genes. Once the average values of the core-genome at each addition step were determined Tettelin and Willenbrock fitted functions, for establishing the core genome, could be applied (Figure 3.2 A) {Tettelin:2005jg, Willenbrock:2007kw}. The fitted function for Tettelin predicted a core genome of 1910 (Figure 3.2 A), which is only slightly larger than the core-genome predicted by GET_HOMOLOGUES. However the core-genome estimated by the Willenbrock fitted function (Figure 3.2 A) was 2078; almost 200 genes greater than that identified by GET_HOMOLOGUES. Despite this the random sampling predictions (Figure 3.2 A) still suggest that the core would remain relatively constant even with the addition of other genomes. The theoretical pan-genome involves the random addition of each *Streptomyces* genome to the pan-genome pool. Again each random addition of a genome is considered by calculating the difference to the total pan-genome dependent on the genome added at each step. This is represented by the black circles, with the variance greater when only 1 genome is considered compared to all 20, as the total number of genes is much greater therefore the difference is less distinct (Figure 3.2 B). The theoretical pan-genome implied an open pan-genome (Figure 3.2 B), as the pan-genome will increase indefinitely with the subsequent addition of other genomes, as the pan-genome consists of all genes present in the test sample.

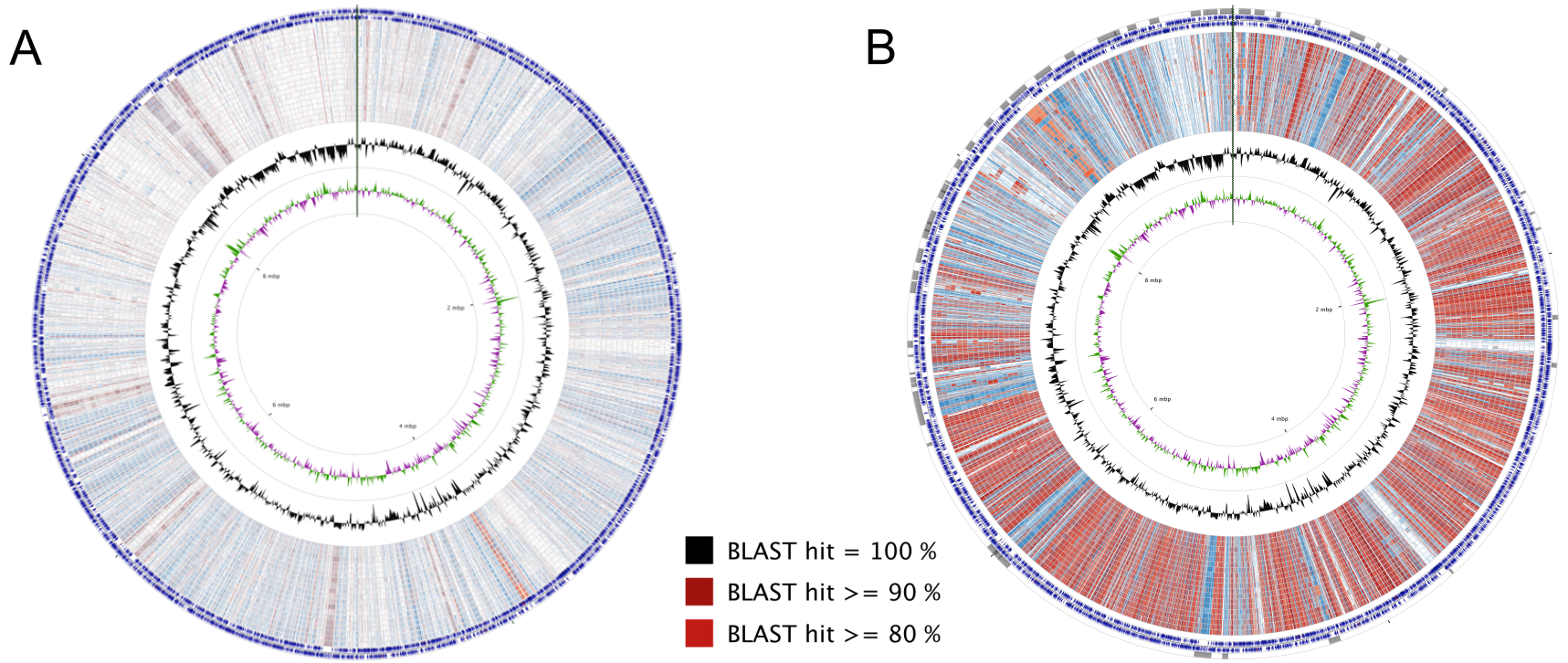


Figure 3.1 CCT map of *S. rimosus* SR1 compared to 19 complete *Streptomyces* genomes. Rings from the outside inward: SR1 plus-strand, SR1 minus-strand, *S. rapamycinicus* NRRL 5491, *S. bingchenggensis* BCW 1, *S. violaceusniger* Tu 4113, *S. hygroscopicus jinggangensis* 5008, *S. hygroscopicus jinggangensis* TL01, *S. davawensis* JCM 4913, *S. avermitilis*, *S. scabiei* 87 22, *S. coelicolor* A3(2), *S. griseus* NBRC 13350, *S. venezuelae* ATCC 10712, *S. fulvissimus* DSM 40593, *S. collinus* Tu 365, *S. sp.* SirexAA, *S. sp.* PAMC26508, *S. flavogriseus* ATCC 3331, *S. albus* J1074, *S. cattleya* DSM 46488, *S. cattleya* NRRL 8057, GC content plot, GC skew and genomic position. BLAST hits that are a light red/orange colour = 50 – 79 % identity, a blue colour = 0 – 49 % identity. **A.** nBLAST comparison. DNA sequences were extracted from Genbank files of genomes listed in Table 1. **B.** pBLAST comparison. Amino acid sequences were extracted from Genbank files of genomes listed in Table 1.

The estimated percentage of proteins that contributed to the core of each *Streptomyces* genome ranged from 19 % for *S. bingchenggensis* BCW-1 to 32 % for *S. cattleya* NRRL 8057, with an average of 1971±24 core proteins per genome. Of the unique proteins an average of 1170±570 per genome were identified, with *S. flavogriseus* ATCC 3331 found to have the lowest percentage, of just 2.5 %, that contributed to its proteome, with *S. rimosus* SR1 having the highest percentage of 26 %. It should be noted that due to the presence of two strains of *S. cattleya* and *S. hygroscopicus jinggangensis* identification of unique proteins was achieved by determining unique proteins compared to all genomes, then all genomes plus the alternative strain, and then finally strain versus strain. Performing the analysis this way enabled the identification of unique proteins belonging to *S. cattleya* or *S. hygroscopicus jinggangensis* species as well as unique proteins between the strains. After establishing the proteins that contributed to the core and unique genome the accessory genome could be ascertained. *S. violaceusniger* Tu 4113 was found to have the highest percentage of proteins that contributed to its accessory genome, with 67.8 %, followed closely by *S. flavogriseus* ATCC 3331 at 67.6 %, and the accessory genome of *S. cattleya* DSM 46488 was the smallest at only 43 %. Overall the average accessory genome for each species or strain was 4500±985. Therefore on average 25 % of each *Streptomyces* genome consisted of core genes, with 60 % made up by accessory genes and a final 15 % comprised of unique genes, if we consider each encoding protein corresponds to a core, accessory or unique gene.

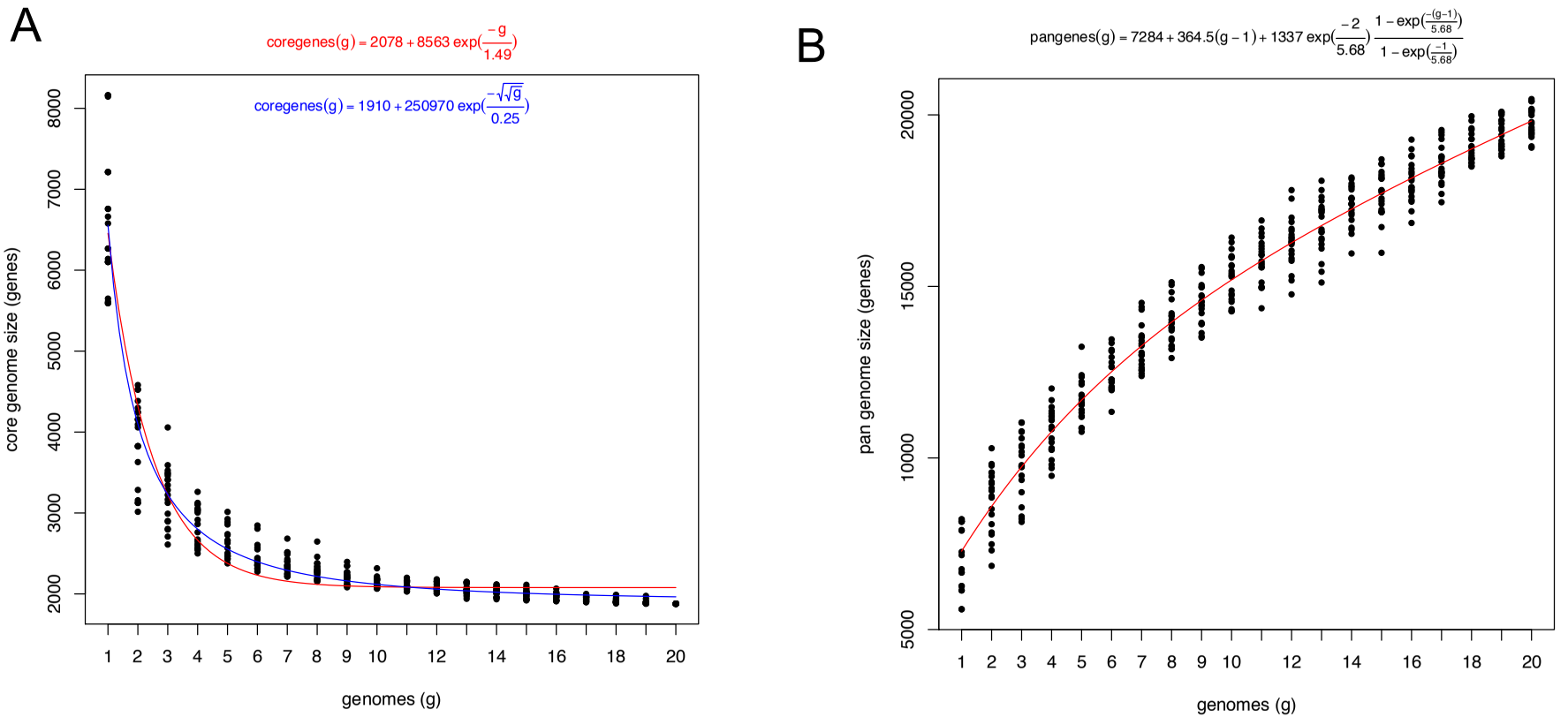


Figure 3.2 Random sampling plots of the core and pan-genome. Black circles represent difference to the core or pan genome with the addition of each different genome for each total genome number. **A.** Core genome size prediction using Tettelin (blue) and Willenbrock (red) fitted function, represented by the corresponding line, blue and red respectively (Tettelin *et al.,* 2005; Willenbrock *et al.,* 2007). Each corresponding fitted function is shown above the graph, with blue representing Tettelin and red representing Willenbrock. Residual standard error for Tettelin = 299.67, residual standard error for Willenbrock = 256.46. **B.** The red line represents the pan-genome size prediction using Tettelin fitted function, as shown above the graph. Residual standard error = 670.68.

3.3.2 *Streptomyces* core genome

After identifying the proteins that contributed to the core genome, these proteins were reverse translated and concatenated to produce a nucleotide core genome. The nucleotide core genome was uploaded to RAST in order to determine the putative function of each gene, using the SEED subsystem (Overbeek 2005). Of the 1879 genes identified, 1433 were assigned to SEED categories, with 446 uncategorised (Table 3.2). Of the clustered genes that were categorised, the majority belonged to amino acid, and carbohydrates SEED subsystems. This was to be expected, as these subsystems are associated with primary metabolism, the core process in all bacterial species.

The regulation and cell signaling SEED subsystem was the next group that was highly represented after amino acids and derivatives and carbohydrates. The majority of the genes within this category were transcriptional regulators of the TetR family, with an average of 35 per genome, followed by two component systems (TCS), containing response regulators (RR) paired with sensor kinases (SK) (Table 3.3). Examination of other members of the actinomycetes found that within the core genome of Nocardiopsis, and the complete genomes of *Saccharopolyspora erythraea* NRRL 23338 and *Amycolatopsis mediterranei* U32, TetR family transcriptional regulators were the most abundant (Oliynyk *et al.,* 2007; Zhao *et al.,* 2010; Li *et al.,* 2013a). The majority of the TetR homologues from *S. coelicolor* A3(2), that were within the core genome, were uncharacterised. SCO2374 the gene that encodes one of the uncharacterised TetR proteins, lies adjacent to SCO2373, a tetracenomycin C efflux protein and this was observed for all the Streptomyces with the exception of *S. albus* J1704, in which a MarR transcriptional regulator lies upstream. SCO4118 was another TetR-like regulator that was found within all the *Streptomyces* and encodes AtrA, a regulator of *actII-ORF4,* the pathway specific regulator of ACT (Uguru *et al.,* 2005). Unlike the TetR proteins the TCS associated proteins were generally characterised, with PhoP and PhoR, AfsQ1 and AfsQ2, AbrC1 with

AbrC2 and AbrC3, and CutS with CutR all of which are associated with antibiotic production (Rodríguez *et al.,* 2013). There was also CseB with CseC and OsaA with OsaB, TCS linked with regulating cell wall integrity and osmoadaption, respectively (Bishop *et al.,* 2004; Hutchings 2004; Lu *et al.,* 2007).

Of the core genes identified only 5 were associated with secondary metabolism, however antiSMASH analysis identified 3 secondary metabolite clusters. One of the predicted terpenes was geosmin a metabolite known to be produced by *Streptomyces* and responsible for the characteristic fragrance of moist soil (Cane *et al.,* 2006). The other terpene has a squalene-hopene cyclase within its biosynthetic cluster, suggesting the product is a hopenoid (Siedenburg & Jendrossek 2011). The final secondary metabolite cluster was for a predicted ectoine. Ectoine is biosynthesised through the action of three proteins, EctA, EctB and EctC, in addition ectoine can be converted to 5-hydroxyectoine by EctD (Bursy *et al.,* 2008; Sadeghi *et al.,* 2014). Examination of the predicted ectoine cluster in the core genome revealed genes that encode homologues of EctA, EctB, EctC and EctD. Ectoine and 5-hydroxyectoine both act as osmoprotectants as well as thermoprotectants in *S. coelicolor* A3(2) and a number of other *Streptomyces* (Bursy *et al.,* 2008; Sadeghi *et al.,* 2014).

Table 3.2 Functional SEED categories of the *S. rimosus* SR1 core genome.

SEED Subsystem	Total number of genes	Contribution to core genome (%)
Cofactors, Vitamins, Prosthetic Groups, Pigments	115	6
Cell Wall and Capsule	61	3
Virulence, Disease, Defense and Resistance	28	1
Potassium Metabolism	9	0
Miscellaneous	20	1
Membrane Transport	38	2
Iron Acquisition	3	0
RNA Metabolism	54	3
Nucleosides and Nucleotides	76	4
Protein Metabolism	109	6
Cell Division and Cell Cycle	25	1
Motility Chemotaxis	1	0
Regulation and Cell Signalling	133	7
Secondary Metabolism	5	0
DNA Metabolism	65	3
Fatty Acids, Lipids, Isoprenoids	122	6
Nitrogen Metabolism	9	0
Dormancy and Sporulation	4	0
Respiration	97	5
Stress Response	51	3
Metabolism of Aromatic Compounds	15	1
Amino Acids and Derivatives	261	14
Sulphur Metabolism	7	0
Phosphorus Metabolism	24	1
Carbohydrates	230	12
Phages, Prophages, Transposable Elements, Plasmids	0	0
Hypothetical/Uncategorised	317	17

Table 3.3 Core regulators of *Streptomyces* genomes.

	TetR	MarR	GntR	TCS	RR	SK	Other Families*
S. albus J1074	35	13	14	32	7	6	43
S. avermitilis MA-4680	34	13	16	32	7	6	43
S. bingchenggensis BCW-1	37	14	15	36	7	6	44
S. cattleya DSM 46488	35	13	14	31	7	6	43
S. cattleya NRRL 8057	35	13	14	31	7	6	43
S. coelicolor A3(2)	36	13	15	31	7	6	43
S. collinus Tu 365	34	13	14	33	7	6	43
S. davawensis JCM 4913	34	14	16	32	7	6	44
S. flavogriseus ATCC 3331	34	13	14	31	7	6	43
S. fulvissimus DSM 40593	35	13	15	31	7	6	44
S. griseus NBRC 13350	35	13	15	31	7	6	45
S. hygroscopicus jinggangensis 5008	34	13	16	33	7	6	42
S. *hygroscopicus* jinggangensis TL01	34	13	16	33	7	6	42
S. sp. PAMC26508	34	13	14	31	7	6	43
S. rapamycinicus NRRL 5491	35	14	14	33	7	6	42
S. rimosus SR1	34	13	15	31	7	6	43
S. scabiei 87 .22	33	13	14	31	7	6	44
S. sp. SirexAA-E	35	13	15	37	7	6	47
S. venezuelae ATCC 10712	35	13	14	31	7	6	45
S. violaceusniger Tu 4113	36	14	14	33	7	6	44

TCS = Two Component System RR = Lone Response Regulator SK = Lone Sensor Kinase

*Other families include: ROK, IclR, LysR, LacI, CadR, MerR, AsnC, AraC, DeoR, PadR, XRE

3.3.3 *S. rimosus* SR1 accessory genome

The accessory genome of *S. rimosus* SR1 contains 4384 genes and can be used to give an overview of *Streptomyces* accessory genomes. Once again the identified proteins of the accessory proteome of *S. rimosus* SR1 were reverse translated, concatenated and uploaded to RAST to obtain predictions of each gene's function. Of the 4384 genes annotated only 2589 were given SEED subsystem functions, with 1795 being hypothetical or uncharacterised (Table 3.4). Regulation and cell signaling was the principal SEED subsystem represented within the accessory genome followed by, cell wall and capsule metabolism, membrane transport and carbohydrate metabolism. Of these SEED subsystems carbohydrate metabolism was the only one that contributed less to the accessory genome compared with its contribution to the core genome. When compared to the core genome there was a higher percentage of hypothetical/uncharacterised genes, most likely due to the nature of both the accessory and core genome, with more essential, well-characterised genes in the core (Tettelin *et al.,* 2005; Lefébure & Stanhope 2007).

As with the core only a small number of genes were attributed to secondary metabolism, nonetheless antiSMASH annotation predicted 36 biosynthetic clusters. Further to this it had already been established that *S. rimosus* SR1 produces at least three different secondary metabolites, whose combined gene clusters include 35 genes solely associated with biosynthesis or modification (Barona-Gómez *et al.,* 2004; Seco *et al.,* 2004; Petković *et al.,* 2006; Pethick *et al.,* 2013). RAST annotation of the biosynthetic gene clusters alone identified by antiSMASH attributed nine genes to the secondary metabolism SEED subsystem with the majority being associated with cofactors, vitamins, prosthetic groups, and pigments metabolism. As many of these secondary metabolite clusters are putative, it seems reasonable to associate these genes with other cellular functions, as these clusters may synthesis molecules not traditionally associated with secondary metabolism such as siderophores or biotin.

Table 3.4 Functional SEED categories of the *S. rimosus* SR1 accessory genome.

SEED Subsystem	Total number of genes	Contribution to accessory genome (%)
Cofactors, Vitamins, Prosthetic Groups, Pigments	124	2.83
Cell Wall and Capsule	259	5.91
Virulence, Disease, Defense and Resistance	83	1.89
Potassium Metabolism	7	0.16
Miscellaneous	19	0.43
Membrane Transport	250	5.70
Iron Acquisition	60	1.37
RNA Metabolism	64	1.46
Nucleosides and Nucleotides	44	1.00
Protein Metabolism	93	2.12
Cell Division and Cell Cycle	10	0.23
Motility Chemotaxis	0	0.00
Regulation and Cell Signalling	449	1.44
Secondary Metabolism	51	10.24
DNA Metabolism	54	1.16
Fatty Acids, Lipids, Isoprenoids	120	1.23
Nitrogen Metabolism	44	2.74
Dormancy and Sporulation	6	1.00
Respiration	109	0.14
Stress Response	94	2.49
Metabolism of Aromatic Compounds	50	2.14
Amino Acids and Derivatives	197	1.14
Sulphur Metabolism	42	4.49
Phosphorus Metabolism	25	0.96
Carbohydrates	272	0.57
Phages, Prophages, Transposable Elements, Plasmids	63	6.20
Hypothetical/Uncategorised	1795	40.94

3.3.4 *S. rimosus* SR1 unique genes

Of all the unique genes identified *S. rimosus* SR1 had the greatest number of genes at 2283. As described before the proteins identified with the GET_HOMOLOGUES script as unique were prepared for RAST annotation. 1776 genes were within the hypothetical/uncategorised SEED subsystem (Table 3.5), accounting for 78 % of the total genes. Once again the regulation and cell signaling subsystem contributed the most after hypothetical/uncategorised, with cell wall and capsule metabolism, followed by amino acid metabolism. A total of 106 genes were associated with the regulation and cell signaling subsystem, and of those genes the majority were putatively associated with transcriptional regulation. There were 63 transcriptional regulators identified that belonged to a specific transcriptional family, with GntR being the most common followed by TetR. antiSMASH analysis was applied again for the unique genes and identified 6 potential secondary metabolite biosynthetic clusters. The antiSMASH identified clusters were predicted to encode three non-ribosomal peptide synthase (NRPS), one T1PKS, one terpene and a hybrid terpene-NRPS.

Among the unique genes are those associated with bacterial phage, and upon closer examination of the regions containing these genes reveals the location of two potential *S. rimosus* SR1 phage. It has been well established that *S. rimosus* SR1 contains two temperate phage, RP2 and RP3 (Hranueli *et al.,* 1979; Vesligaj *et al.,* 1981; Rausch *et al.,* 1993). Using amino acid sequence comparison CCT maps, targeted at the regions of the unique phage, prophage, transposable elements and plasmids SEED subsystem genes, revealed two regions of with little sequence identity against the other *Streptomyces* species (Figure 3.6). This is not particularly strange given the narrow host range of RP2 and RP3 (Rausch *et al.,* 1993). It should be noted that within these regions are accessory genes that are potentially common to multiple *Streptomyces* phage. Nucleotide blast alignments of *S. rimosus* SR1 chromosomal DNA with RP3 (accession number X80661) from *S. rimosus* R6-554 DNA gave a 99 % identity at

positions 3692470 – 3695238 bp. In addition, nucleotide blast alignments of the *attP* sequence of RP3 (accession number X67955) against the *S. rimosus* SR1 chromosome revealed two sites of homology located at positions 3695063 – 3695238 and 3636012 – 3636182, flanking the region predicted to be RP3 (Figure 3.3). Further to this the gene architectures were similar to that of the actinophage φC31 and φSASD1 (Smith *et al.,* 1999; Wang *et al.,* 2010), with the majority of the genes transcribed in the same direction, unlike bacterial genes, further supporting the identification of the RP2 and RP3 phage. Physical mapping predicted the distance between these two phage, within the *S. rimosus* R6-501 chromosome, to be around 1.3 Mbp (Pandza *et al.,* 1997), whereas the distance is around 1.2 Mbp within *S. rimosus* SR1. The observed size of RP2 and RP3 did not coincide with the predicted sizes of 64 kbp and 62 kbp (Rausch *et al.,* 1993), as the findings here suggest both phage are around 59 kbp in size, however the GC content was the same at 70 %.

Table 3.5 Functional SEED categories of the *S. rimosus* SR1 unique genes

SEED Subsystem	Total number of genes	Contribution to unique genome (%)
Cofactors, Vitamins, Prosthetic Groups, Pigments	21	0.92
Cell Wall and Capsule	52	2.28
Virulence, Disease, Defense and Resistance	27	1.18
Potassium Metabolism	5	0.22
Miscellaneous	0	0.00
Membrane Transport	31	1.36
Iron Acquisition	9	0.39
RNA Metabolism	19	0.83
Nucleosides and Nucleotides	15	0.66
Protein Metabolism	20	0.88
Cell Division and Cell Cycle	4	0.18
Motility Chemotaxis	1	0.04
Regulation and Cell Signalling	106	4.65
Secondary Metabolism	15	0.66
DNA Metabolism	22	0.96
Fatty Acids, Lipids, Isoprenoids	18	0.79
Nitrogen Metabolism	3	0.13
Dormancy and Sporulation	1	0.04
Respiration	16	0.70
Stress Response	14	0.61
Metabolism of Aromatic Compounds	0	0.00
Amino Acids and Derivatives	49	2.15
Sulphur Metabolism	3	0.13
Phosphorus Metabolism	0	0.00
Carbohydrates	29	1.27
Phages, Prophages, Transposable Elements, Plasmids	24	1.05
Hypothetical/Uncategorised	1776	77.89

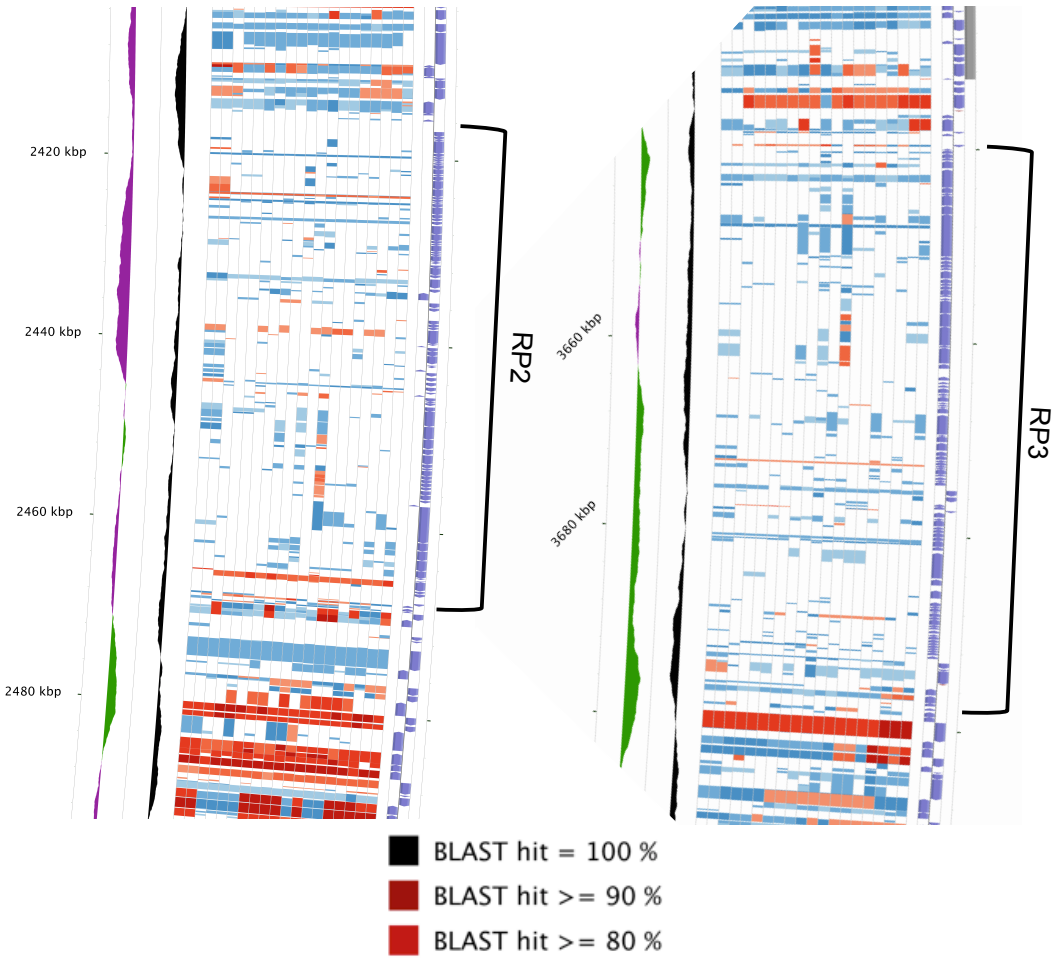


Figure 3.3 CCT map location of phages RP2 and RP3. Zoomed image of the CCT map shown in Figure 3.1 B, at coordinates 2450002 and 3670500 for RP2 and RP3 respectively. Coordinates correspond to *S. rimosus* SR1 genome position 2450002 bp and 3670500 bp. The CCT zoom command was set to 40 for RP2 and 50 for RP3. The pBLAST hits that are a light red/orange colour = 50 – 79 % identity, a blue colour = 0 – 49 % identity. As in Figure 3.1 B, rings from the outside inward: SR1 plus-strand, SR1 minus-strand, *S. rapamycinicus* NRRL 5491, *S. bingchenggensis* BCW 1, *S. violaceusniger* Tu 4113, *S. hygroscopicus jinggangensis* 5008, *S. hygroscopicus jinggangensis* TL01, *S. davawensis* JCM 4913, *S. avermitilis*, *S. scabiei* 87 22, *S. coelicolor* A3(2), *S. griseus* NBRC 13350, *S. venezuelae* ATCC 10712, *S. fulvissimus* DSM 40593, *S. collinus* Tu 365, *S. sp.* SirexAA, *S. sp.* PAMC26508, *S. flavogriseus* ATCC 3331, *S. albus* J1074, *S. cattleya* DSM 46488, *S. cattleya* NRRL 8057, GC content plot, GC skew and genomic position.

3.4 Core and Pan phylogenomic analysis of *Streptomyces*

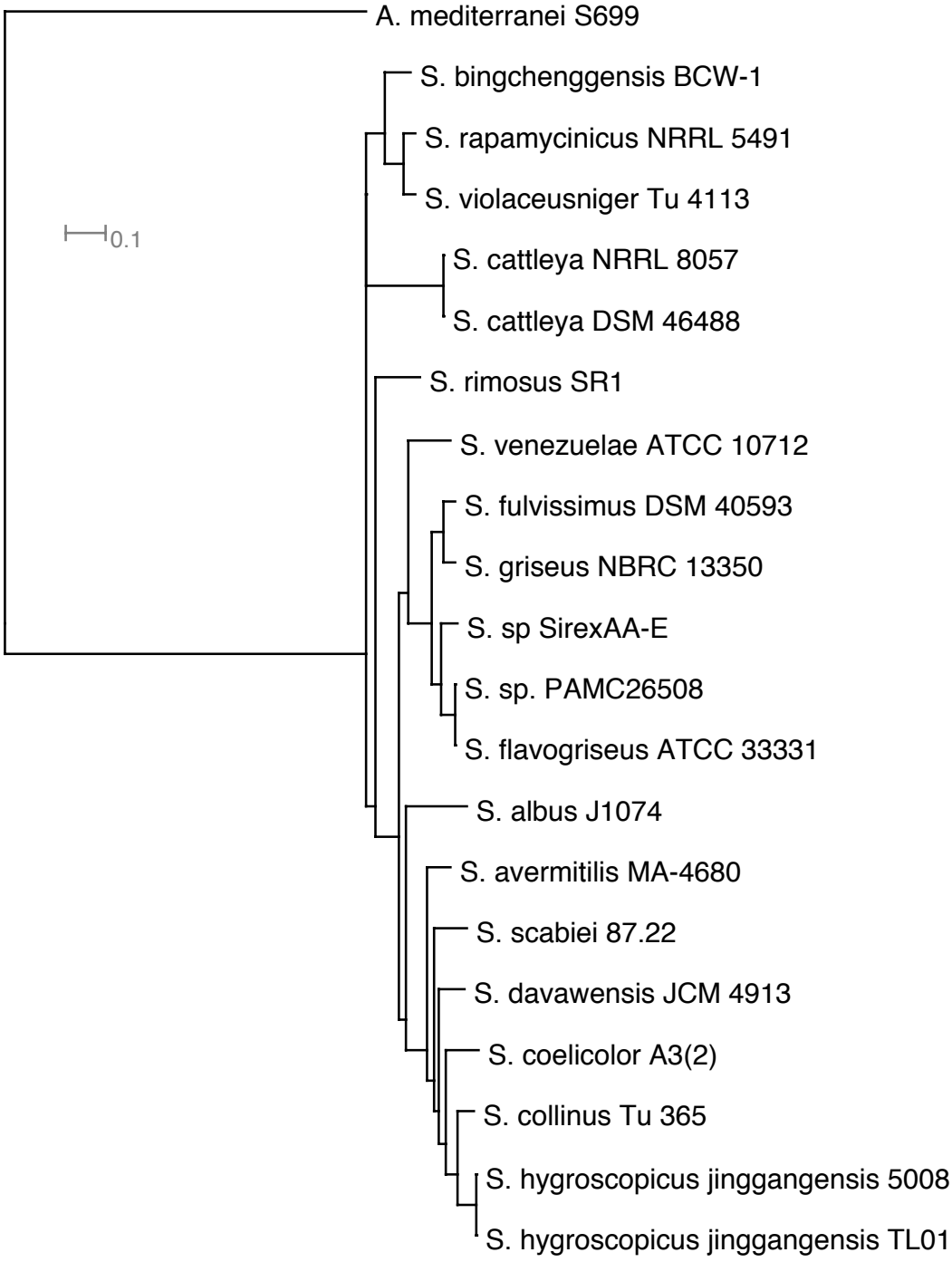
Phylogenetic trees based on either the core or pan proteome were constructed (Figure 3.4 A and B). Multilocus sequence analysis (MLSA) and 16S rRNA phylogenetic trees (Figure 3.4 C and D) were also produced in order to compare the pan and core proteome trees to standard methods of phylogeny (Rong *et al.,* 2013; Oren & Garrity 2013). The core proteome tree was produced from the 1879 homologous clustered proteins reverse translated and then concatenated for each species, followed by alignment using Parsnp (Treangen *et al.,* 2014). Parsnp produces a guide tree based on core single nucleotide polymorphisms, and therefore bootstrapping could not be applied. Using the GET_HOMOLOGUES script a reduced binary format of a pangenomic matrix was produced. This matrix contained information regarding the presence/absence of clusters for each Streptomycete in .phylip format, which was analysed using SplitsTree4 (Huson & Bryant 2006) to generate a rooted neighbour joining (NJ); (Saitou & Nei 1987) tree. The 16S rRNA and MLSA trees were produced from MUSCLE alignments (Edgar 2004), produced in MEGA (Tamura *et al.,* 2011), of each 16S rRNA DNA sequence or concatenated amino acid sequence of AtpD, GyrB, RecA, RpoB and TrpB proteins. NJ trees were then created using SplitsTree4. For all trees *A. mediterranei* S699 (accession no. CP003729) was used as an outlier. The core proteome tree suggests three distinct groups, with the pan proteome tree predicting four groups. Both trees suggest that *S. sp* PAMC26508 and *S. flavogriseus* ATCC 33331 are very close neighbours, being almost indistinct in the core tree (Figure 3.4 A and B). The overall topology of both trees is almost identical, with the exception of *S. scabiei* 87.22 and *S. albus* J1704, as these two species are found within the same branch of the core proteome tree but separate on the pan proteome tree (Figure 3.4 A and B). The pan proteome tree also predicted a greater phylogenetic distance between the different *Streptomyces* species than the core. Furthermore the pan proteome tree (Figure 3.4 B) shows increased definition, as can be seen with both strains of

S. cattleya and *S. hygroscopicus jinggangensis*, as distinct nodes are formed between different strains of the same species (Figure 3.4 A and B). The exact differences between the strains of *S. cattleya*, in this context, are only 526 proteins, with only 195 proteins varying between the *S. hygroscopicus jinggangensis* strains. Once again three distinct groups can be seen in the MLSA tree; however, when compared to the pan proteome tree, both *S. albus* J1704 and *S. venezuelae* ATCC 10712 are closer to the *S. coelicolor* A3(2) rather than being within the same branch as *S. griseus* NBRC 13350 (Figure 3.4 B and C). This is closer to the topology of the core proteome tree as *S. albus* J1704 is placed within the same branch as *S. coelicolor* A3(2) (Figure 3.4 A). The MLSA tree fails to distinguish between the *S. cattleya* and *S. hygroscopicus jinggangensis* strains, and only slightly reveals separate nodes for *S. sp* PAMC26508 and *S. flavogriseus* ATCC 33331 (Figure 3.4 C). The 16S rRNA tree was the least accurate of the phylogenetic trees, as its bootstrapping scores were the lowest and most varied compared to the other trees (Figure 3.4 D). Furthermore, while all the other trees were relatively consistent with one another, the 16S rRNA tree varied the most (Figure 3.4 D). A recent study by Labeda *et al* compared the 16S rRNA of 615 species from the family *Streptomycetaceae*, 582 of which were *Streptomyces* species (Labeda *et al.,* 2011). Of the species studied here only 8 were missing from the Labeda analysis, *S. binchenggensis* BCW-1, both *S. cattleya* strains, *S. davawensis* JCM 4913, both *S. hygroscopicus subsp jinggangensis* strains, *S. sp.* PAMC26508 and *S. sp.* SirexAA-E. The Labeda analysis found that the clades containing *S. rapamycinicus* NRRL 5491 and *S. violaceusniger* Tu 4113 were closely related, consistent with the trees produced here. Examining clustering, when considering morphology of each strain, using the analysis of Labeda *et al*, finds that species that cluster together within the core, pan, MLSA and 16S rRNA trees generally share spore morphology.

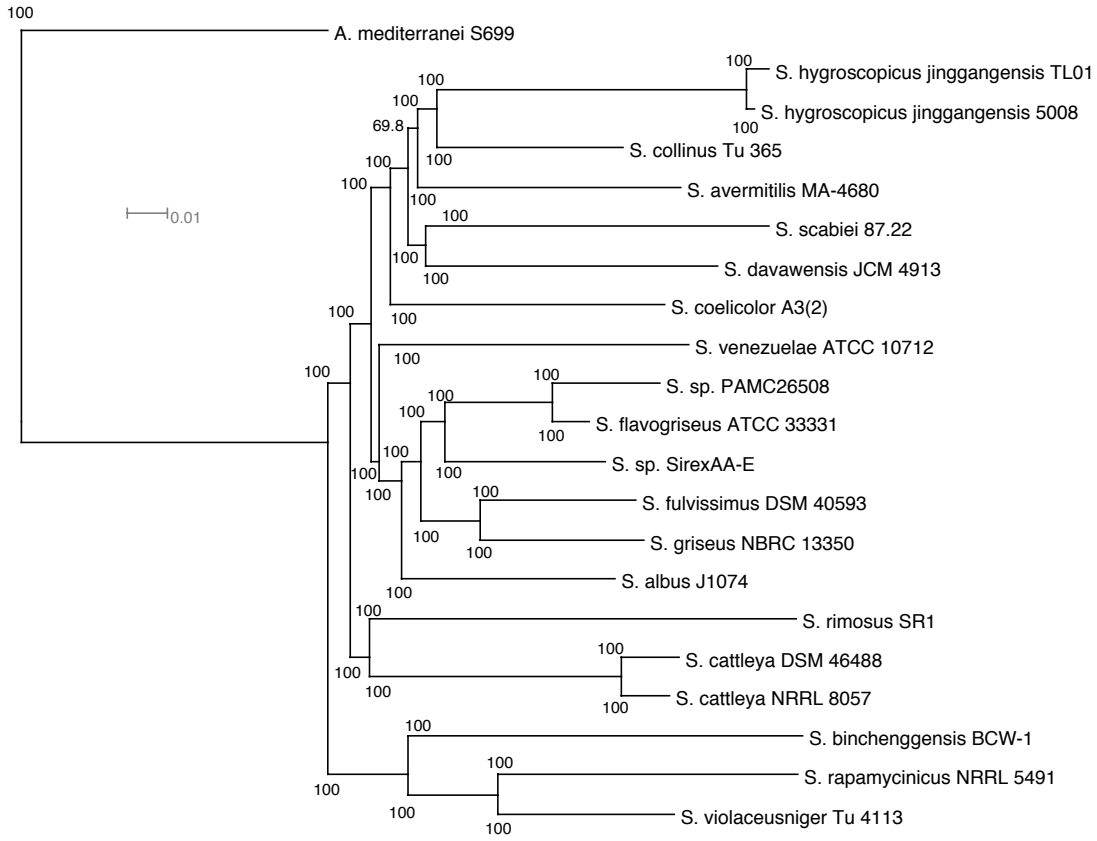
In addition to the phylogenetic tree a pairwise proteome plot was produced. Pairwise similarity was determined by Sij/Gi as described by Doroghazi and Metcalf, 2013, where Sij represents all similar proteins

between proteome i and j, and Gi is the total number of proteins within proteome i (Doroghazi & Metcalf 2013). Comparing the phylogenetic trees to a pairwise proteome similarity plot (Figure 3.5) shows correlation between all the maps and whole proteome similarity for defined pairs i.e *S. rapamycinicus* NRRL 5491 and *S. violaceusniger* Tu 4113 (Figure 3.4 A, B C and D). Those trees with three or four distinct groups (Figure 3.4 A, B and C) also correlate well with the pairwise proteome similarity plot, as the species within those groups generally share the most similarity on the plot. Overall the core genome and pan proteome phylogenetic trees give comparable results to the MLSA tree. In addition the pan tree gives greater definition for those species or strains that are very similar, whereas the MLSA and core trees do not distinguish between members of the same species. Taken together the pan proteome tree appears to produce a more precise phylogeny.

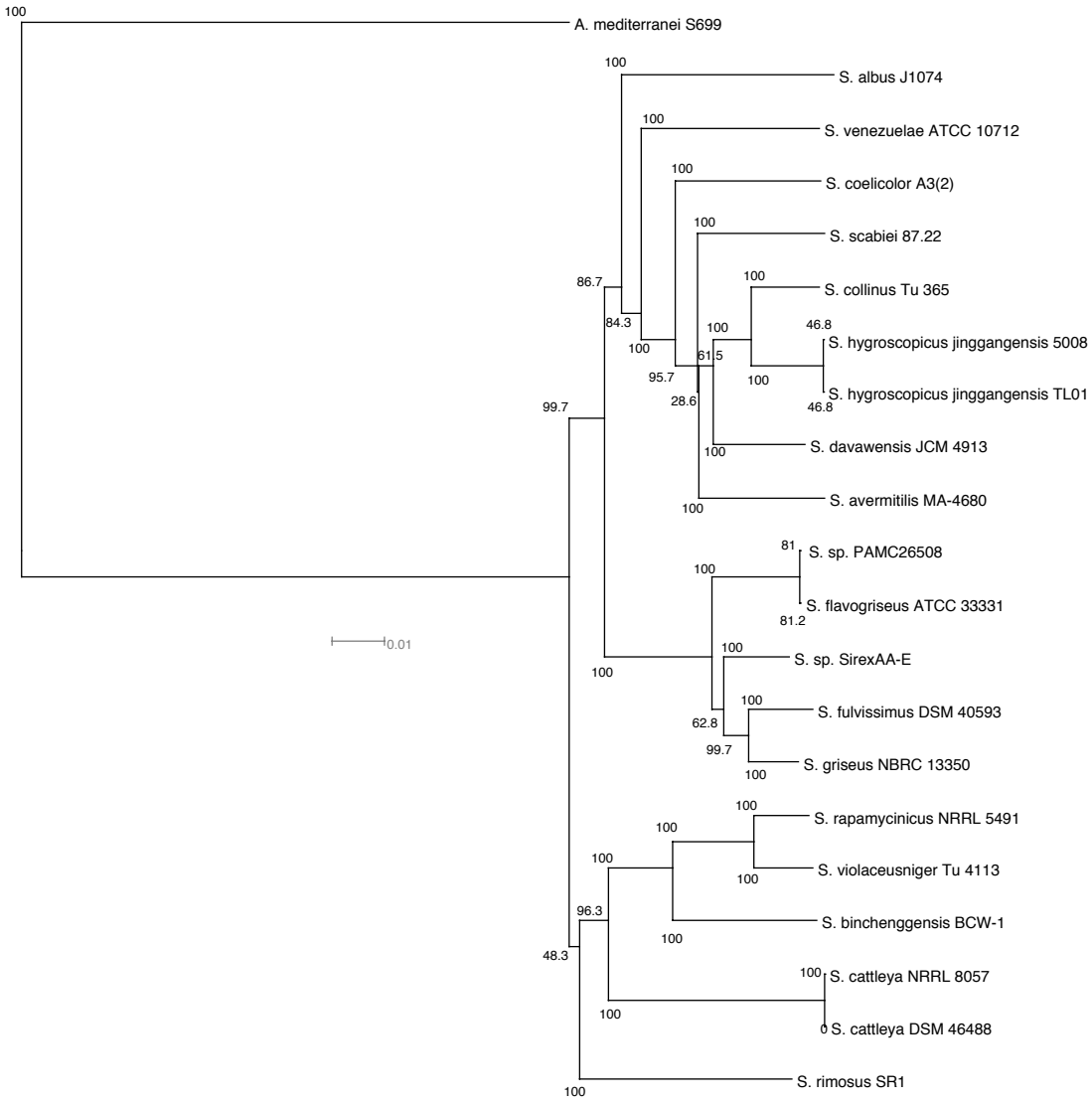
A


A. mediterranei S699
S. bingchenggensis BCW-1
S. rapamycinicus NRRL 5491
S. violaceusniger Tu 4113
0.1
S. cattleya NRRL 8057
S. cattleya DSM 46488
S. rimosus SR1
S. venezuelae ATCC 10712
S. fulvissimus DSM 40593
S. griseus NBRC 13350
S. sp SirexAA-E
S. sp. PAMC26508
S. flavogriseus ATCC 33331
S. albus J1074
S. avermitilis MA-4680
S. scabiei 87.22
S. davawensis JCM 4913
S. coelicolor A3(2)
S. collinus Tu 365
S. hygroscopicus jinggangensis 5008
S. hygroscopicus jinggangensis TL01

B


A. mediterranei S699
S. hygroscopicus jinggangensis TL01
S. hygroscopicus jinggangensis 5008
S. collinus Tu 365
S. avermitilis MA-4680
S. scabiei 87.22
S. davawensis JCM 4913
S. coelicolor A3(2)
S. venezuelae ATCC 10712
S. sp. PAMC26508
S. flavogriseus ATCC 33331
S. sp. SirexAA-E
S. fulvissimus DSM 40593
S. griseus NBRC 13350
S. albus J1074
S. rimosus SR1
S. cattleya DSM 46488
S. cattleya NRRL 8057
S. binchenggensis BCW-1
S. rapamycinicus NRRL 5491
S. violaceusniger Tu 4113
0.01

C


A. mediterranei S699
S. albus J1074
S. venezuelae ATCC 10712
S. coelicolor A3(2)
S. scabiei 87.22
S. collinus Tu 365
S. hygroscopicus jinggangensis 5008
S. hygroscopicus jinggangensis TL01
S. davawensis JCM 4913
S. avermitilis MA-4680
S. sp. PAMC26508
S. flavogriseus ATCC 33331
S. sp. SirexAA-E
S. fulvissimus DSM 40593
S. griseus NBRC 13350
S. rapamycinicus NRRL 5491
S. violaceusniger Tu 4113
S. binchenggensis BCW-1
S. cattleya NRRL 8057
S. cattleya DSM 46488
S. rimosus SR1
0.01

D

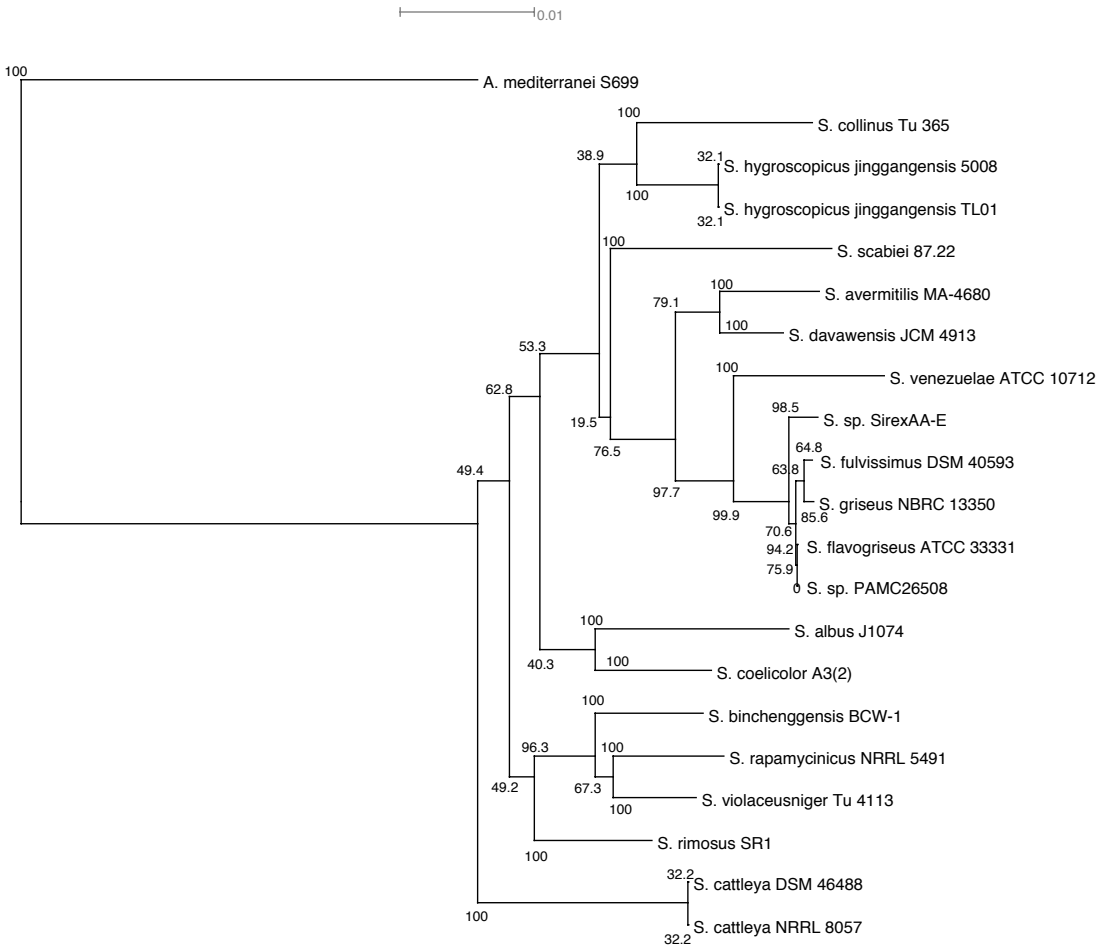


Figure 3.4 Phylogeny of the *Streptomyces* species. Neighbouring joining trees were generated in SplitsTree4, following MUSCLE alignment in MEGA version 5.2.2 for 16S rRNA and MLSA trees, and bootstrapped with 1000 replicates. **A.** Core phylogenetic tree. **B.** Pan phylogenetic tree. **C.** MLSA proteins tree. The AtpD, GyrB, RecA, RpoB and TrpB proteins were identified for each *Streptomyces* species and concatenated. **D.** 16S rRNA tree.

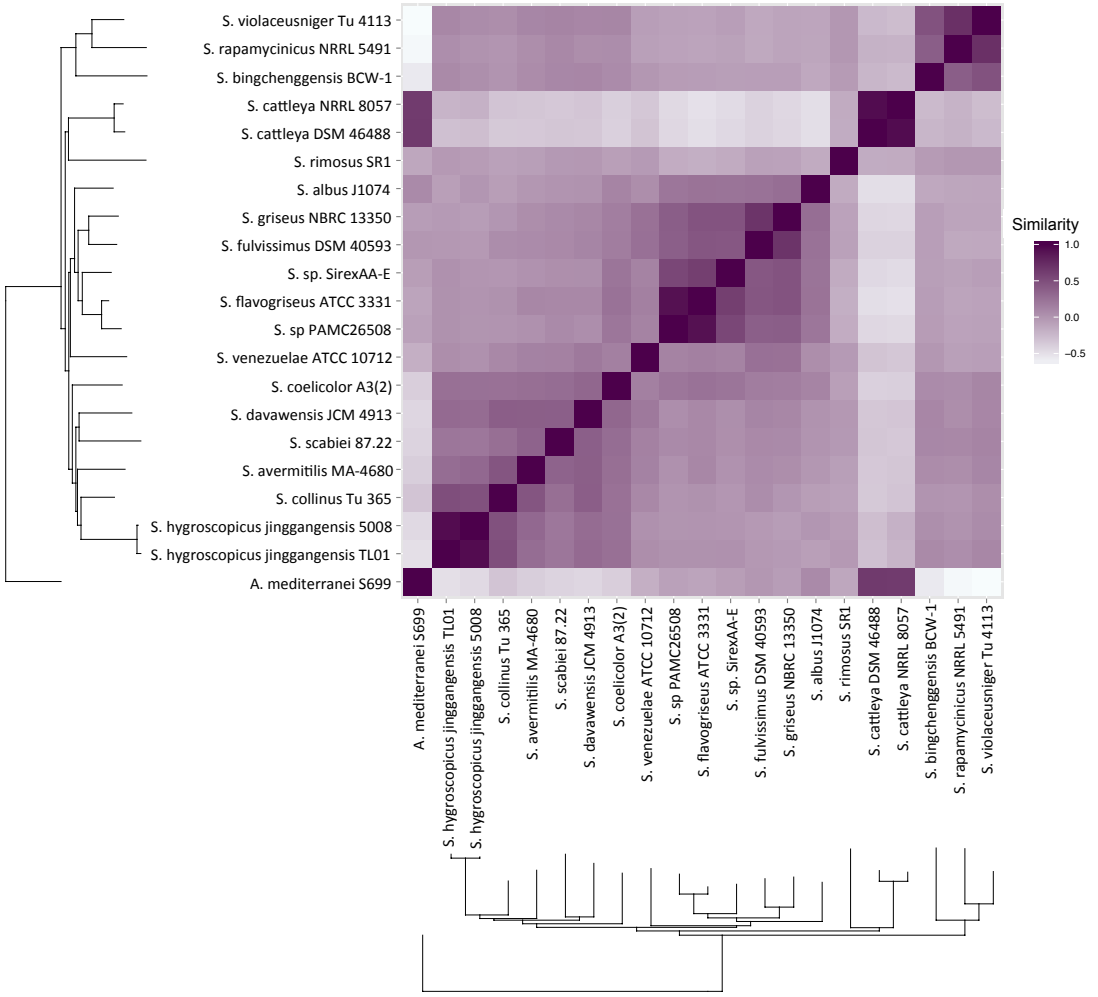


Figure 3.5 Putative whole proteome similarity. The order for the comparison is based on the pan proteome phlogeny. The diagonal line from bottom left to top right represents comparisons of a proteome with itself. Whole proteome similarity was determined by the number of similar proteins between two proteomes, used as the numerator, with the proteome represented by each row used as the denominator, for each comparison. The pan proteome phylogenetic tree is aligned at the sides.

3.5 Additional features of the accessory genome of *Streptomyces*

3.5.1 Examination of the secondary metabolites associated with each Streptomycete

S. rimosus SR1 is known to produce the secondary metabolites OTC and rimocidin (Phornphisutthimas *et al.,* 2010); however, little is known about other secondary metabolites that *S. rimosus* SR1 can produce. antiSMASH version 2.0 was used to identify additional biosynthetic clusters within the *S. rimosus* SR1 genome and predicted a total of 54 potential clusters (Figure 3.6). The predicted biosynthetic clusters were primarily within the arms of the chromosome rather than the core central region (Figure 3.6), with each region being associated with non-essential and essential genes respectively (Bentley *et al.,* 2002). Further to this, the majority were towards the end of the chromosome, it is also within this region that there is the least protein sequence identity between *S. rimosus* SR1 proteins and the rest of the *Streptomyces*. Of the biosynthetic clusters predicted only the OTC cluster had been fully identified previously (Butler *et al.,* 1989; Binnie *et al.,* 1989; Pandza *et al.,* 1997). NRPS clusters were the most abundant biosynthetic clusters identified, with T1PKS and terpenes being the second most. However the only T2PKS identified was the OTC cluster. One of the T1PKS predicted was that of rimocidin, and until now only the putative loading domain of the pathway and modifying proteins had been discovered in *S. diastaticus* var. 108 (Seco *et al.,* 2004). Further to this a hybrid NRPS-T1PKS, upstream of the rimocidin biosynthetic cluster, was identified and from previous work was concluded to be associated with polyketide mnp (Seco *et al.,* 2004). The previously deduced PKS domains Mnp2, Mnp4 and Mnp5 lie very close to a potential NRPS, and therefore antiSMASH predicted a hybrid NRPS-T1PKS. The polyketide product of the *mnp* PKS has yet to be identified, so whether the product is a simple polyketide or a hybrid is unknown. antiSMASH cluster annotation does not give a precise

representation, as there are overlapping clusters, however this may be due to the draft nature of the *S. rimosus* SR1 genome.

antiSMASH analysis was applied to the *Streptomyces* genomes (Table 3.1) and resulted in an overview of the secondary metabolite clusters present in each species (Figure 3.7 A). The most common secondary metabolite clusters predicted were NRPS, T1PKS and bacteriocin. *S. albus* J1704 had the least number of predicted biosynthetic clusters, with only 24 and *S. rimosus* SR1 had the most with 54 (Figure 3.7 A). Plotting the total number of secondary metabolite biosynthetic clusters against genome size shows a trend for increasing cluster number with increasing genome size (Figure 3.7 B). Further to this *S. coelicolor* A3(2) has a genome almost 2 Mbp larger than *S. albus* J1704 but has only 2 more biosynthetic clusters. Though it should be noted that *S. bingchenggensis* BCW 1, whose genome is the largest examined, has only one less predicted cluster when compared to *S. rimosus* SR1 (Figure 3.7 A). *S. bingchenggensis* BCW 1 has the most hybrid biosynthetic clusters predicted, with *S. coelicolor* A3(2) predicted to have the least.

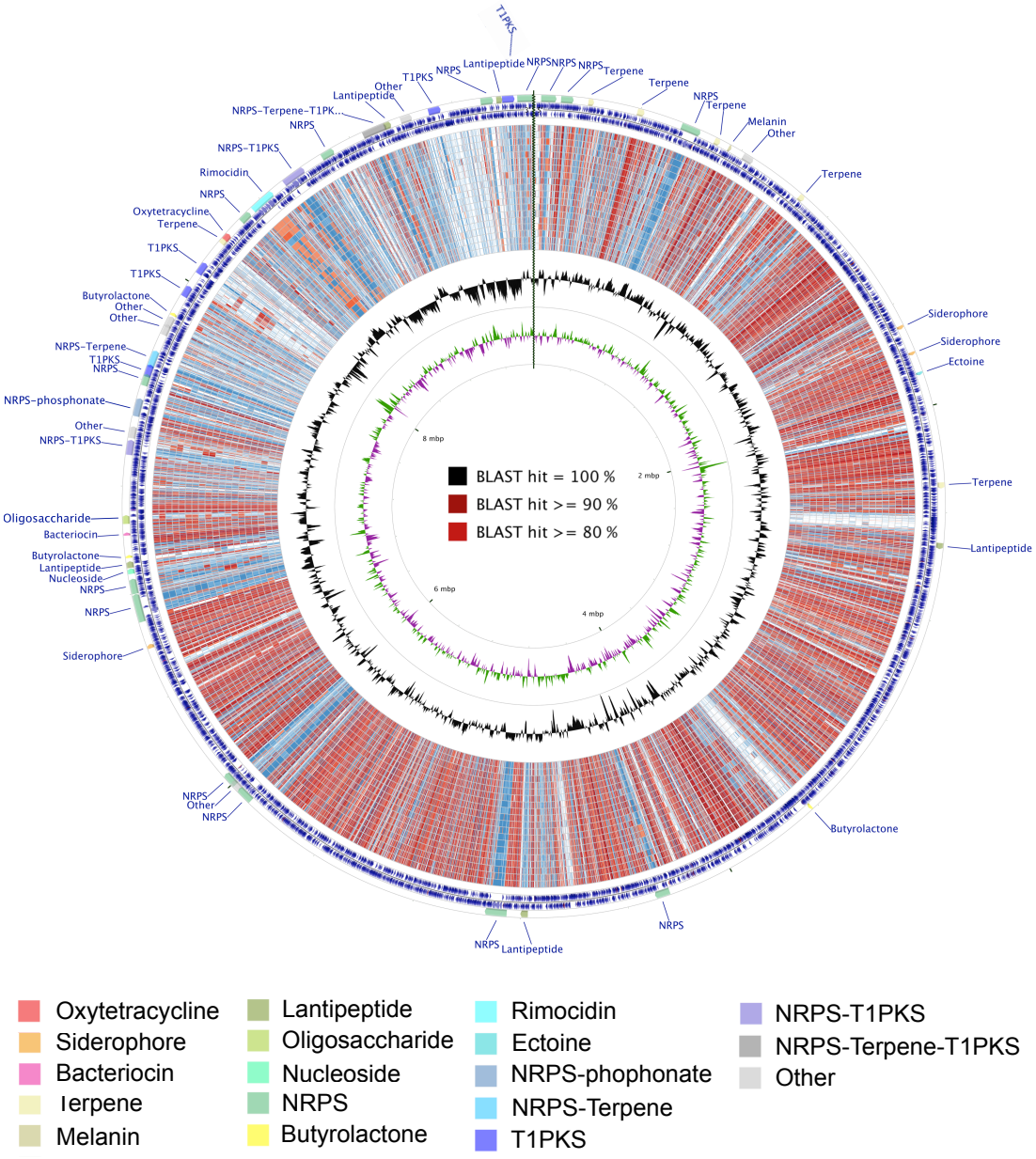


Figure 3.6 The location of *S. rimosus* SR1 biosynthetic clusters. pBLAST CCT map as seen in Figure 3.1 B, with additional tags representing the location of biosynthetic clusters as identified by antiSMASH. BLAST hits that are a light red/orange colour = 50 – 79 % identity, a blue colour = 0 – 49 % identity.

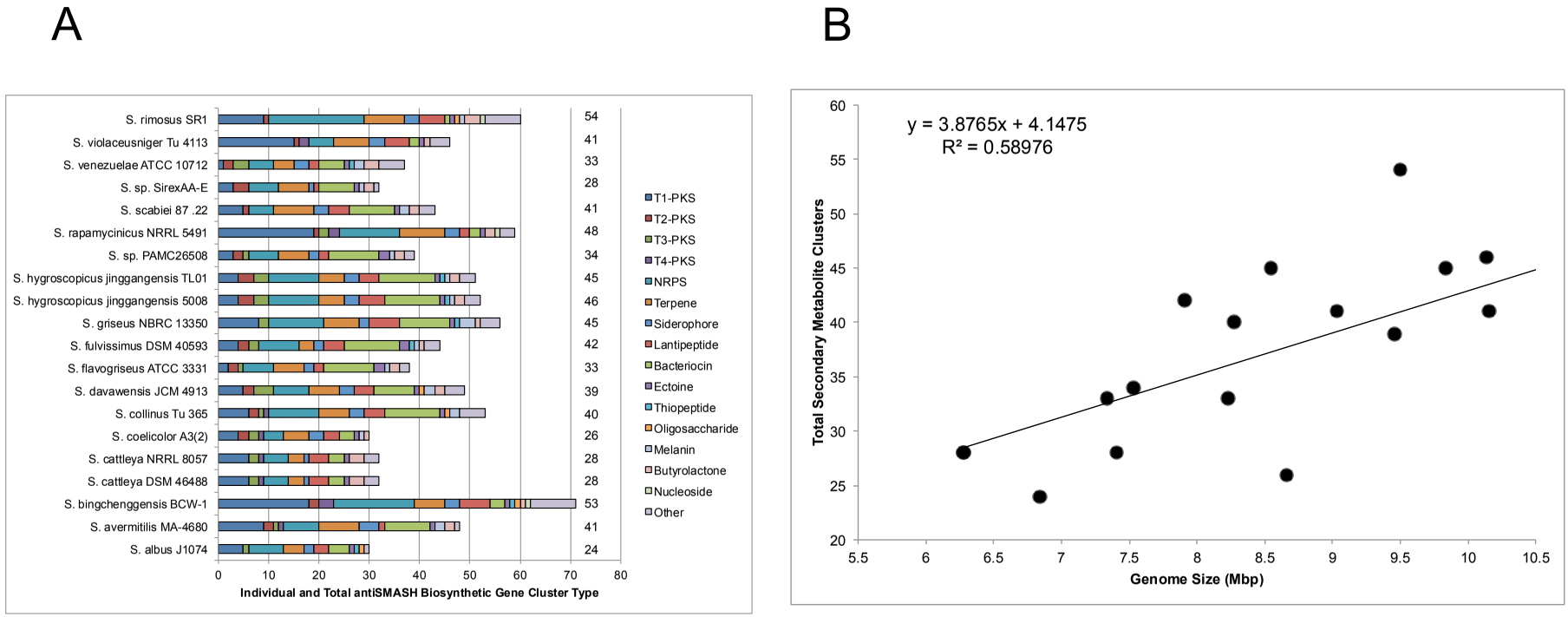


Figure 3.7 Overview of biosynthetic clusters for *Streptomyces* species. A. antiSMASH predicted biosynthetic cluster for each of the *Streptomyces* species examined. Hybrid biosynthetic clusters are counted as individual clusters, for clarity total predicted biosynthetic clusters are labeled at the end of each bar. **B.** Plot of total biosynthetic clusters identified versus genome size for each of the *Streptomyces*. Within the chart the equation of the trendline is displayed as well as the coefficient of determination (R^2).

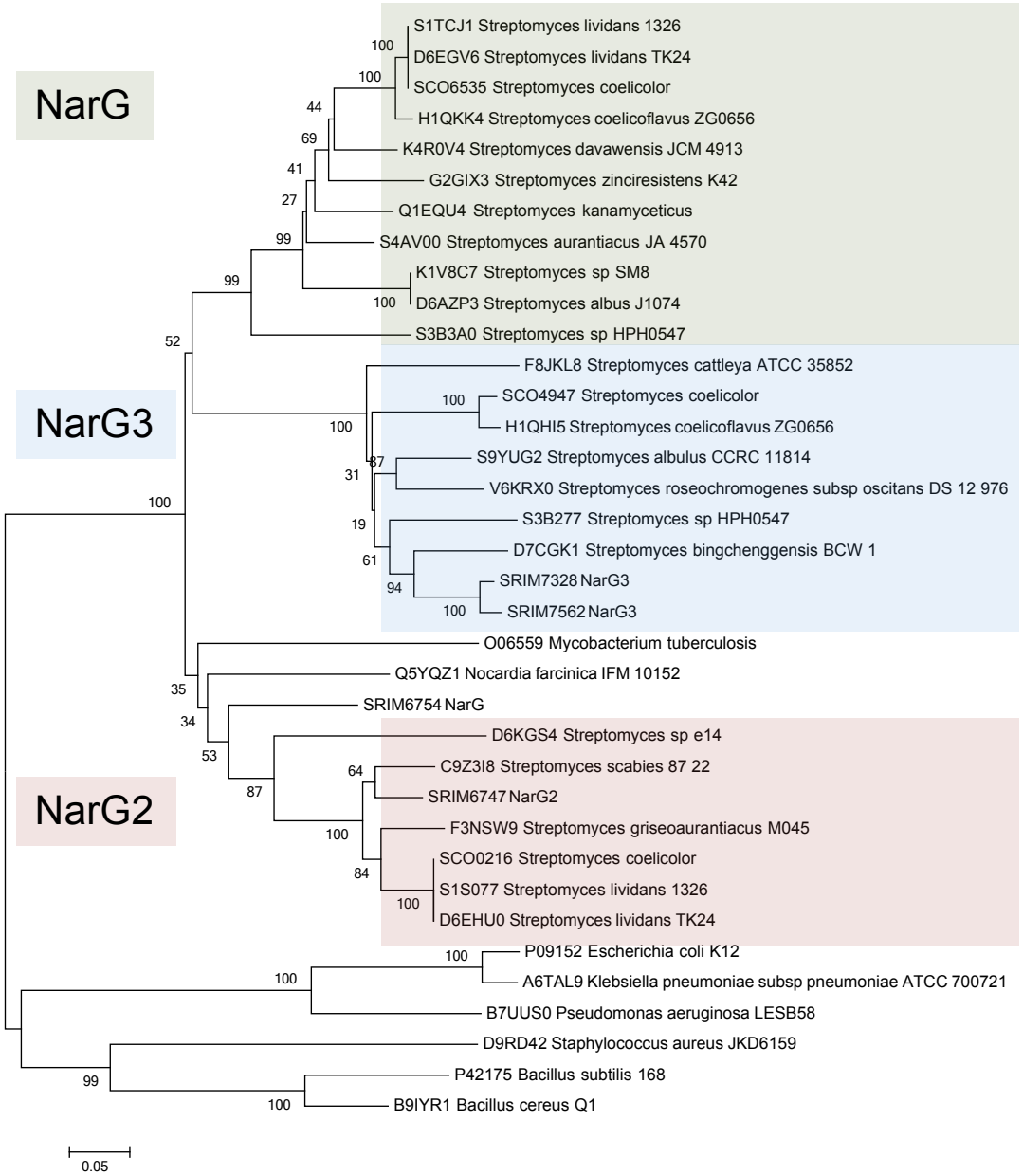
3.5.2 *S. rimosus* SR1 has 4 sets of respiratory nitrate reductase genes

Four sets of respiratory nitrate reductase (Nar) genes were found within the accessory genome of *S. rimosus* SR1, this in itself was interesting as only *S. coelicolor* A3(2) was known to contain multiple copies of the *nar* genes, and even then only three sets (van Keulen *et al.,* 2005). Of the 20 *Streptomyces* species analysed, only 6 had a complete set of *nar* genes (*narGHIJ*) and only 2 had more than one set. In order to determine how closely related each set of *nar* genes were and to elucidate which species had which set, phylogenetic trees were produced (Figure 3.8). As would be expected, the *nar* genes from the actinobacteria cluster together within the trees, with members of the proteobacteria and firmicutes grouped separately (Figure 3.8). Each tree displays distinct subgroups for Nar1, Nar2 and Nar3 type, highlighting that, although these proteins perform a similar function they may operate under different physiological conditions. For example, the Nar1 set functions exclusively in resting spores, and it is predicted that the Nar2 set would function in both spores and hyphae, with Nar3 functioning solely for hyphae (Fischer *et al.,* 2010; Fischer *et al.,* 2013).

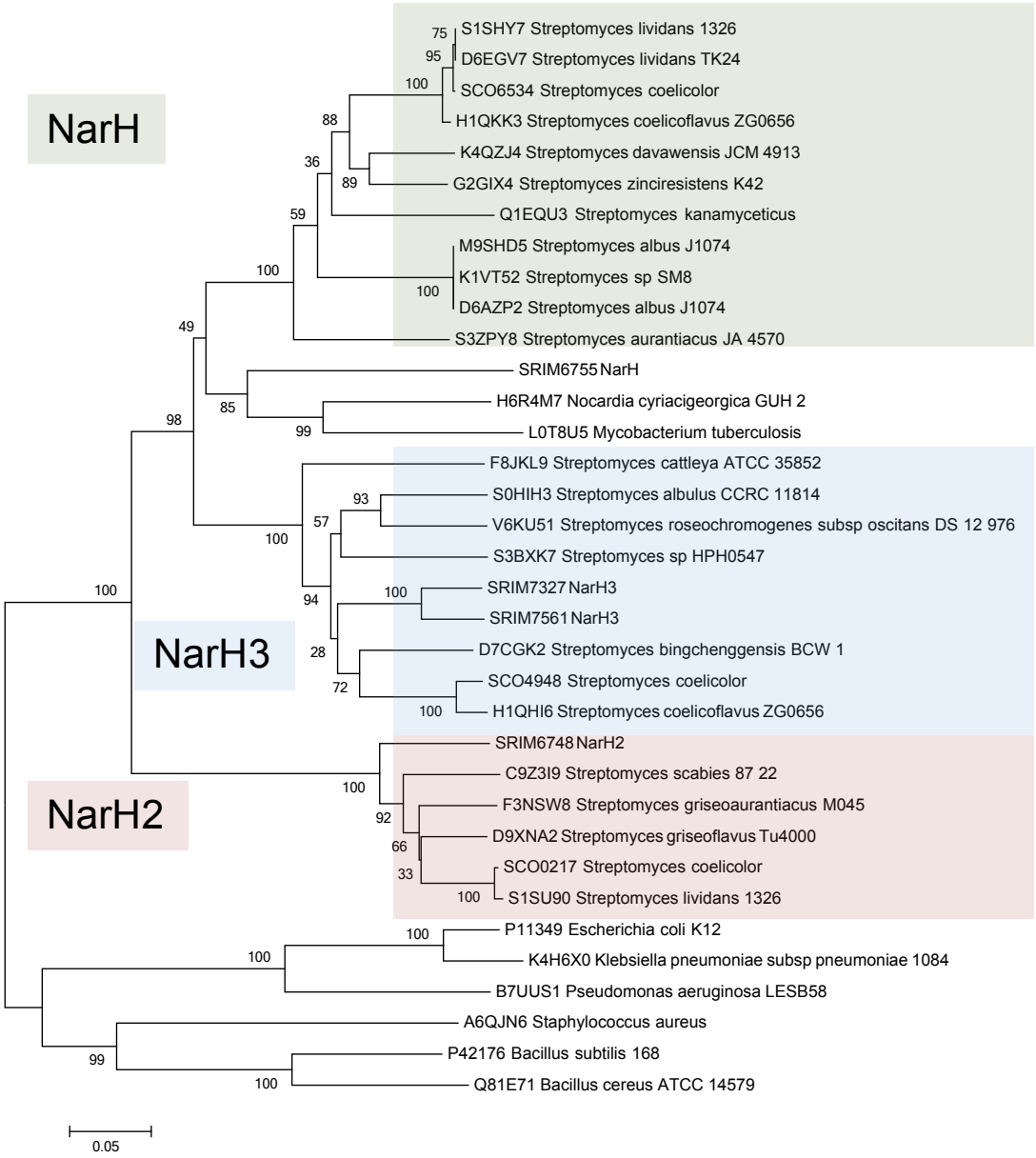
The set of *nar* genes belonging to *S. albus* J1074 and *S. davawensis* JCM 4913 were of the Nar1 type, with the Nar gene set of *S. scabiei* 87 22 being Nar2 and a Nar3 set belonging to *S. binchenggensis* BCW-1 (Figure 3.8). The phylogenetic analysis revealed that *S. rimosus* SR1 has two sets of the Nar3 gene type, one set of Nar2 and a distantly related set of the Nar1 type, with the exception SRIM6754 which is more closely related to NarG2 (Figure 3.8 A). Of the two genes that encode NarG3 type proteins, SRIM7562 appears to be truncated, missing the central portion of the NarG3 protein, this is a possible explanation for the extra copy of the *nar3* genes. Upstream of *SRIM6754* is *SRIM6751*, a homologue of *SCO0213* which encodes a nitrate/nitrite exchanger *narK2* (van Keulen *et al.,* 2005; Zheng *et al.,* 2013). Within the genome of *S. coelicolor* A3(2) there are 3 *narK* like genes with *SCO2959* and *SCO3809*, termed here *narK1* and *narK3* respectively, in addition to *SCO0213* (Bentley *et al.,* 2002; Fischer *et al.,*

2013). All of the *Streptomyces* species examined here have at least one copy of a *narK* gene. *narK2* lies upstream of the *nar2* operon in the genomes of both *S. coelicolor* A3(2) and *S. scabiei* 87 22, however in *S. rimosus* SR1 *SRIM6751* lies upstream of the potential *nar1* operon and downstream of the *nar2* operon. *S. davawensis* JCM 4913 along with *S. griseus* NBRC 13350, *S. venezuelae* ATCC 10712, *S. fulvissimus* DSM 40593, *S. sp*. SirexAA E, *S. sp*. PAMC26508, *S. flavogriseus* ATCC 3331 have homologues of *SCO0213* but no set of *nar2* genes. Further to this all of the *Streptomyces* species examined here have a homologue of *SCO2959* (*narK1*), apart from *S. sp* SirexAA E, *S. albus* J1074, *S. cattleya* DSM 46488, and *S. cattleya* NRRL 8057. *S. albus* J1704 only has one homologue of *SCO3809* (*nark3*). This indicates that although these species do not have the ability to reduce nitrate they can still transport nitrate and nitrite. In *S. bingchenggensis* BCW 1, *S. davawensis* JCM 4913, *S. rimosus* SR1, *S. scabiei* 87 22, and *S. coelicolor* A3(2), the *narK1* gene, lies separately from the *nar* operons. The phylogenetic analysis further indicated that *S. hygroscopicus jinggangensis* 5008 and *S. hygroscopicus jinggangensis* TL01 have a single copy of the *narJ* gene (Figure 3.8 D).

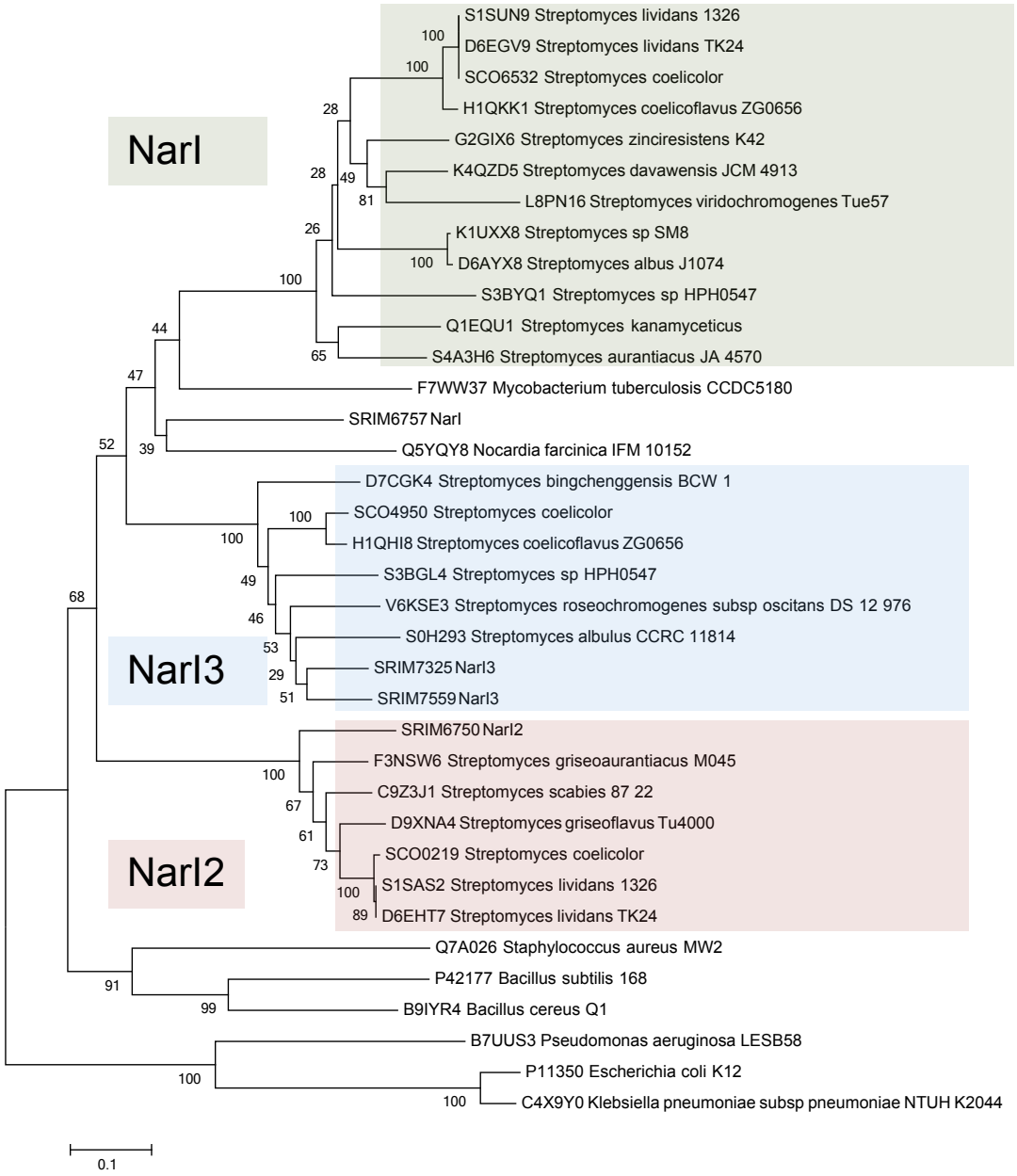
A


NarG
NarG3
NarG2
S1TCJ1 Streptomyces lividans 1326
D6EGV6 Streptomyces lividans TK24
SCO6535 Streptomyces coelicolor
H1QKK4 Streptomyces coelicoflavus ZG0656
K4R0V4 Streptomyces davawensis JCM 4913
G2GIX3 Streptomyces zinciresistens K42
Q1EQU4 Streptomyces kanamyceticus
S4AV00 Streptomyces aurantiacus JA 4570
K1V8C7 Streptomyces sp SM8
D6AZP3 Streptomyces albus J1074
S3B3A0 Streptomyces sp HPH0547
F8JKL8 Streptomyces cattleya ATCC 35852
SCO4947 Streptomyces coelicolor
H1QHI5 Streptomyces coelicoflavus ZG0656
S9YUG2 Streptomyces albulus CCRC 11814
V6KRX0 Streptomyces roseochromogenes subsp oscitans DS 12 976
S3B277 Streptomyces sp HPH0547
D7CGK1 Streptomyces bingchenggensis BCW 1
SRIM7328 NarG3
SRIM7562 NarG3
O06559 Mycobacterium tuberculosis
Q5YQZ1 Nocardia farcinica IFM 10152
SRIM6754 NarG
D6KGS4 Streptomyces sp e14
C9Z3I8 Streptomyces scabies 87 22
SRIM6747 NarG2
F3NSW9 Streptomyces griseoaurantiacus M045
SCO0216 Streptomyces coelicolor
S1S077 Streptomyces lividans 1326
D6EHU0 Streptomyces lividans TK24
P09152 Escherichia coli K12
A6TAL9 Klebsiella pneumoniae subsp pneumoniae ATCC 700721
B7UUS0 Pseudomonas aeruginosa LESB58
D9RD42 Staphylococcus aureus JKD6159
P42175 Bacillus subtilis 168
B9IYR1 Bacillus cereus Q1
0.05

B


NarH
NarH3
NarH2
S1SHY7 Streptomyces lividans 1326
D6EGV7 Streptomyces lividans TK24
SCO6534 Streptomyces coelicolor
H1QKK3 Streptomyces coelicoflavus ZG0656
K4QZJ4 Streptomyces davawensis JCM 4913
G2GIX4 Streptomyces zinciresistens K42
Q1EQU3 Streptomyces kanamyceticus
M9SHD5 Streptomyces albus J1074
K1VT52 Streptomyces sp SM8
D6AZP2 Streptomyces albus J1074
S3ZPY8 Streptomyces aurantiacus JA 4570
SRIM6755 NarH
H6R4M7 Nocardia cyriacigeorgica GUH 2
L0T8U5 Mycobacterium tuberculosis
F8JKL9 Streptomyces cattleya ATCC 35852
S0HIH3 Streptomyces albulus CCRC 11814
V6KU51 Streptomyces roseochromogenes subsp oscitans DS 12 976
S3BXK7 Streptomyces sp HPH0547
SRIM7327 NarH3
SRIM7561 NarH3
D7CGK2 Streptomyces bingchenggensis BCW 1
SCO4948 Streptomyces coelicolor
H1QHI6 Streptomyces coelicoflavus ZG0656
SRIM6748 NarH2
C9Z3I9 Streptomyces scabies 87 22
F3NSW8 Streptomyces griseoaurantiacus M045
D9XNA2 Streptomyces griseoflavus Tu4000
SCO0217 Streptomyces coelicolor
S1SU90 Streptomyces lividans 1326
P11349 Escherichia coli K12
K4H6X0 Klebsiella pneumoniae subsp pneumoniae 1084
B7UUS1 Pseudomonas aeruginosa LESB58
A6QJN6 Staphylococcus aureus
P42176 Bacillus subtilis 168
Q81E71 Bacillus cereus ATCC 14579
0.05

C


NarI
NarI3
NarI2
S1SUN9 Streptomyces lividans 1326
D6EGV9 Streptomyces lividans TK24
SCO6532 Streptomyces coelicolor
H1QKK1 Streptomyces coelicoflavus ZG0656
G2GIX6 Streptomyces zinciresistens K42
K4QZD5 Streptomyces davawensis JCM 4913
L8PN16 Streptomyces viridochromogenes Tue57
K1UXX8 Streptomyces sp SM8
D6AYX8 Streptomyces albus J1074
S3BYQ1 Streptomyces sp HPH0547
Q1EQU1 Streptomyces kanamyceticus
S4A3H6 Streptomyces aurantiacus JA 4570
F7WW37 Mycobacterium tuberculosis CCDC5180
SRIM6757 NarI
Q5YQY8 Nocardia farcinica IFM 10152
D7CGK4 Streptomyces bingchenggensis BCW 1
SCO4950 Streptomyces coelicolor
H1QHI8 Streptomyces coelicoflavus ZG0656
S3BGL4 Streptomyces sp HPH0547
V6KSE3 Streptomyces roseochromogenes subsp oscitans DS 12 976
S0H293 Streptomyces albulus CCRC 11814
SRIM7325 NarI3
SRIM7559 NarI3
SRIM6750 NarI2
F3NSW6 Streptomyces griseoaurantiacus M045
C9Z3J1 Streptomyces scabies 87 22
D9XNA4 Streptomyces griseoflavus Tu4000
SCO0219 Streptomyces coelicolor
S1SAS2 Streptomyces lividans 1326
D6EHT7 Streptomyces lividans TK24
Q7A026 Staphylococcus aureus MW2
P42177 Bacillus subtilis 168
B9IYR4 Bacillus cereus Q1
B7UUS3 Pseudomonas aeruginosa LESB58
P11350 Escherichia coli K12
C4X9Y0 Klebsiella pneumoniae subsp pneumoniae NTUH K2044
0.1

D

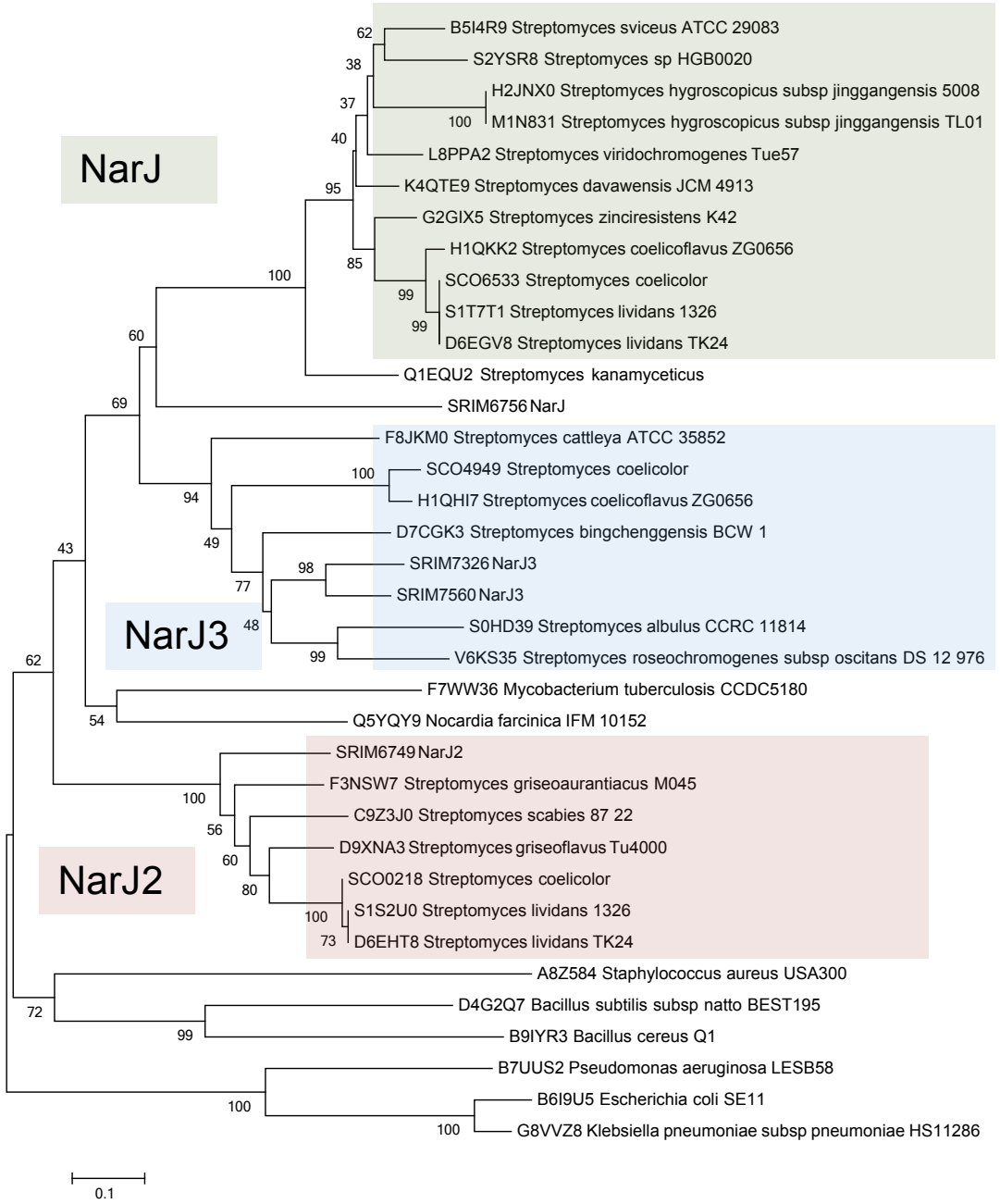


Figure 3.8 Phylogenetic Trees of Nar Proteins. The phylogenetic trees are NJ trees generated in MEGA version 5.2.2, following MUSCLE alignment, and bootstrapped with 1000 replicates. The Uniprot accession number for each protein is listed before the species name, with the exception of *S. coelicolor* A3(2), whose protein gene name is given. **A.** NarG proteins. **B.** NarH proteins. **C.** NarI proteins. **D.** NarJ proteins.

3.6 Discussion

Pfizer originally isolated *S. rimosus* SR1 from soil in 1950 (Finlay & Hobby 1950) and has since then been studied extensively (Petković *et al.,* 2006). The genome sequence of *S. rimosus* SR1 revealed a genome size of 9.5 Mbp, with a GC content of 71.88 % and 8651 protein-coding genes (Pethick *et al.,* 2013). It was decided to analyse the core and pan genome of *Streptomyces* to better understand the differences and similarities between the *Streptomyces* and *S. rimosus* SR1. Although around 130 *Streptomyces* genomes are available at this time, it was decided to limit this analysis to complete genomes, due to computing power available and the general variation in annotation quality. Though for a more extensive analysis more genomes should be examined. Previous core/pan genome analysis of 5 complete *Streptomyces* genomes resulted in 17,362 clusters within the pan genome which consisted of 3096 core gene clusters, 5066 accessory clusters and 9200 unique gene clusters (Zhou *et al.,* 2012b). With the addition of 15 more genomes these results expanded to 35,972 for the pan clusters, with 15,254 accessory clusters, 18,839 unique and a reduction in the core genome to 1879. The average make up of each *Streptomyces* genome was about 25 % core genes, 60 % dedicated to accessory genes and 15 % for unique genes. Compared to other single genus core and pan genome analysis, the contribution of the *Streptomyces* core genome was smaller than 17 Nocardiopsis genomes whose core was around 43 but larger than 10 Mycobacterium genomes whose core was around 16 % (Li *et al.,* 2013a; Prasanna & Mehra 2013). *S. cattleya* NRRL 8057 had the smallest number of protein-coding genes and had the largest core contribution to its genome, which was the reverse for *S. bingchenggensis* BCW 1 in which the largest pool of protein-coding genes were present and the smallest core contribution. Along with *S. albus* J1704, the *S. cattleya* strains have the smallest genomes of the *Streptomyces*. However *S. albus* J1704 is generally considered the best heterologous host and example of a minimal genome (Zaburannyi *et al.,* 2014), most likely due to the presence of a

megaplasmid, pSCAT, in *S. cattleya*. Additionally the *S. cattleya* strains are the most diverse, as demonstrated by the phylogenomic analysis and whole genome similarity (Figure 3.4 and 3.5), these strains also have the second greatest unique gene contribution, 23 – 24 %. *S. cattleya* DSM 46488 had the least number of accessory genes within its genome. As 33 % of its genome contains core genes and 24 % is unique, leaving 43 % of its genome to be composed of accessory genes. This further supports the diverse nature of the *S. cattleya* species, as it has the least overlap with other *Streptomyces* species. The accessory gene contribution to the *S. violaceusniger* Tu 4113 genome was the greatest with 66 %, suggesting this species has an abundance of genes common to *Streptomyces* within its gene pool and therefore more of an overlap with the other species. The unique gene contribution to the genome of *S. flavogriseus* ATCC 33331 was 2.5 %. Whole genome pairwise comparisons (Figure 3.5) and the phylogenomic trees (Figure 3.4) suggest that *S. flavogriseus* ATCC 33331 and *S. sp* PAMC26508 are closely related. When defining a bacterial species 5 techniques are generally used, phenotypic properties, DNA-DNA hybridisation, 16S rRNA sequence, multilocus sequence analysis (MLSA) and genomics (Staley 2006; Rong & Huang 2012; Oren & Garrity 2013). Of these techniques 16S rRNA sequence comparison is regarded as being useful to classify a bacterium to the genus level, with MLSA, phenotypic properties and genomics being able to reach the level of species and sub-species (Staley 2006; Oren & Garrity 2013). Taking MLSA (Figure 3.4 C) and genomics into consideration when discussing *S. flavogriseus* ATCC 33331 and *S. sp* PAMC26508, then these bacteria can be considered the same species. Nucleotide blast alignments of the MLSA genes (Rong & Huang 2012; Rong *et al.,* 2013) reveal 100 % identity and coverage in all but *recA*, with the same applying to amino acid sequence alignments. Further to this whole genome alignments give an identity of 99 % with coverage of 96 %. The majority of the proteins from both bacteria gave a sequence identity of 98 % and it was thought this might be a reason for so few unique genes within the *S. flavogriseus* ATCC 33331 genome. However comparisons

excluding *S. sp* PAMC26508 only increased the percentage of unique genes to 7 % for *S. flavogriseus* ATCC 33331. It is interesting how closely related both *S. sp* PAMC26508 and *S. flavogriseus* ATCC 33331 are, as the former is an endosymbiotic bacterium that was originally isolated from the Antarctic lichen Cladonia borealis, with the later being a soil organism isolated from Laval, Canada (Doroghazi & Buckley 2010; Shin *et al.,* 2013). Even with such distinct habitats there are only 290 genes unique to *S. flavogriseus* ATCC 33331 and 922 genes unique to *S. sp* PAMC26508. Considering the degree of similarity between *S. flavogriseus* ATCC 33331 and *S. sp* PAMC26508 at the DNA and amino acid level for MLSA and genomics, it seems likely they are the same species. However the definitive answer would come only from combining the MLSA, genomic and phenotypic property data for both bacteria.

Previous studies have suggested that *S. rimosus* SR1 has a different genome structure compared to other *Streptomyces* species, proposing a unique evolutionary heritage (Kirby *et al.,* 2008). The phylogenomic trees support this theory, however the pan tree suggests *S. rimosus* SR1 and *S. cattleya* share a common ancestor, whereas the core tree suggests that these two species diverged at an earlier point in time (Figure 3.4 A). The high number of unique genes further supports the idea of *S. rimosus* SR1 having a more distinctive origin. At the protein identity level *S. rapamycinicus* NRRL 5491 was the most closely related to *S. rimosus* SR1 (Figure 3.1 B), and both are found on the same clade on the both phylogenomic trees (Figure 3.4). Even though *S. rapamycinicus* NRRL 5491 shares the greatest sequence identity to *S. rimosus* SR1 proteins there are still 3821 and 3241 genes unique to each species, respectively, which is 37 % of the *S. rimosus* SR1 genome. Of the genes that were unique to *S. rimosus* SR1, when compared to all the *Streptomyces*, around 78 % were uncharacterized/hypothetical, with 5 % of the characterised proteins belonging to regulation and cell signaling (Table 3.5). Along with TCS RR and SK there were 63 transcriptional regulators, of which the GntR family was most common. As with TetR family regulators, GntR regulators are

involved in modulating gene expression in response to environmental factors, however GntR regulators bind effector molecules that modify the proteins activity (Hillerich & Westpheling 2006; Hoskisson & Rigali 2009). GntR regulators can function as either a repressor or activator and within *Streptomyces* are most commonly attributed to morphology and antibiotic production (Hoskisson *et al.,* 2006; Bai *et al.,* 2013), with DasR and WhiH the most well known (Rigali *et al.,* 2008; Shen *et al.,* 2007). Quite often genes that are controlled by a GntR regulator are located adjacent to the regulator on the chromosome (Hoskisson & Rigali 2009), so the genes located next to the *S. rimosus* SR1 GntRs were examined. Most of the adjacent genes were hypothetical; however, the GntR encoding genes *SRIM3283*, *SRIM5447* and *SRIM6345* lay next to genes encoding a FAD-dependent pyridine nucleotide-disulphide oxidoreductase, magnesium and cobalt transport protein and lanthionine biosynthetic genes, respectively. As GntR regulators are known to control many different cellular processes, it is not unreasonable to suggest these GntRs may be involved in regulation of these genes or clusters. Most likely these regulators are involved in specific cellular responses or activities that give *S. rimosus* SR1 its unique biological features.

The most abundant categories of genes within the core genome were carbohydrates, amino acids and derivatives and regulation and cell signaling. The genes associated with carbohydrates and amino acids and derivatives were linked to primary cell functions, such as primary metabolism, and for this reason were not examined extensively as these processes are present in all bacteria. As dicussed previously (Chapter 1.1), primary metabolism is of great importance when considering secondary metabolism. From a variety of different carbon sources, 12 precursor metabolites are used for secondary metabolism but are also used in the production of all cellular components (Rokem *et al.,* 2007). As the 12 precursor metabolites; glucose-6-P, fructose-6-P, ribose-5-P, erythrose-4-P, glyceralde-hyde-3-P, 3-phosphoglycerate, phosphoenolpyruvate, pyruvate, acetyl-CoA, 2-oxoglutarate, succinyl-CoA and oxaloacetate, are also used in other cellular functions there is a switch from balanced utilisation to specific use of one or

two metabolites during secondary metabolism (Rokem *et al.,* 2007; Hwang *et al.,* 2014). This can be observed as a change in primary metabolism carbon flux as demonstrated by comparisons between an antibiotic producing and non-producing strain (Coze *et al.,* 2013). Additionally ATP, NADH and NADPH also provide a integral connection between primary and secondary metabolism as these cofactors are shared (Rokem *et al.,* 2007). At the induction of secondary metabolism there is an increase in production of NADPH due to increased flux into the pentose phosphate pathway (PPP) (Hwang *et al.,* 2014). Therefore identifying the core genes involved in primary metabolism and any differences between the *Streptomyces* species may explain why certain species produce different antibiotic types compared to others e.g. non-ribosomal polypeptides rather than Type-I polyketides.

A large number of genes were associated with regulation and cell signaling that are not ubiquitous. Of the regulatory and cell signaling associated genes the most prominent were regulators. The primary regulator class within the core genome of *Streptomyces* was TetR, followed by TCS. Primarily TetR family regulators control genes involved in efflux pumps, environmental stress, multidrug resistance, biosynthesis of antibiotics, and pathogenicity (Ramos *et al.,* 2005). Proteins in this family have also been found to influence morphology/development and small active molecule production in both Gram negative and positive bacteria (Xu *et al.,* 2010; Shimada *et al.,* 2012; Yin *et al.,* 2013; Longo *et al.,* 2013). The proximity of *SCO2374* to the TcmA homologue, *SCO2373*, suggests that *SCO2374* encodes a TcmR-like TetR regulator (Ramos *et al.,* 2005). TcmR from *S. glaucescens* represses both *tcmR* and *tcmA* in the absence of tetracenomycin C, by binding to the shared -10 to -35 region of each genes promoter (Guilfoile & Hutchinson 1992). This suggests that all these *Streptomyces*, apart from *S. albus* J1704, share tetracenomycin C resistance and common regulation, as all have both *tcmA* and *tcmR* that are divergently transcribed. Although *S. albus* J1704 possesses *tcmA* there is no apparent regulator, other than the MarR family transcriptional regulator encoded by *XNR_3114* that lies upstream. As MarR transcriptional regulators are

associated with repression of antibiotic resistance genes (Alekshun *et al.,* 2001), then it is likely XNR_3114 fulfills the role of TcmR in *S. albus* J1704 by regulating the expression of *XNR_3115*, the *tcmA* homologue. Another TetR-family protein that was conserved among the *Streptomyces* encoded a homologue of AtrA, a global regulator rather than a local repressor. This is quite significant as AtrA from *S. coelicolor* A3(2) has been associated with actinorhodin (ACT) production, as the essential activator of *actII-ORF4*, and can bind to the promoter of *strR*, the activator of streptomycin production in *S. griseus* (Uguru *et al.,* 2005; Liu *et al.,* 2013). Interestingly AtrA influences only ACT production, as *atrA* gene knockouts decreased ACT production but had no visible effect on undecylprodigiosin (RED) or calcium-dependent antibiotic (CDA) (Uguru *et al.,* 2005). In *S. coelicolor* A3(2) *atrA* has been shown to regulate N-acetylglucosamine (GlcNAc) transport, which, via the GlcNAc-signalling pathway, counteracts the negative affect DasR has on *actII-ORF4* expression (Nothaft *et al.,* 2010; van Wezel & McDowall 2011). Furthermore disruption of *atrA* in *S. griseus* decreased streptomycin production under some growth conditions (Hirano *et al.,* 2008). Taken together it is likely that the homologue of AtrA, within each Streptomycete, plays a regulatory role for at least one antibiotic biosynthetic cluster.

Within the *Streptomyces* core an average of 32 TCS were present with 7 lone RR and 6 lone SK. TCS systems generally consist of a cytoplasmic RR that is regulated by a transmembrane SK (Galperin 2010; Yamamoto *et al.,* 2013). Bacterial TCS are one of the key controllers of adaption to changes in environmental conditions, as TCS can influence genes and proteins throughout the entire cell (Galperin 2010). In *Streptomyces* TCS control a broad range of secondary metabolite biosynthetic clusters, either directly or indirectly (Rodríguez *et al.,* 2013). The previously characterised TCS identified covered both these areas, with 4 TCS connected to antibiotic production and 2 TCS coupled to environmental stress. OsaB is required for osmoadaption in *Streptomyces*, as gene knockouts are unable to produce aerial hyphae on media supplemented with an osmolyte (Bishop *et al.,* 2004). OsaA is the SK associated with OsaB,

however *osaA* knockouts in *S. coelicolor* A3(2) do not have as severe a phenotype as *osaB* knockouts, and it has been suggested this may be due to the presence of a lone SK paralogue, *SCO7327* (Bishop *et al.,* 2004). If this is the case then *osaA* knockouts of *S. hygroscopicus jinggangensis* 5008, *S. hygroscopicus jinggangensis* TL01, *S. avermitilis* MA-4680, *S. scabiei* 87 22, *S. venezuelae* ATCC 10712, *S. collinus* Tu 365, *S. sp*. SirexAA and *S. rimosus* SR1 should have delayed osmoadaption and development, as seen with *osaA* knockouts of *S. coelicolor* A3(2), as each species contains a *SCO7327* homologue. Furthermore, both OsaA and OsaB are associated with a regulatory protein OsaC, which is also required for osmoadaption (Martínez *et al.,* 2009). The expression of *osaB* occurs from two promoters, *osaBp1* and *osaBp2*, associated with osmoadaption response and osmotic stress response respectively (Martínez *et al.,* 2009). The activation of osaBp2 is in a σB-dependent manner and during periods of osmotic stress *sigB* expression is induced via the *sigBp1*. *sigB* expression, induced in this manner, results in a positive feedback loop as activation of the *sigBp1* promoter is σB dependent (Martínez *et al.,* 2009). It is the role of OsaC to modulate this response, thus returning *sigB* and *osaB* expression to basal levels after physiological recovery of the cell in response to osmotic stress (Martínez *et al.,* 2009). Given the role played by OsaA, OsaB and OsaC in osmoadaption and in response to osmotic stress, as well as effect the consequent gene knockouts have on morphological differentiation, it is understandable that all the *Streptomyces* would have these proteins present. In addition to the TCS of OsaA and OsaB, the RR CseB and the SK CseC are also associated with adaption to environmental stress. As cell envelope integrity changes due to stress, CseB and CseC, are responsible for the activation of *sigE*, an ECF sigma factor that is recruited by RNA polymerase to transcribe genes associated with cell wall glycan production and general cell envelope functions (Paget *et al.,* 1999; Hong *et al.,* 2002). The whole *sigE* operon contains *cseA, cseB* and *cseC* and all are present in the core genome of the *Streptomyces*. The antibiotic production associated TCS within the core are well characterized and understood, most likely due to the

global effect these systems have on antibiotic production. Of the TCS within the core CutR and CutS varied for *S. rimosus* SR1, as there were two copies of CutS, potentially providing an alternative SK for CutR. As both TetR family transcriptional regulators and TCS are generally involved in global cell functions, it is reasonable for these genes to be found within the core genome of the *Streptomyces* examined.

antiSMASH analysis of the *Streptomyces* genomes predicted NRPS, T1PKS and bacteriocin as the most common biosynthetic clusters present. However for the majority of the *Streptomyces*, these are not the main type of secondary metabolite associated to them. It is not uncommon for cryptic clusters to be present in *Streptomyces* genomes (Doroghazi & Metcalf 2013) and it is possible particular clusters are activated more readily due to central metabolism, as many of the precursor molecules required for secondary metabolism are provided via central carbon metabolism (Rokem *et al.,* 2007). Therefore depending on the presence and expression of genes associated with central carbon metabolism, precursor supply will vary and this may prevent the activation of a secondary metabolite cluster but allow the expression of another. Most of the *Streptomyces* genomes contain the same central carbon metabolism genes however there is variation in the number of homologues of each gene. It is possible the variation is related to the main secondary metabolite associated with each species, either because central carbon metabolism influenced the secondary metabolites produced or vice versa. Though it should be noted that the principal secondary metabolite linked with each species is also influenced by how readily the antibiotic is produced and isolated in the lab, for example the principal secondary metabolites of *S. rimosus* SR1 are rimocidin and OTC. It is possible in the natural habitat of *S. rimosus* SR1 other secondary metabolites are more readily produced. OTC and rimocidin are produced under the same growth conditions and both also incorporate malonyl-CoA during biosynthesis, which is supplied via the anaplerotic node by the conversion of pyruvate into acetyl-CoA (Ryu *et al.,* 2006). Pyruvate is the product of glycolysis in which glucose is broken down initially to glucose-6-P which continues through

glycolysis however glucose-6-P can be converted to either 6-phosphoglucolactone or glucose-1-P (Ryu *et al.,* 2006). The two enzymes required for these conversions are glucose-6-P dehydrogenase (Zwf) and phosphoglucomutase and are required for the first steps of the pentose phosphate pathway (PPP) or glycogen synthetic pathway respectively (Ryu *et al.,* 2006). Examination of the genes associated with these pathways in each *Streptomyces* species showed a correlation between carbon flux and the production of two polyketides, as *S. rimosus* SR1 only has one copy of the *zwf* gene rather than two; however, there are two copies of phosphoglucomutase. This suggests a potential link between central carbon metabolism and the primary antibiotics produced by *S. rimosus* SR1.

During the *Streptomyces* comparison it was established that *S. rimosus* SR1 has 4 sets of *nar* genes, and only *S. rimosus* SR1 and *S. coelicolor* A3(2) have multiple sets. *nar* genes are associated with anaerobic growth as they can allow for the use of nitrate as an electron acceptor rather than oxygen (Richardson *et al.,* 2001; Gilberthorpe & Poole 2008). *Streptomyces* are generally considered obligate aerobes, as are Mycobacterium but both have species within their genus that contain a set of *nar* genes (Cole *et al.,* 1998; van Keulen *et al.,* 2003). *S. coelicolor* A3(2) can reduce nitrate under both aerobic and anaerobic conditions and that the *nar* genes are responsible for this activity, confirming a biological role for these genes in anaerobic conditions (Fischer *et al.,* 2010). This suggests that the presence of *nar* genes within a *Streptomyces* genome would enable that species to reduce nitrate aerobically and anaerobically. As each set of Nar genes are predicted to function at particular points in the cell cycle it is interesting that those *Streptomyces* that contain *nar* genes do not have all 3 sets. It is possible there is a certain amount of redundancy between the Nar proteins, and they are simply developmentally regulated. If this were the case then it would allow a strain containing one set of *nar* genes to reduce nitrate at any point in the cell cycle, assuming the genes are expressed. Engineering a strain of *S. coelicolor* A3(2) with the *nar* promoters switched would allow for this hypothesis to be tested, e.g. placing the *nar3* set under

the promoter of *nar1* and examining if nitrate can be reduced in spores. An alternative explanation is that a complete set is not essential for survival but can give a competitive edge, given that only 2 of the *Streptomyces* species contained a complete set of *nar* genes, and gene knockouts of *nar123* in *S. coelicolor* A3(2) had no detrimental effect on spore survival (Fischer *et al.,* 2010). As there is another set of nitrate reductases present in the *S. coelicolor* A3(2) genome, the assimilatory nitrate reductases, NasA1 (SCO2473) and NasA2 (SCO7374) (Wang & Zhao 2009; Fischer *et al.,* 2013), an alternative method of nitrate reduction is potentially available for those species that do not have nar genes. However, not all of the species without *nar* genes contain a copy of the *nasA1* or *nasA2* genes, so most likely nitrate reduction is not necessary for survival or another means is utilised. The presence of *narJ* genes within the genomes of *S. hygroscopicus jinggangensis* 5008 and *S. hygroscopicus jinggangensis* TL01 suggests that at one point both these strains had a set of *nar* genes, as NarJ is required for the maturation of NarI and NarGH complex (Lanciano *et al.,* 2007).

From the information provided here there is a clear benefit to comparative genomics in providing in depth understanding of those features that are unique to a genus. Increasing the breadth of species would allow for the identification of those proteins that constitute the core of a genus or even phylum. Further to this, whole genome comparison can provide insight into the evolutionary origin of a species with greater definition than MLSA or 16S rRNA comparison. Although biochemical analysis would still need to be carried out to confirm predicted species.

Chapter 4

4 The genomic differences between *Streptomyces rimosus* SR1 and the production strains

4.1 Introduction

When *S. rimosus* SR1 (G7) was first isolated in 1950 (Finlay & Hobby 1950), it produced only 1 g/L of oxytetracycline (OTC); however, over the many years that followed the productivity of *S. rimosus* increased to at least 55 g/L of OTC, as seen with *S. rimosus* SR55. This was achieved through a strain improvement program involving treatment of *S. rimosus* cells with N-methyl-N'-nitro-N'-nitrosogaunidine (MNNG) and UV irradiation (Petković *et al.,* 2006), followed by selection of those bacterial cells that produced more OTC. Although strains of *S. rimosus* were isolated that had increased OTC productivity, the exact genetic changes that lead to this increase are unknown due to the method employed. There is now an opportunity to analyse the precise genomic differences that occurred in each of the production strains and link these changes to OTC productivity. However, in order to investigate the genetic basis for increased OTC productivity it was evident that it was necessary to obtain a genome sequence of the original soil isolate *S. rimosus* SR1, in addition to the subsequent production strains. As *S. rimosus* SR1 was to be used as the reference sequence, a whole-genome shotgun sequencing approach was performed, using a Roche 454 GS Junior. This approach was used as the coverage and average read length was optimal for assembling a scaffold genome sequence (Loman *et al.,* 2012), allowing genomic comparisons between *S. rimosus* SR1 and the subsequent strains in the lineage to be carried out.

The later strains in the OTC strain improvement lineage; *S. rimosus* SR20 (M4018), SR50S (15883S) and SR55 (23383), were sequenced previously using different methods. *S. rimosus* SR20 was sequenced using

a Roche 454 GS FLX (454) system, whilst *S. rimosus* SR50S and SR55 were sequenced using an Illumina Genome Analyzer II (Illumina) with 200 bp paired end reads. In order to make comparisons between *S. rimosus* SR1 and the other strains, alignment files were made. Burrows-Wheeler Alignment (BWA) (Li & Durbin 2009; Li & Durbin 2010) and Sequence Alignment/Map (SAMtools) (Li *et al.,* 2009) were used to produce alignment files that could be used within the Artemis software for both the 454 and Illumina reads. BWA was chosen as the alignment tool as the option to map long or short reads was available. This was necessary as the 454 reads were an average of 427 bp long compared to the 31 bp length of the Illumina reads.

After comparing the production strains with *S. rimosus* SR1, chromosomal rearrangements, in the form deletions, and mutations, such as single nucleotide change (SNC) or insertions and deletions (InDEL), were identified. In order to understand the basis of increased OTC productivity in the production strains the deletions, SNCs and InDELs were examined in greater detail. In the case of genomic deletions, genes that had undergone deletion were examined further to determine any possible association with increased OTC production. When SNCs were examined, it was necessary to identify whether the SNC was intergenic or intragenic. The placement of a SNC can influence a gene or its protein, either by altering expression or modifying the protein produced in the form of a non-synonymous SNC. Examination of InDELs required the determination of whether the InDELs was intragenic before ascertaining the affect this had on the encoded protein. As such, when genomic comparisons were made, the number of deleted genes, and the percentage of intergenic, synonymous and non-synonymous SNCs were noted, as well as the genes affected by non-synonymous SNCs and disruptive InDELs. These comparisons formed the basis of the experiments that make up the subsequent chapters of this thesis.

4.2 Generation of comparison files for the identification of SNCs and InDELs

In order to identify SNCs and InDELs between the production strains and SR1, comparison files were generated. As mentioned previously BWA was the alignment tool used for mapping the SR20 454 reads and the SR50S and SR55 Illumina reads to the SR1 genome. It was first necessary to merge the raw .fasta and .qual files for both the 454 and Illumina reads first needed to be merged to produce .fastq files; this was achieved using a fastaQual2fastq perl script. Once this was accomplished these files were then entered into the BWA pipeline. From this .sam files were generated that were then converted into .bam files and allow visualisation in Artemis (Carver *et al.,* 2012) by mapping to a SR1 Genbank file. Identification of SNCs and InDELs was performed using the VarScan v2.3.6 (Koboldt *et al.,* 2009; Koboldt *et al.,* 2012) pipeline. SNCs and InDELs were then filtered by read frequency (95%), read depth (≥ 10) and read quality (≥ 17) before progressing to the identification of intergenic, synonymous and non-synonymous SNCs using TRAMS (Reumerman *et al.,* 2013) and manual identification.

4.3 Overview of genomic differences between SR1 and all strains

4.3.1 Deletion profiles of each production strain

SR50S is an OTC non-producing derivative of SR50 due to a spontaneous chromosomal deletion (Iain Hunter, personal communication). Due to this observation it was logical to establish the degree of genome loss from the production strains, as it was likely spontaneous deletions had occurred in SR20 and SR55, but did not result in the loss of OTC production. Genomic deletions were identified from the *de novo* assemblies of SR20, SR50S and SR55 when mapped to SR1 (Figure 4.1) via a nucleotide BLAST CCT alignment. The de novo assemblies of SR20, SR50S and SR55 were

achieved using Mira (Chevreux 2004). Sequence with no identity is represented by white space within the map and from this it was established that SR50S had lost the greatest amount of chromosomal DNA when compared to SR1. Given that SR55 stems from SR50, rather than SR50S, this observation was not unusual, as the spontaneous deletion that generated SR50S would not have occurred in SR50. Two regions of chromosomal loss with respect to SR1 were observed for SR20, totaling 346,991 bp, two regions for SR50S, a total loss of 680,500 bp, and four regions for SR55, with a total loss of 559,506 bp (Figure 4.1). However these totals include deletions that occurred in previous strains i.e SR55 shares the same genomic deletions as SR20. In order for a more informative assessment of the relationship between genomic deletions and OTC production, the deletions associated with each strain improvement step were determined. By examining the deletions in this way four strain profiles were deduced and new total genomic loss was determined. The deletion profile of SR50 was determined by identifying deletion regions shared only by SR50S and SR55, this was required, as the deletion profile of SR50S cannot solely be linked to increased OTC production, as the spontaneous deletion removed more than the OTC cluster. SR20 has the greatest gene loss, followed by SR50 then SR50S, with SR55 having the least (Figure 4.2). Once the deletion profile of each strain was resolved, further examination of the genes within was undertaken.

In order to obtain a better understanding of the influence a deletion had on OTC production for each production strain, it was necessary to identify the genes within. To aid this process each gene was categorised into a SEED subsystem (theSEED.org) (Overbeek 2005) rather than examining each gene individually. Each gene's SEED subsystem was determined using RAST annotation (Aziz *et al.,* 2008). The total number of genes within the SEED subsystems for each deletion profile is listed in Table 4.1. The hypothetical/uncategorised SEED subsystem contained the most for all production strains, followed by secondary metabolism, and then regulation and cell signaling for SR20 and SR50S. However, there are 3287

genes annotated as hypothetical/uncategorised within the *S. rimosus* SR1 genome. This means the 117 hypothetical/uncategorised genes lost from SR20 only account for 3.5 % of the total genes making up this category. Therefore the deleted genes were examined as a percentage of the total number of SR1 genes in each SEED subsystem, rather than overall numbers (Table 4.1). The SEED subsystems with the highest percentage of deleted genes, therefore secondary metabolism, iron acquisition, and miscellaneous SEED subsystems for SR20 (Table 4.1), were considered significant. For SR50 these SEED subsystems were miscellaneous, phages, prophages, transposable elements, and plasmids, and cell wall and capsule. Secondary metabolism, dormancy and sporulation, and phages, prophages, transposable elements, and plasmids for SR50S contained the highest percentage of genes. For SR55 the SEED subsystems were phages, prophages, transposable elements, and plasmids, nitrogen metabolism, secondary metabolism, cell wall and capsule, and miscellaneous. These are the SEED subsystems for each production strain that will be discussed in more detail below.

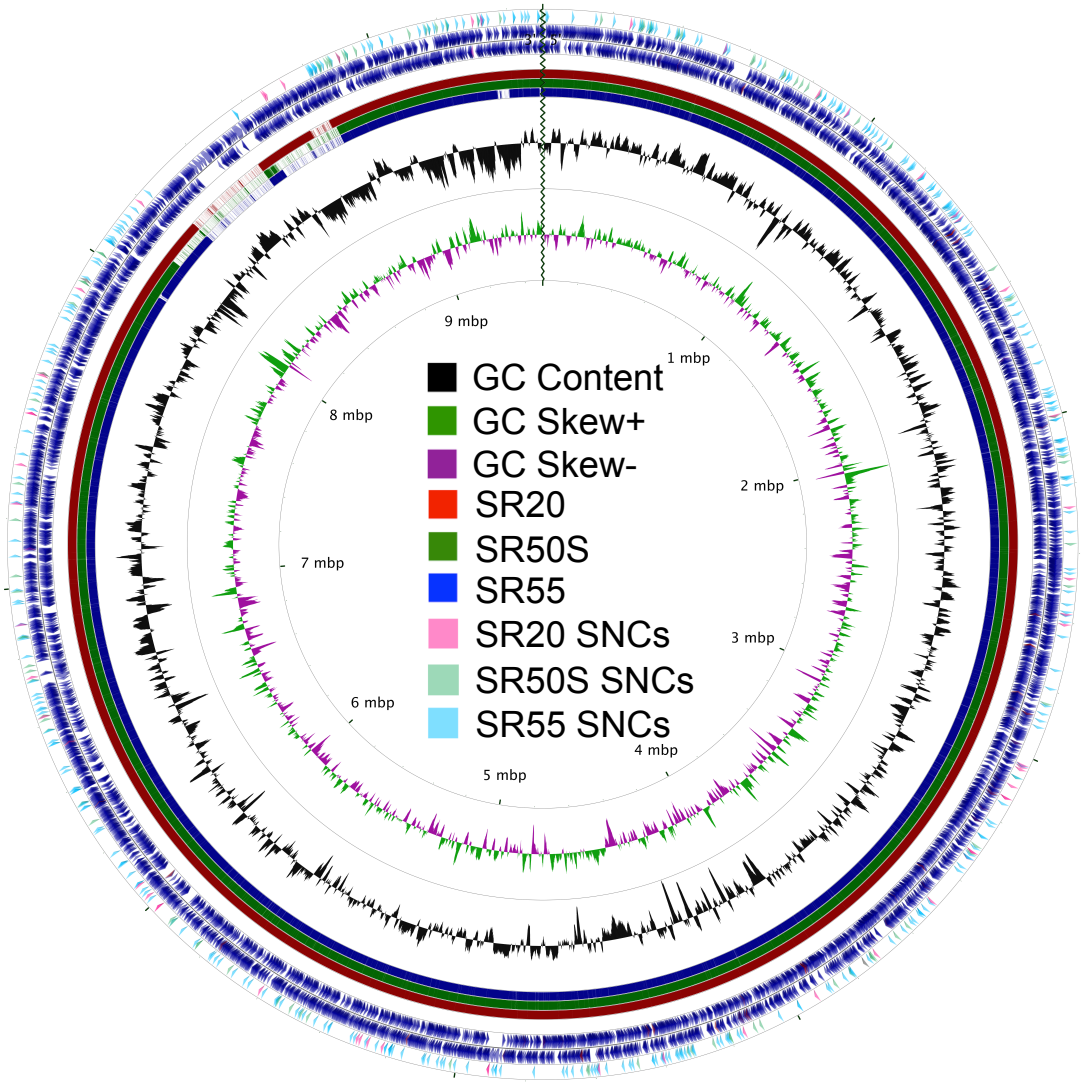


Figure 4.1 CCT Map of nBLAST of SR20, SR50S and SR55 versus SR1. Rings from the outside inward: SR20 (red), SR50S (green) and SR55 (blue). The vertical black jagged line represents ends of a linear chromosome. The relative position of non-synonymous SNCs for each production strain is shown around the outside edge of the map as specifically coloured arrows.

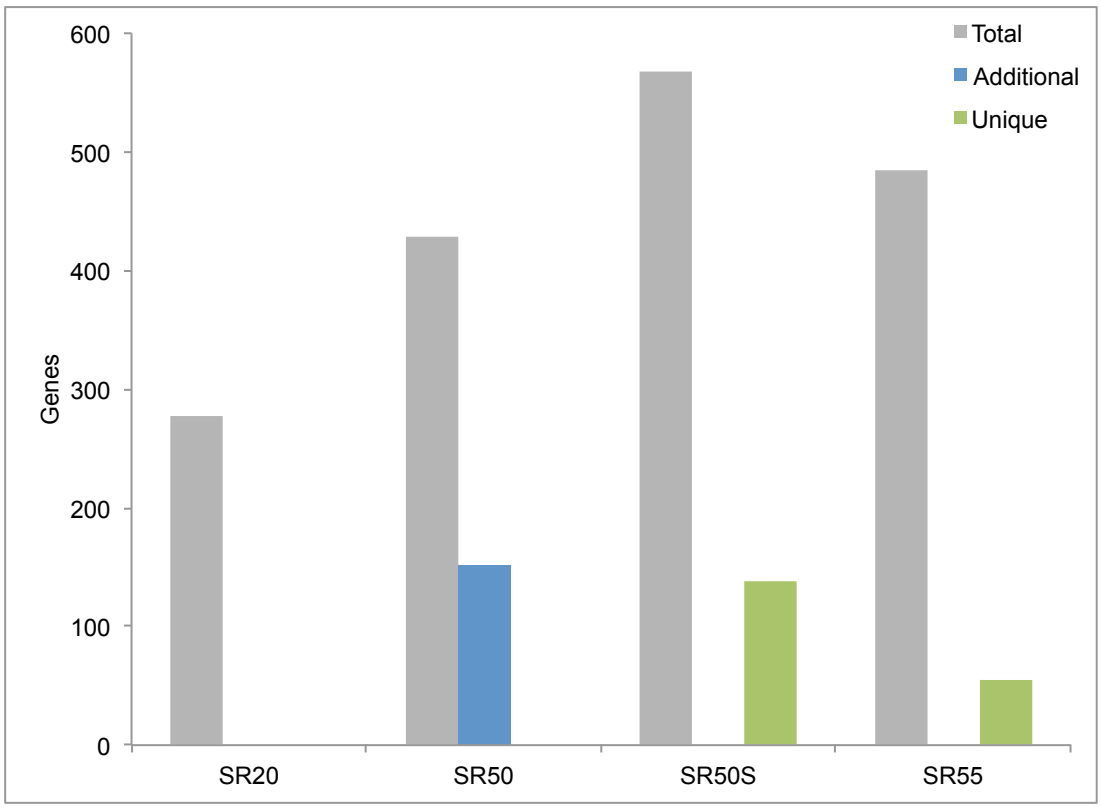


Figure 4.2 Overview of gene loss for each deletion profile. Total gene loss is cumulative and includes genes lost from previous production strains. Additional genes represents genes subsequently lost in the deletion profile but are shared by other strains. Unique signifies genes within a deletion that are unique to the associated strain.

Table 4.1 Total number of genes within each deletion profile and as a percentage of each SEED subsystem.

Subsystem	Total number of genes				Precentage of SEED subsystem			
	SR20	SR50	SR50S	SR55	SR20	SR50	SR50S	SR55
Cofactors, Vitamins, Prosthetic Groups, Pigments	7	0	0	2	2	0	0	0
Cell Wall and Capsule	9	6	0	2	5	3	0	1
Virulence, Disease and Defense (Resistance)	7	2	4	0	9	2	5	0
Potassium Metabolism	0	0	0	0	0	0	0	0
Miscellaneous	18	17	9	2	10	10	5	1
Phages, Prophages, Transposable Elements, Plasmids	2	2	7	6	4	4	12	11
Membrane Transport	4	1	0	1	1	0	0	0
Iron Acquisition and Metabolism	8	0	0	0	8	0	0	0
RNA Metabolism	2	1	0	0	2	1	0	0
Nucleosides and Nucleotides	1	1	0	0	1	1	0	0
Protein Metabolism	4	1	2	0	1	0	1	0
Cell Division and Cell Cycle	0	0	0	0	0	0	0	0
Motility and Chemotaxis	0	0	0	0	0	0	0	0
Regulation and Cell Signalling	26	14	13	3	4	2	2	0
Secondary Metabolism	25	1	28	1	30	1	34	1
DNA Metabolism	1	2	2	0	1	1	1	0
Regulons	0	0	0	0	0	0	0	0
Fatty Acids, Lipids, Isoprenoids	5	6	2	1	1	2	1	0
Nitrogen Metabolism	2	1	3	3	5	2	7	7
Dormancy and Sporulation	0	0	1	0	0	0	25	0
Respiration	8	1	3	0	3	0	1	0
Stress Response	10	2	1	0	6	1	1	0
Metabolism of Aromatic Compounds	2	0	0	0	3	0	0	0
Amino Acids and Derivatives	4	2	2	4	1	0	0	1
Sulphur Metabolism	0	0	0	0	0	0	0	0
Phosphorus Metabolism	1	0	0	0	1	0	0	0
Carbohydrates	11	6	8	0	2	1	1	0
Hypothetical	117	74	49	30	4	2	2	1
Sigma Factors	3	2	4	0	3	2	4	0
Unknown	0	0	0	0	0	0	0	0
Total	277	152	138	55	3	2	2	1

4.3.2 Overview of single nucleotide changes for each production strain

SNCs are single nucleotide differences between otherwise identical gene sequences. SNCs can be categorized as either synonymous, in which the nucleotide difference does not affect the encoded amino acid, non-synonymous, where the SNC causes a change in amino acid, or intergenic in which the SNC lies outside the coding sequence (CDS). An overview of the SNCs for each strain is shown in Figure 4.3 and illustrates that total SNCs increase over the strain improvement process and non-synonymous SNCs are the most abundant type. Non-synonymous SNCs were focused on, rather than intergenic or synonymous SNCs, as these polymorphisms have greatest impact on a CDS's product. The relative position of non-synoymous SNCs for each production strain is shown in Figure 4.1. Once again SEED subsystem categories were used to class SNCs and Table 4.2 lists the total number of non-synonymous SNCs for each strain within these groups. The most represented group was hypothetical/uncategorised, as seen previously with the deleted genes, so SNCs were examined as a percentage of a SEED subsystem, as before. From Table 4.2 it can be determined that the top three SEED subsystems for each strain were; secondary metabolism, sigma factors and RNA metabolism for SR20, secondary metabolism, nitrogen metabolism and miscellaneous for SR50S, and secondary metabolism, miscellaneous and with cell wall and capsule, DNA metabolism, nucleotides and nucleosides, and virulence, disease and defense equally for SR55.

All of the production strains had secondary metabolism within the top SEED categories for deletion profiles and non-synonymous SNCs. This category is the only SEED subsystem found within the top SEED subsystems for both deletion profiles and non-synonymous SNCs, with the exception of miscellaneous for SR55. This suggests that loss and disruption of non-OTC secondary metabolite clusters plays a role in increased OTC productivity.

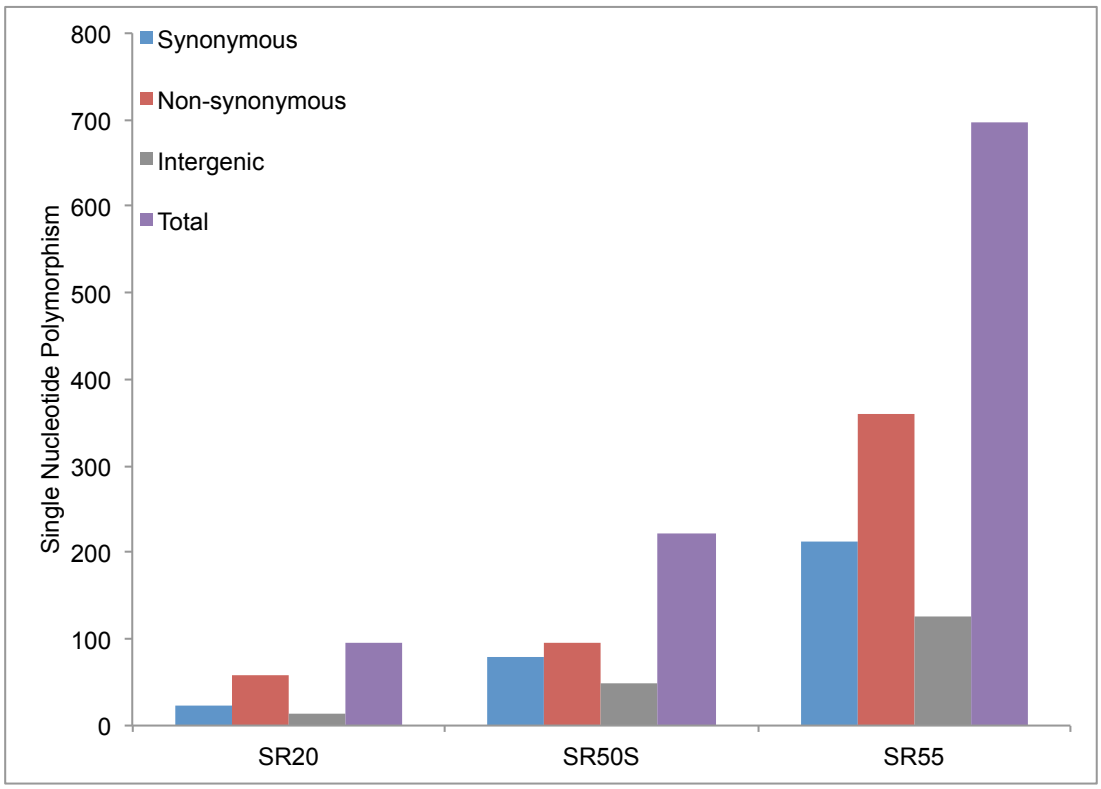


Figure 4.3 Categorisation of the single nucleotide polymorphisms for strains SR20, SR50S and SR55.

Table 4.2 Total number and percentage of non-synonymous SNCs within each SEED subsystem for SR20, SR50S and SR55.

Subsystem	Total number of genes			Percentage of SEED Subsystem		
	SR20	SR50S	SR55	SR20	SR50S	SR55
Cofactors, Vitamins, Prosthetic Groups, Pigments	0	1	5	0	0	1
Cell Wall and Capsule	1	4	14	1	2	8
Virulence, Disease and Defense	0	2	6	0	2	7
Potassium Metabolism	0	0	1	0	0	6
Miscellaneous	3	6	27	2	3	15
Phages, Prophages, Transposable Elements, Plasmids	0	1	3	0	2	5
Membrane Transport	5	7	17	1	1	3
Iron Acquisition	0	0	4	0	0	4
RNA Metabolism	2	2	6	2	2	5
Nucleosides and Nucleotides	1	1	11	1	1	8
Protein Metabolism	2	3	16	1	1	5
Cell Division and Cell Cycle	0	1	2	0	3	6
Motility and Chemotaxis	0	0	0	0	0	0
Regulation and Cell Signalling	11	7	33	2	1	5
Secondary Metabolism	2	10	14	2	12	17
DNA Metabolism	2	0	11	1	0	8
Regulons	0	0	0	0	0	0
Fatty Acids, Lipids, Isoprenoids	0	4	11	0	1	3
Nitrogen Metabolism	0	3	1	0	7	2
Dormancy and Sporulation	0	0	0	0	0	0
Respiration	2	3	7	1	1	3
Stress Response	1	1	5	1	1	3
Metabolism of Aromatic Compounds	1	1	1	2	2	2
Amino Acids and Derivatives	5	2	32	1	0	5
Sulphur Metabolism	0	0	0	0	0	0
Phosphorus Metabolism	0	2	0	0	3	0
Carbohydrates	2	5	24	0	1	4
Hypothetical	15	27	109	0	1	3
Sigma Factors	2	1	0	2	1	0
Unknown	0	1	0	0	2	0
Total	57	95	360	1	1	4

4.4 Genomic comparisons between *S. rimosus* SR1 and *S. rimosus* SR20

4.4.1 SR20 has lost genes associated with two secondary metabolite biosynthetic clusters

During the strain improvement process *S. rimosus* SR1 was randomly treated with mutagens, eventually leading to the isolation of SR20, which produced 20 – 30 g/L of OTC, compared with less than 1 g/L for SR1. As the process for increased production was random it was never established at a genetic level what lead to this increase. From preliminary genomic examination it was determined that two regions of DNA was lost in SR20, one of 277,657 bp and another of 69,334 bp, resulting in the deletion of 277 genes. A zoomed CCT map of the 277 kbp region is shown in Figure 4.4, and within this region two secondary metabolite clusters are present, the rimocidin and mnp clusters, as well as genes associated with iron acquisition and arsenic resistance. The 69 kbp region contains mainly hypothetical genes, though stress response genes are also present. It is interesting to note that both regions contain genes associated with genetic mobility, i.e transposases, and suggests that transposition might explain deletions at these locations.

Of the 277 genes deleted it was determined that the main SEED subsystems in which these genes belonged were; secondary metabolism, iron acquisition, virulence, disease and defense, and miscellaneous. Deleted genes that fell within these categories, as well as the top pBLAST hit, are listed in Table 4.3. The genes *SRIM7641, SRIM7642, SRIM7648, SRIM7649* and *SRIM7650* were placed within the iron acquisition SEED subsystem, originally due to the RAST annotation, the pBLAST hits of the encoded proteins also support this. *SRIM7589, SRIM7590*, and *SRIM7591* were identified to be genes associated with arsenic resistance. In addition to those genes previously mentioned there was also a loss of genes associated with the rimocidin and *mnp* biosynthetic clusters. *SRIM7746*, *SRIM7748* and *SRIM7749* encode the polyketide synthases (PKS) of the mnp polyketide,

Mnp2, Mnp4 and Mnp5 respectively (Seco *et al.,* 2004). Initial analysis of each of the PKS, using SBSPKS (Anand *et al.,* 2010) identified the signatures of acyltransferases (AT) that incorporate methylmalonyl-CoA and malonyl-CoA, and supports the findings of *Seco et al,* 2004. Loss of these genes might reduce resource competition as the extender units in OTC production are derived from malonyl-CoA and suggests that the loss of the *mnp* biosynthetic cluster is more significant than the other genes within the deletion. The loss of the rimocidin biosynthetic cluster was equally significant; however, this will be discussed in greater detail in the next chapter.

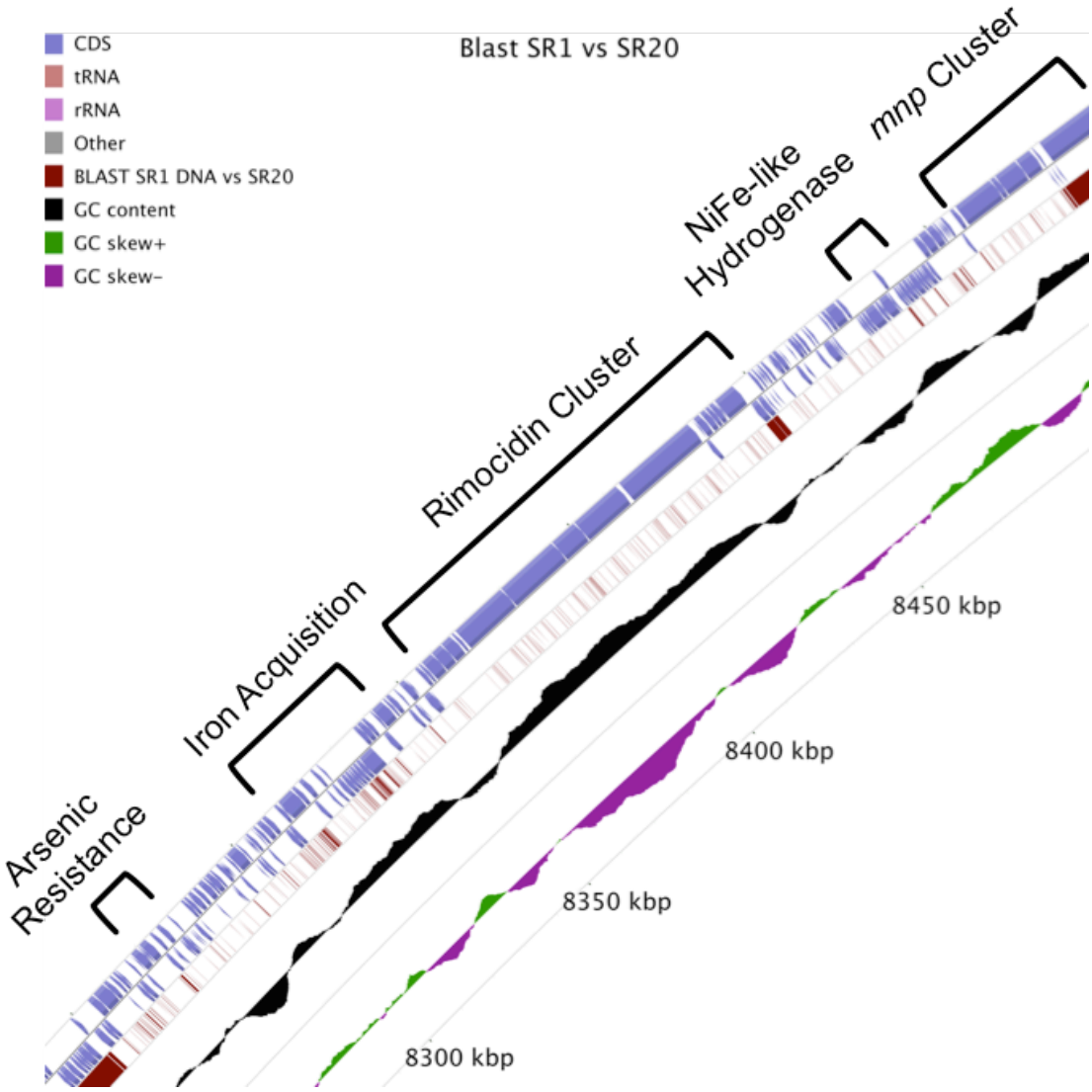


Figure 4.4 Zoomed CCT nBLAST map of the 277 kbp deletion region of SR20. Small regions of homology can be seen as thin red lines, however viewing SR20 sequence reads mapped to SR1 in Artemis shows no mapping in these areas.

Table 4.3 Genes within the SR20 deletion contributing to the main SEED subsystems

Gene	RAST Annotation	Homologue	Species	Identity
SRIM7566	Putative multidrug efflux ABC transporter	Putative ABC transporter	*Leifsonia aquatica* ATCC 14665	62 %
SRIM7588	Secreted Protein	Secreted protein	*Streptomyces violaceusniger* Tu 4113	88 %
SRIM7589	Arsenate reductase	Protein tyrosine phosphatase	*Streptomyces griseus* XylebKG-1	78 %
SRIM7590	Arsenate reductase	Arsenical-resistance protein	*Streptomyces violaceusniger* Tu 4113	87 %
SRIM7591	Arsenical resistance operon repressor	ArsR family transcriptional regulator	*Streptomyces himastatinicus* ATCC 53653	87 %
SRIM7601	Permease of the drug/metabolite transporter (DMT) superfamily	Putative uncharacterized protein	*Streptomyces violaceusniger* Tu 4113	72 %
SRIM7602	Cupin domain, putative	Cupin 2 conserved barrel domain protein	*Kribbella flavida* DSM 17836	70 %
SRIM7628	Fe3+ dicitrate transport system permease protein FecD	Uncharacterized protein	*Streptomyces sp.* HPH0547	72 %
SRIM7629	ABC-type Fe3+ siderophore transport system, permease component	Transport system permease protein	*Streptomyces sp.* SirexAA-E	75 %
SRIM7630	Periplasmic binding protein	Iron(III) dicitrate ABC transporter, periplasmic protein	*Streptomyces sp* e14	68 %
SRIM7631	Isochorismatase of siderophore biosynthesis	Isochorismatase	*Streptomyces sp.* e14	75 %
SRIM7636	Putative acetyltransferase	Putative acetyltransferase	*Streptomyces griseus* subsp. griseus	76 %
SRIM7637	Alpha/beta hydrolase fold	Alpha/beta hydrolase superfamily	*Herbaspirillum frisingense* GSF30	53 %
SRIM7641	ABC-type Fe3+ siderophore transport system, permease 2	ABC-type transporter, integral membrane subunit	*Streptomyces violaceusniger* Tu 4113	58 %

SRIM7642	ABC-type Fe3+ siderophore transport system, permease component	ABC-type transporter, integral membrane subunit	*Streptomyces violaceusniger* Tu 4113	67 %
SRIM7648	Glycosyltransferase IroB	Glycosyltransferase, MGT family	*Yokenella regensburgei* ATCC 43003	49 %
SRIM7649	Enterochelin esterase	Putative Ferric enterochelin esterase	*Streptomyces fulvissimus* DSM 40593	51 %
SRIM7650	Fe3+ dicitrate transport system, periplasmic iron-binding protein FecB	ABC-type Fe3+-hydroxamate transport system, periplasmic component	*Rothia mucilaginosa* DY-18	42 %
SRIM7706	Putative hydrolase	Hydrolase	*Saccharopolyspora erythraea*	77 %
SRIM7711	Predicted Protein	Uncharacterized protein	*Microlunatus phosphovorus* ATCC 700054	39 %
SRIM7714	Lead, cadmium, zinc and mercury transporting ATPase	ATPase	*Streptomyces collinus* Tu 365	74 %
SRIM7736	Beta lactamase	Beta-lactamase	*Rhodococcus sp*. EsD8	61 %
SRIM7744	3-oxoacyl-[acyl-carrier protein] reductase	Short chain dehydrogenase	*Streptomyces mobaraensis* NBRC 13819	83 %
SRIM7745	putative cytochrome P450 hydroxylase (mnp1)	Cytochrome P450 monooxygenase	*Streptomyces diastaticus*	100 %
SRIM7746	Malonyl CoA-acyl carrier protein transacylase (EC 2.3.1.39) (mnp2)	Type I polyketide synthase	*Streptomyces diastaticus*	100 %
SRIM7747	putative cytochrome P450 hydroxylase (mnp3)	Cytochrome P450 monooxygenase	*Streptomyces diastaticus*	100 %
SRIM7748	Malonyl CoA-acyl carrier protein transacylase	Type I polyketide synthase	*Streptomyces zinciresistens* K42	51 %
SRIM7749	Malonyl CoA-acyl carrier protein transacylase	Modular polyketide synthase NorC	*Streptoverticillium orinoci*	42 %

SRIM7751	acetyltransferase	Acetyltransferase	*Streptomyces bingchenggensis* BCW-1	88 %
SRIM7935	Putative secreted protein	Putative secreted protein	*Streptomyces sp*. HK1	68 %
SRIM7975	Oxygenase (Secreted protein)	Oxygenase	*Streptomyces sp*. C	52 %

The amino acid sequence encoded by each gene was subjected to pBLAST via Uniprot. Top pBLAST hits are listed

4.4.2 SR20 has non-synonymous SNCs within genes associated with (p)ppGpp

A total of 96 SNCs were identified between SR1 and SR20, with 24 being synonymous, 58 non-synonymous and 14 intergenic. Along with SNC analysis, short InDELs (sInDEL) were also examined. sInDELs were defined as any InDELs unassociated with the major deletion events that contributed to the deletion profiles. A total of 47 sInDEL were identified, with a total of 27 within CDS. As previously mentioned the major SEED subsystems in which SNCs were present were secondary metabolism, sigma factors, regulation and cell signaling and RNA metabolism, as these subsystems were overrepresented it was decided to focus on these groups. Table 4.4 lists the CDSs found within these subsystems, as well as all 27 sInDEL and SNCs that resulted in frame-shifts. Of the 42 SNCs and sInDEL identified, 29 were predicted to cause a frame shift, or introduce a stop or non-stop codon, with only 13 SNCs predicted to cause single amino acid substitutions. As the aim of this work was to identify mutations that potentially impact on OTC production, mutations that occurred out with the above SEED subsystem that were predicted to cause frame shifts and stop or non-stop codon changes were also examined.

Table 4.5 summarises the proteins with SNC or sInDEL mutations and the closest pBLAST homologues. From top pBLAST hits for high impact mutations, there are six mutations within the hypothetical category, four within cell wall and capsule, three associated with regulation, two connected with amino acids, secondary metabolism, DNA metabolism and carbohydrates, with one mutation for nitrogen metabolism, fatty acids, lipids, isoprenoids, membrane transport, cofactors, vitamins, prosthetic groups, pigments and miscellaneous. It is interesting to note that the genes *SRIM4545* and *SRIM6224*, both encoding nonribosomal peptide synthases (NRPS), have mutations that cause a stop to non-stop and frame-shift change, respectively. NRPS are similar to type I PKS as the genes encoding the NRPS are generally in a multi-gene cluster, however nonribosomal

peptides (NRP) are composed of amino acids (Galm *et al.,* 2008; Bushley *et al.,* 2013). Each NRPS consists of a domain for the activation of the amino acid, the adenylation (A) domain, a domain for the elongation of the peptide chain (peptidyl carrier protein (PCP) or thiolation domain), and a domain required for the condensation of the amino acids (condensation (C) domain) (Stack *et al.,* 2007; Strieker *et al.,* 2010). During the NRP synthesis process the PCP domain must be primed and this is achieved by the attachment of its 4'-phospopantetheine (ppant) cofactor via a ppant-transferase (Jia *et al.,* 2007; Strieker *et al.,* 2010). Coenzyme A (CoA) is the precursor of the ppant cofactor and it was shown that almost 80 % of CoA from *E. coli* grown on glucose was acetylated. antiSMASH (Medema *et al.,* 2011) analysis of the NRPS clusters associated with SRIM4545 and SRIM6224 identified 7 and 10 PCP domains, respectively. This suggests a similar association to other secondary metabolic pathways and OTC production, as seen with the deletions, if the CoA used for the production of the ppant, which primes these PCP domains, is acetylated.

The most prominent mutation category shared between the SNC and sInDELs were within proteins related to regulation, either as predicted regulators or sigma factors. Of these proteins SRIM1204 and SRIM7514, were predicted to have an association with increased OTC production, due to predicted functions and homologues.

SRIM1204 is a homologue of RelA, sharing 86 % identity with RelA from *S. coelicolor* A3(2), and RelA can influence antibiotic production, either positively or negatively (Sun *et al.,* 2001; Jin *et al.,* 2004; Saito *et al.,* 2006; Gomez-Escribano *et al.,* 2008). The non-synonymous SNC was within the amino-terminal (N-terminal) of SRIM1204, where the catalytic domain of the RelA protein in *E. coli* is located (Gropp *et al.,* 2001; Sajish *et al.,* 2009). The significance of the amino acid substitution, from lysine to glutamic acid (Table 4.4), is not obvious however it is still worth noting that the lysine is highly conserved in RelA among Gram positive and negative bacteria (Figure 4.5). SRIM7514 is a predicted cAMP binding protein and is a close homologue of SCO7699 and SCO5249, EshA and EshB respectively. Due to the shared

synteny of *SRIM7514* with *eshA* of other *Streptomyces* species (Figure 4.6), SRIM7514 was considered a homologue of EshA. EshA was found to be a positive regulator of ACT production but not undecylprodigiosin (RED) and is produced during late growth phase in *Streptomyces* species (Kawamoto *et al.,* 2001). Additionally for *S. griseus* an accumulation of dGTP occurred in an *eshA* mutant which led to the inability for the strain to produce aerial mycelium (Saito 2002). However, in a *S. coelicolor* A3(2) *eshA* mutant, lower levels of guanosine 5'-diphosphate 3'-diphosphate (ppGpp) accumulated and it was demonstrated that this was the cause of the non ACT producing phenotype (Saito *et al.,* 2006). Further to this Saito *et al.,* also demonstrated that it was the nucleotide binding domain that was essential for the activity of EshA. The non-synonymous SNC converts a leucine to a phenylalanine and was within the N-terminal cyclic nucleotide (cNMP) binding domain of SRIM7514 (Kwak *et al.,* 2001; McCue 2000). This leucine residue is not predicted to be involved in cNMP binding but is one of the conserved amino acids at its position, other than valine and isoleucine (McCue 2000). As EshA is generally associated with positive regulation of antibiotic production it is unlikely this mutation prevents activity.

Table 4.4 CDS in which frame-shift, stop or non-stop condon changes were predicted as well as SNCs contributing to the main SEED subsystems.

Mutation Type	Reference Position	Gene	Coding Region Change	Amino Acid Change
slnDEL	304360	*SRIM0252*	n.858C > +G	a.Pro286 > Fs
slnDEL	433807	*SRIM0371*	n.555C > +CGGACGT	a.Arg186 > Fs
slnDEL	495028	*SRIM0432*	n.976A > +G	a.Arg327 > Fs
slnDEL	694254	*SRIM0587*	n.816C > -CG	a.Pro273 > Fs
slnDEL	931528	*SRIM0805*	n.660C > +G	a.Thr222 > Fs
slnDEL	1067327	*SRIM0921*	n.56G > +GC	a.Pro20 > Fs
slnDEL	1073973	*SRIM0927*	n.454C > +GCGGCTGACGGGGCTGCCCGCCGA	a.Arg155 > Fs
slnDEL	1133554	*SRIM0982*	n.1132A > +CGATCAGCGAGTGGGCGCCCAGGC	a.Leu378 > Fs
slnDEL	1283125	*SRIM1134*	n.562C > +GGCGTGCCAGG	a.Gly188 > Fs
slnDEL	1666194	*SRIM1499*	n.230C > -G	a.Gly78 > Fs
slnDEL	1817136	*SRIM1633*	n.1516G > +CTA	a.Gly506 > Fs
slnDEL	2270395	*SRIM2037*	n.385C > +GTG	a.Gly129 > Fs
SNC	3133947	*SRIM2867*	n.696C > T	a.Arg233 > *
slnDEL	3267421	*SRIM2998*	n.641C > -G	a.Gly215 > Fs
slnDEL	3267445	*SRIM2998*	n.665A > -C	a.Arg223 > Fs
slnDEL	3419561	*SRIM3121*	n.69G > -G	a.Leu23 > Fs
slnDEL	3700985	*SRIM3437*	n.1843C > -G	a.Ala617 > Fs

SNC	4908782	*SRIM4545*	n.12421A > G and n12422A > G	a.*4141 > Trp
sInDEL	5175989	*SRIM4817*	n.690A > -G	a.Pro230 > Fs
sInDEL	5176092	*SRIM4817*	n.587T > -C	a.Gly196 > Fs
sInDEL	5612340	*SRIM5234*	n.798C > +ACGGCGGCT	a.Arg267 > Fs
sInDEL	5729552	*SRIM5347*	n.1750T > -C	a.Gly585 > Fs
sInDEL	5924804	*SRIM5524*	n.633C > -G	a.Asp214 > Fs
sInDEL	6067596	*SRIM5631*	n.1740C > -G	a.Gln583 > Fs
sInDEL	6070156	*SRIM5635*	n.592T > -C	a.Asp198 > Fs
sInDEL	6342322	*SRIM5890*	n.989T > +C	a.Glu331 > Fs
sInDEL	6756636	*SRIM6224*	n.2025G > -C	a.Pro676 > Fs
sInDEL	7363331	*SRIM6748*	n.957T > -G	a.Trp320 > *
sInDEL	9434499	*SRIM8595*	-GCGGGT	a.Arg188 and a.Thr189 > Del
SNC	9325969	*SRIM8505*	n.4914C > T	a.Leu1639 > Phe
SNC	831732	*SRIM0711*	n.1302C > T	a.Leu435 > Phe
SNC	5286231	*SRIM4920*	n.334G > A	a.Ser112 > Asn
SNC	1208866	*SRIM1055*	n.156G > C	a.Gly53 > Arg
SNC	1357766	*SRIM1204*	n.1185A > G	a.Lys396 > Glu
SNC	3133903	*SRIM2867*	n.652A > G	a.Asp218 > Gly
SNC	3719024	*SRIM3455*	n.712A > G	a.Glu238 > Gly
SNC	3759585	*SRIM3486*	n.1737C > T	a.Leu580 > Phe

SNC	4158608	*SRIM3866*	n.616A > G	a.His206 > Arg
SNC	5262280	*SRIM4897*	n.342T > C	a.Phe115 > Leu
SNC	5605498	*SRIM5229*	n.625C > A	a.Ala209 > Asp
SNC	5983812	*SRIM5567*	n.670C and n.669G > T	a.Ala224 > Phe
SNC	8192034	*SRIM7514*	n.408C > T	a.Leu137 > Phe

n = nucleotide and a = amino acid, numbers following represent the nucleotide number of the gene or amino acid number within the protein.

Table 4.5 Top pBLAST hits for proteins SNCs or sInDEL mutations within the main SEED subsystem.

Mutation Type	Protein	RAST Annotation	Homologue	Species	Identity
sInDEL	SRIM0252	Molybdenum transport system permease protein ModB	ABC transporter ATPase	*Streptomyces albulus* CCRC 11814	79 %
sInDEL	SRIM0371	D-amino-acid oxidase	Putative D-amino acid oxidase	*Streptomyces auratus* AGR0001	69 %
sInDEL	SRIM0432	Putative integral membrane protein	Integral membrane protein	*Streptomyces auratus* AGR0001	81 %
sInDEL	SRIM0587	Hypothetical protein	PAS fold-4 domain protein	*Streptomyces violaceusniger* Tu 4113	53 %
sInDEL	SRIM0805	Methylated-DNA--protein-cysteine methyltransferase	Putative bifunctional transcriptional activator/DNA repair enzyme AlkA	*Streptomyces davawensis* JCM 4913	61 %
sInDEL	SRIM0921	Dolichol-phosphate mannosyltransferase	Glycosyl transferase family 1	*Streptomyces albulus* CCRC 11814	80 %
sInDEL	SRIM0927	N-acetylmuramoyl-L-alanine amidase	N-acetylmuramyl-L-alanine amidase, negative regulator of AmpC, AmpD	*Streptomyces auratus* AGR0001	75 %
sInDEL	SRIM0982	Serine phosphatase RsbU, regulator of sigma subunit	Putative Magnesium or manganese-dependent protein phosphatase	*Streptomyces viridochromogenes* Tue57	64 %
sInDEL	SRIM1134	Putative ROK-family transcriptional regulator	ROK-family transcriptional regulator	*Streptomyces sp.* e14	63 %
sInDEL	SRIM1499	Hypothetical protein	Uncharacterized protein	*Streptomyces bingchenggensis* BCW-1	57 %
sInDEL	SRIM1633	Hypothetical protein	Putative integral membrane protein	*Streptomyces coelicolor* A3(2)	52 %
sInDEL	SRIM2037	Hypothetical protein	Putative uncharacterized protein	*Roseiflexus sp.* RS-1	31 %
SNC	SRIM2867	Transcriptional regulator, IclR family	IclR family transcriptional regulator	*Streptomyces auratus* AGR0001	87 %
sInDEL	SRIM2998	Putative transport protein	Uncharacterized protein	*Streptomyces albulus*	71 %

sInDEL	SRIM3121	Histidine ammonia-lyase	Phenylalanine/histidine ammonia-lyase	*Streptomyces auratus* AGR0001	63 %
sInDEL	SRIM3437	ATP-dependent DNA helicase RecQ	ATP-dependent DNA helicase RecQ	*Streptomyces albulus* CCRC 11814	83 %
SNC	SRIM4545	Long-chain-fatty-acid--CoA ligase	Amino acid adenylation	*Frankia sp.* CcI3	47 %
sInDEL	SRIM4817	Predicted N-acetyl-glucosamine kinase 2, ROK family	Sugar kinase	*Streptomyces sp.* C	83 %
sInDEL	SRIM5234	Hypothetical protein	Uncharacterized protein	*Streptomyces albulus* CCRC 11814	68 %
sInDEL	SRIM5347	Hypothetical protein	Group 1 glycosyl transferase	*Streptomyces auratus* AGR0001	63 %
sInDEL	SRIM5524	Putative secreted protein	Uncharacterized protein	*Streptomyces roseochromogenes subsp. oscitans*	74 %
sInDEL	SRIM5631	Hypothetical protein	Uncharacterized protein	*Streptomyces auratus* AGR0001	59 %
sInDEL	SRIM5635	Putative alkanal monooxygenase (luciferase)	Alkanal monooxygenase	*Streptomyces bingchenggensis* BCW-1	85 %
sInDEL	SRIM5890	Dihydroxyacetone kinase family protein	DAK2 domain fusion protein YloV	*Streptomyces auratus* AGR0001	65 %
sInDEL	SRIM6224	Long-chain-fatty-acid--CoA ligase	Amino acid adenylation enzyme/thioester reductase family protein	*Streptomyces sp.* SM8	47 %
sInDEL	SRIM6748	Respiratory nitrate reductase beta chain	Respiratory nitrate reductase beta chain	*Streptomyces lividans* 1326	87 %
sInDEL	SRIM8595	Probable NreB protein	Uncharacterized protein	*Streptomyces aurantiacus* JA 4570	69 %
SNC	SRIM8505	Long-chain-fatty-acid--CoA ligase	Non-ribosomal peptide synthetase	*Streptomyces clavuligerus* ATCC 27064	65 %
SNC	SRIM0711	Putative RNA polymerase ECF-subfamily sigma factor	Putative RNA polymerase ECF-subfamily sigma factor	*Streptomyces venezuelae* ATCC 10712	77 %
SNC	SRIM4920	Anti-sigma F factor antagonist	Anti-sigma factor antagonist	*Streptomyces auratus* AGR0001	94 %
SNC	SRIM1055	Putative transcriptional regulatory protein	DeoR family transcriptional regulator	*Streptomyces auratus* AGR0001	90 %

SNC	SRIM1204	GTP pyrophosphokinase	ppGpp synthetase	*Streptomyces natalensis*	92 %
SNC	SRIM3455	Putative DNA-binding protein	XRE family transcriptional regulator	*Streptomyces albulus* CCRC 11814	89 %
SNC	SRIM3486	Regulatory protein	Regulator	*Streptomyces albulus* CCRC 11814	78 %
SNC	SRIM3866	Aromatic hydrocarbon utilization transcriptional regulator CatR	Uncharacterized protein	*Streptomyces albulus* CCRC 11814	83 %
SNC	SRIM4897	Transcriptional regulator, TetR family	TetR family transcriptional regulator	*Streptomyces auratus* AGR0001	85 %
SNC	SRIM5229	Serine phosphatase RsbU, regulator of sigma subunit	Uncharacterized protein	*Streptomyces auratus* AGR0001	95 %
SNC	SRIM5567	DNA-binding Protein	DNA-binding protein	*Streptomyces cattleya* NRRL 8057	81 %
SNC	SRIM7514	cAMP-binding proteins	Crp/Fnr family transcriptional regulator	*Streptomyces albulus* CCRC 11814	86 %



Figure 4.5 CLC amino acid alignment of the N-terminal catalytic domain of RelA from various Gram positive and Gram negative organisms. The numbers listed on the right hand side correspond to the amino acid position of the last amino acid listed. The amino acid sequence was obtained from Uniprot, with numbers listed. The organisms from which the sequences were obtained were: *S. clavuligeris, S. coelicolor* A3(2), *S. natalensis, Mycobacterium tuberculosis, Streptococcus dysgalactiae subsp. equisimilis, Bacillus subtilis, Clostridium difficile, Escherichia coli, Salmonella typhi, Klebsiella pneumoniae, Yersinia pestis, Pseudomonas aeruginosa.*

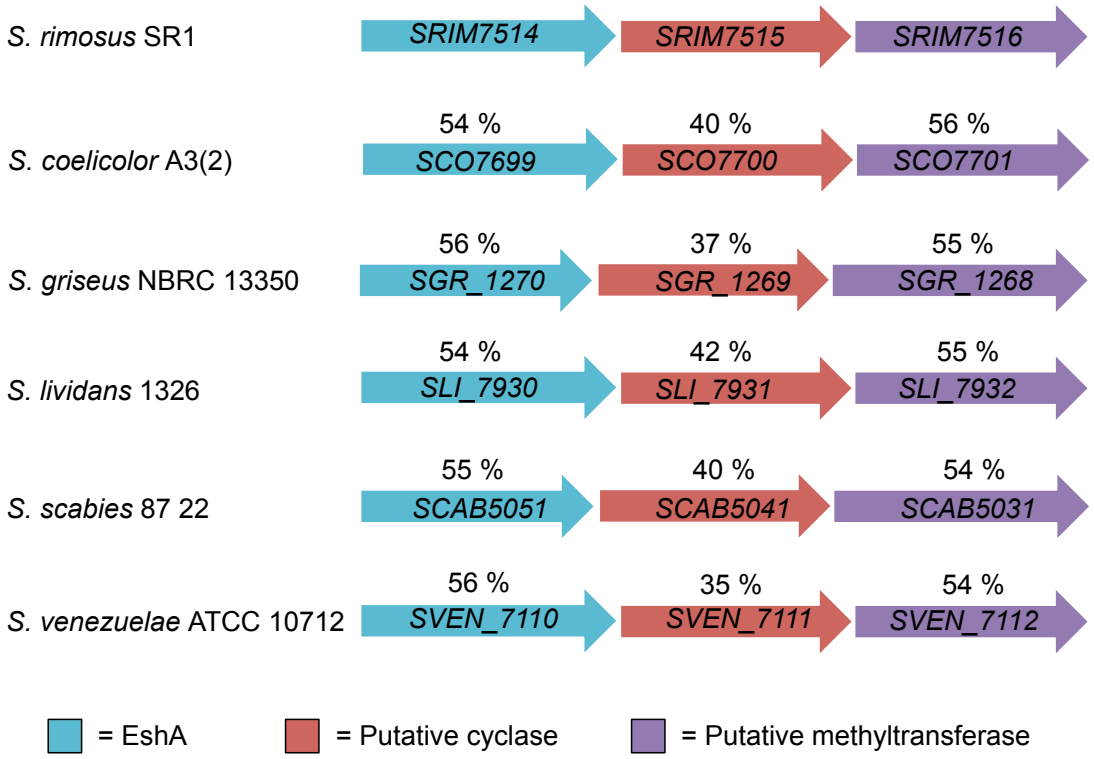


Figure 4.6 Chromosomal synteny of *eshA* (*SRIM7514*) between selected *Streptomyces* species. Percentage identity of the encoded proteins by these genes to the *S. rimosus* SR1 representatives are shown above the corresponding gene. Products encoded are shown.

4.5 Genomic comparisons between *S. rimosus* SR1 and *S. rimosus* SR50S

4.5.1 SR50S has lost a gene putatively linked with CutRS

As discussed previously SR50S is an OTC minus strain that was generated through a spontaneous loss of the OTC cluster from SR50. This provided two deletion profiles that were examined; however, for the purposes of this study, to determine the basis of increased OTC productivity, the SR50 deletion profile was focused on. In addition to the genomic deletions previously discussed for SR20 two additional regions were identified for SR50, one of 142,506 bp and 11,912 bp (Figure 4.7). These deletion regions occurred at the same point as the 69,334 bp deletion of SR20, as seen in Figure 4.7, with an additional 152 genes missing for SR50. Figure 4.8 represents an overview of the gene clusters within the 142 kbp region, present are genes encoding putative vancomycin resistance, DNA repair, membrane transport proteins and a NRPS cluster, with the central region containing mainly unassociated miscellaneous and hypothetical genes. The 12 kbp region consisted mainly of hypothetical genes, with some stress response and regulation associated genes present as well. As with the NRPS polymorphisms for SR20, there is the potential for a similar influence on OTC production from the loss of this NRPS cluster.

The main SEED subsystems identified for the 152 genes were; miscellaneous, cell wall and capsule, virulence, disease and defense, and phages, prophage, mobile elements and plasmids, as listed in Table 4.6. *SRIM7772, SRIM7786, SRIM7788, SRIM7796, SRIM7806, SRIM7808, SRIM7817,* and *SRIM7847* were all membrane associated, either as putative tranporters or integral membrane proteins, however only *SRIM7808* was predicted to have a connection with OTC production. *SRIM7808* was annotated as a sensor kinase similar to *cutS*, however unlike most two component regulatory systems *SRIM7808* was an orphan sensor kinase as no regulatory protein was found up or downstream of it. As SRIM7808 was

predicted to encode a CutS like protein an amino acid alignment with CutS from *S. coelicolor* A3(2) was performed. SRIM7808 only shared 63 % identity with CutS, however a nucleotide BLAST alignment of the *cutS* gene from *S. coelicolor* against the *S. rimosus* SR1 genome identified *SRIM6174*. SRIM6174 shared 72 % identity with CutS and lay downstream of a *cutR* homologue, *SRIM6173*. SRIM6173 shared 94 % identity with CutR, therefore this suggests that SRIM6173 and SRIM6174 are the primary homologues of CutS and CutR, respectively. The *cutRS* two component system negatively regulates actinorhodin (ACT) biosynthesis, though direct binding to *actII-ORF4* has yet to be shown (Tseng & Chen 1991; Rodríguez *et al.,* 2013). Although there has been no deletion of *SRIM6173* and *SRIM6174*, it is possible that SRIM7808 acts as a secondary sensor kinase for SRIM6173, given that SRIM7808 and SRIM6174 share 81 % identity, potentially increasing its repressive effect on antibiotic production.

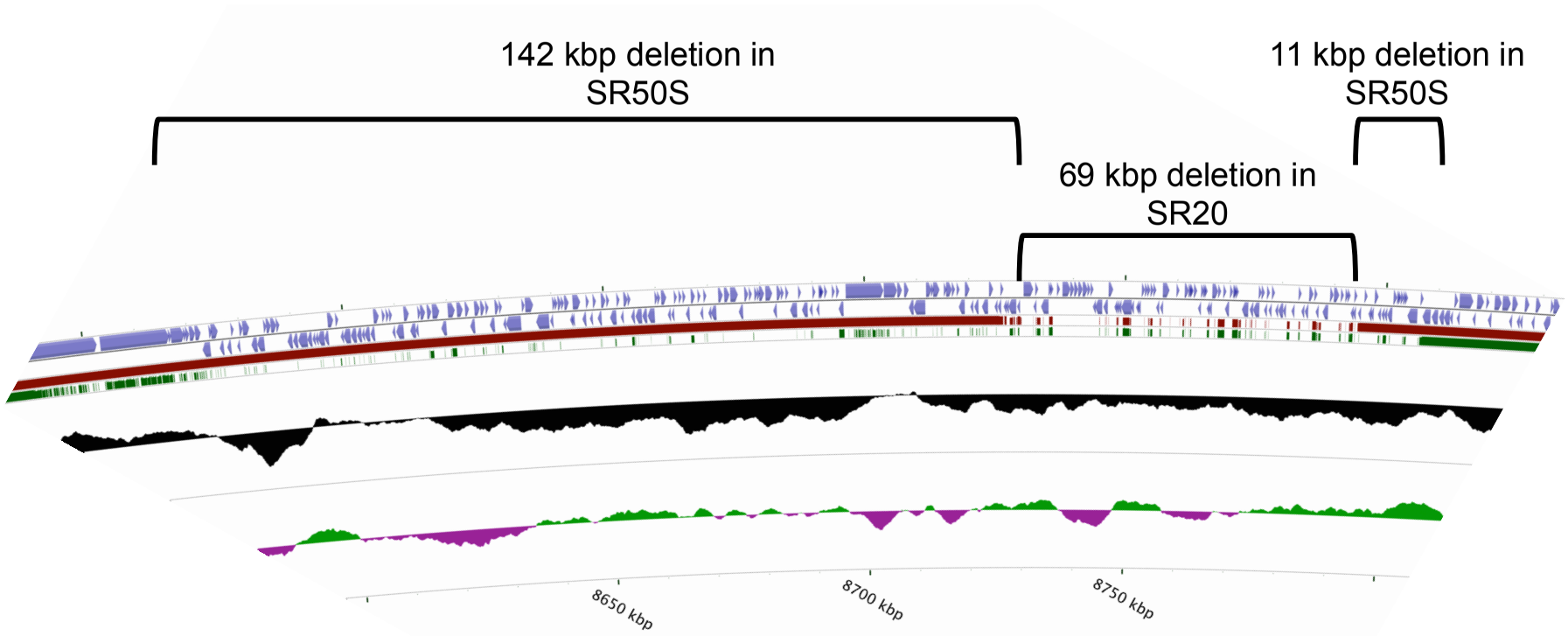


Figure 4.7 SR20 and SR50S sequences mapped to SR1 at the 69 kbp deletion point. The additional 142 kbp and 11 kbp deletion points of SR50S are shown, along with the original 69 kbp deletion in SR20. The red mapped reads represent SR20, with the green mapped reads representing SR50S.

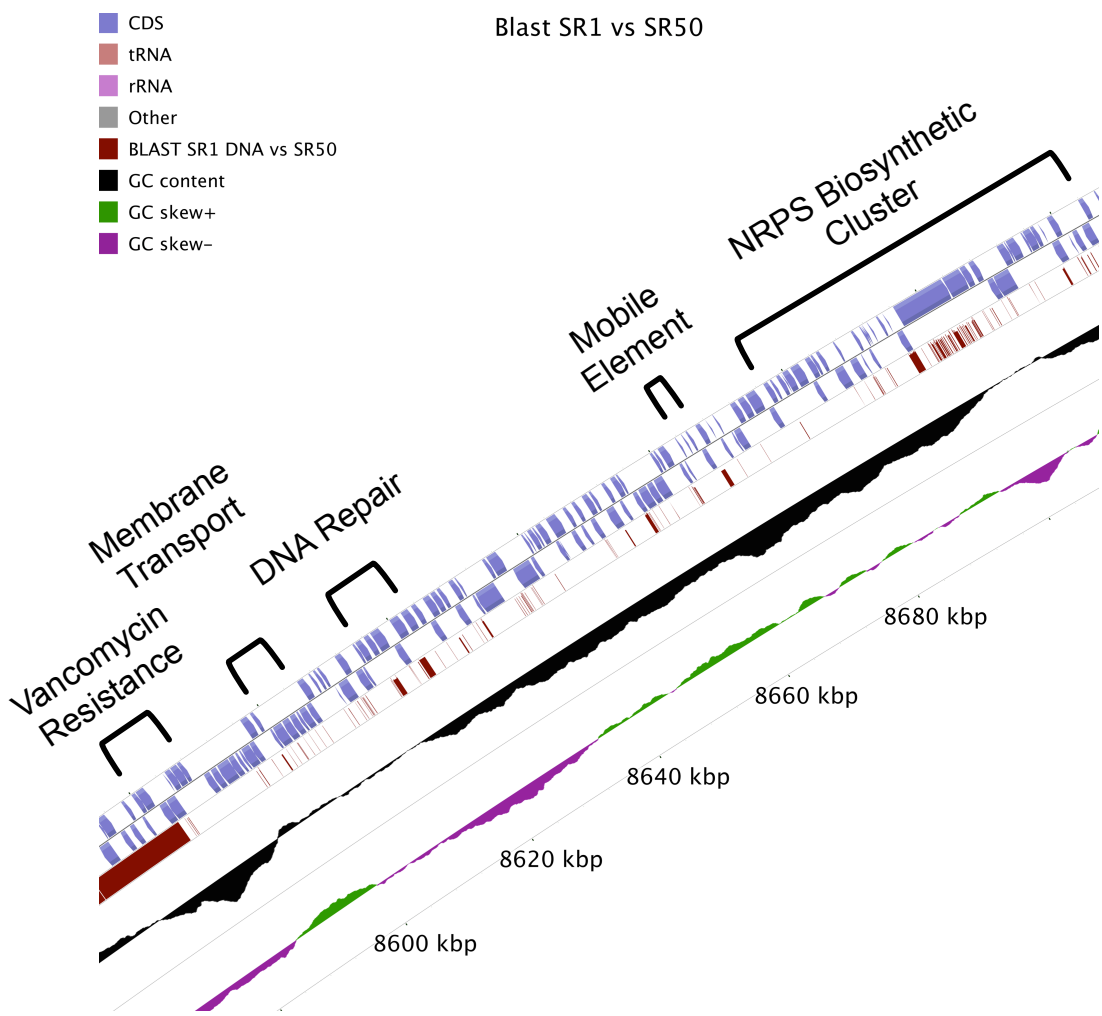


Figure 4.8 Zoomed CCT nBLAST map of the 142 kbp deletion region of SR50, before the start of the 69 kbp deletion region shared with SR20. As before small regions of homologus sequence can be seen as thin red lines, however viewing SR50S sequence reads mapped to SR1 in Artemis shows no mapping in these areas.

Table 4.6 Genes within the SR50 deletion profile contributing to the main SEED subsystems.

Gene	RAST Annotation	Homologue	Species	Identity
SRIM7772	UDP-N-acetylmuramoylalanyl-D-glutamyl-2,6-diaminopimelate--D-alanyl-D-alanine ligase	UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D-alanine ligase	uncultured bacterium esnapd26	91 %
SRIM7786	Secreted protein	Predicted protein	*Streptomyces sp.* C	54 %
SRIM7788	Aerobic C4-dicarboxylate transporter for fumarate	Putative C4-dicarboxylate transport protein	*Streptomyces aurantiacus* JA 4570	81 %
SRIM7796	protein of unknown function DUF1360	Integral membrane protein	*Streptomyces griseoaurantiacus* M045	65 %
SRIM7806	integral membrane protein	Uncharacterized protein	*Streptomyces aurantiacus* JA 4570	73 %
SRIM7808	Sensor protein cutS	Integral membrane sensor signal transduction histidine kinase	*Streptomyces auratus* AGR0001	77 %
SRIM7817	transporter	Uncharacterized protein	*Streptomyces scabies* 87.22	75 %
SRIM7820	Secreted protein	Uncharacterized protein	*Streptomyces niveus* NCIMB 11891	47 %
SRIM7828	Hemolysin-type calcium-binding region	Uncharacterized protein	*Streptomyces aurantiacus* JA 4570	72 %
SRIM7838	Cupin 2, conserved barrel domain protein	Cupin 2 conserved barrel domain protein	*Streptomyces violaceusniger* Tu 4113	80 %
SRIM7839	D-alanyl-D-alanine carboxypeptidase	Beta-lactamase	*Streptomyces himastatinicus* ATCC 53653	53 %
SRIM7842	Inositol-1-monophosphatase	Putative phosphatase SCF76.19c	*Streptomyces himastatinicus* ATCC 53653	91 %
SRIM7847	MFS permease protein	Putative integral membrane protein	*Streptomyces cattleya* DSM 46488	73 %
SRIM7854	putative transposase	Uncharacterized protein	*Streptomyces niveus* NCIMB 11891	77 %
SRIM7855	Transposase	Putative transposase	*Streptomyces tenjimariensis*	51 %

SRIM7865	L-carnitine dehydratase/bile acid-inducible protein F	Uncharacterized protein	*Streptomyces sp.* HPH0547	85 %
SRIM7874	phytanoyl-CoA dioxygenase family protein	Phytanoyl-CoA dioxygenase	*Pseudomonas sp.* EGD-AK9	70 %
SRIM7876	similar to membrane protein	Putative uncharacterized protein	*Streptomyces cinnamonensis*	89 %
SRIM7877	probable oxidoreductase	Oxidoreductase	*Mycobacterium abscessus subsp bolletii*	55 %
SRIM7892	probable short chain dehydrogenase	Short-chain dehydrogenase	*Pseudomonas aeruginosa* BL06	59 %
SRIM7893	Dehydrogenases with different specificities	Short-chain dehydrogenase/reductase SDR	*Streptomyces sp.* Mg1	50 %
SRIM7903	blr4874; putative alcohol dehydrogenase	Alcohol dehydrogenase	*Streptomyces hygroscopicus subsp. jinggangensis* TL01	92 %
SRIM7907	D-alanyl-D-alanine carboxypeptidase	Uncharacterized protein	*Streptomyces niveus* NCIMB 11891	79 %
SRIM8005	probable oxidoreductase/Short-chain dehydrogenase	Short-chain dehydrogenase	*Actinomadura madurae* LIID-AJ290	93 %
SRIM8010	monooxygenase	Monooxygenase	*Streptomyces collinus* Tu 365	82 %
SRIM8013	integrase	Integrase	*Streptomyces sp.* PAMC26508	76 %

The amino acid sequence encoded by each gene was subjected to pBLAST via Uniprot. Top pBLAST hits are listed

4.5.2 SR50S contains a sInDEL within a homologue of *sigQ*

Of the twelve sInDEL that were identified between SR1 and SR50S, six were present in CDS and a total of 223 SNCs were identified. Of the SNCs 79 were synonymous, 96 non-synonymous and 48 intergenic. The major SEED subsystems identified for these SNCs were secondary metabolism, nitrogen metabolism, and miscellaneous. The CDSs found within these subsystems, as well as the six sInDEL and genes in which a SNC caused a non-stop or stop codon substitution, are listed in Table 4.7, along with the predicted impact of the mutations. Only ten SNCs and sInDEL were considered to cause frame-shifts or stop or non-stop codon changes. pBLAST homologues were ascertained for these SNCs and sInDEL as listed in Table 4.8. Additional SEED subsytems identified from pBLAST hits for high impact mutations were; virulence, defense and disease, hypothetical, sigma factors, amino acids and derivatives, cell wall and capsule, protein metabolism and secondary metabolism.

One frame-shift mutation was within SRIM5103, a putative RNA polymerase ECF subfamily sigma factor. Further examination determined that SRIM5103 was a homologue of SigQ, sharing 69 % identity with SCO4908. *SRIM5103* lies downstream of *SRIM5102* and *SRIM5101*, homologues of *afsQ1* and *afsQ2* respectively, sharing synteny with *S. coelicolor* A3(2) and other *Streptomyces* species. In *S. coelicolor* A3(2) deletion of *sigQ* increases antibiotic production and delays the formation of aerial hyphae when grown on minimal media with glutamate as the sole nitrogen source (Shu *et al.,* 2008). It is likely that the frameshift predicted in SRIM5103 mimics a *sigQ* deletion and therefore leads to an increase in OTC production, however given the role SigQ plays, this would be in a medium-dependent manner. Another CDS a predicted frame-shift mutation was *SRIM6098*, a predicted tRNA delta(2)-isopentenylpyrophosphate transferase and close homologue of *SCO5791*. Both SRIM6098 and SCO5791 share homology with the tRNA dimethylallyltransferase of *E. coli*, named MiaA. Work undertaken to discern the role of MiaA in tRNA modification showed

that insertion mutation of *miaA* caused transversion changes GC > TA opposed to GC > AT transitions with a *mutL* mutation (Connolly & Winkler 1991). It was hypothesised that the mutation present in *SRIM6098* would cause a similar preference for transversion over transition mutations and in order to establish this the percentage of transition/transversion mutatons for SR20, SR50S and SR55 were examined. From the results shown in Table 4.9 there is a slight decrease in the percentage of transversion mutations for strains SR50S and SR55, but also a more concentrated and increased percentage of C > T and G > A transitions, rather than A > G or T > C. These results do not correlate with the hypothesis, though there was a distinct difference in transition mutations.

Table 4.7 CDS in which frame-shift, stop or non-stop condon changes were predicted as well as SNCs contributing to the main SEED subsystems.

Mutation Type	Reference Position	Gene	Coding Region Change	Amino Acid Change
SNC	283710	*SRIM0236*	n.1047C > T	a.350Gln > *
sInDEL	623015	*SRIM0519*	n.1469A > -C	a.490Arg > Fs
SNC	758633	*SRIM0645*	n.1124G > A	a.375Trp > *
SNC	2123050	*SRIM1903*	n.383G > A	a.128Trp > *
sInDEL	3067931	*SRIM2810*	n.3077A > +C	a.1029Ser > Fs
sInDEL	5471205	*SRIM5103*	n.108C > +G	a.39Thr > Fs
sInDEL	5924030	*SRIM5523*	n.642A > -C	a.215Thr > Fs
sInDEL	6582771	*SRIM6098*	n.920G > -C	a.310Gly > Fs
sInDEL	7842807	*SRIM7178*	n.900G > -A	a.301Asp > Fs
SNC	9500828	*SRIM8650*	n.792C > T	a.265 Arg > *
SNC	403203	*SRIM0342*	n.217C > T	a.73Ala > Val
SNC	5930862	*SRIM5528*	n.3051G > A	a.1018Glu > Lys
SNC	6732462	*SRIM6208*	n.4305C > T	a.1436Pro > Ser
SNC	6760111	*SRIM6225*	n.11097G > A	a.3700Val > Met
SNC	6775330	*SRIM6226*	n.2352C > T	a.785Pro > Ser
SNC	6859246	*SRIM6298*	n.225T > C	a.76Trp > Leu
SNC	7197152	*SRIM6596*	n.453G > A	a.152Ala > Thr
SNC	8870359	*SRIM8075*	n.786G > A	a.263Val > Ile
SNC	8875085	*SRIM8075*	n.5512C > T	a.1838Ala > Val
SNC	3807356	*SRIM3537*	n.591G > A	a.198Glu > Lys
SNC	8007573	*SRIM7328*	n.1084A > G	a.362His > Arg
SNC	8007988	*SRIM7328*	n.669G > T	a.224Ala > Ser
SNC	3597851	*SRIM3292*	n.1011G > A	a.338Gly > Arg
SNC	5711956	*SRIM5329*	n.205C > T	a.69Ala > Val
SNC	6614162	*SRIM6124*	n.24G > A	a.9Asp > Asn

SNC	7301677	*SRIM6700*	n.426G > A	a.143Gly > Ser
SNC	8862853	*SRIM8069*	n.778C > T	a.260Thr > Ile
SNC	8901185	*SRIM8090*	n.1105C > G	a.369Ser > Gly

n = nucleotide and a = amino acid, numbers following represent the nucleotide number of the gene or amino acid number within the protein.

Table 4.8 Top pBLAST hits for proteins SNCs or sInDEL mutations within the main SEED subsystem.

Mutation Type	Protein	RAST Annotation	Homologue	Species	Identity
SNC	SRIM0236	Putative tellurium resistance protein	Mrr restriction system protein	Streptomyces auratus AGR0001	77%
sInDEL	SRIM0519	Hypothetical protein	ATPase-like protein	Streptosporangium roseum ATCC 12428	39%
SNC	SRIM0645	Tyrosinase	O-aminophenol oxidase	Streptomyces griseus subsp. griseus	64%
SNC	SRIM1903	Putative membrane protein	Putative ABC transporter integral membrane subunit	Streptomyces auratus AGR0001	56%
sInDEL	SRIM2810	Hypothetical protein	Uncharacterized protein	Streptomyces viridochromogenes Tue57	32%
sInDEL	SRIM5103	Putative RNA polymerase ECF subfamily sigma factor	ECF subfamily RNA polymerase sigma factor	Streptomyces auratus AGR0001	88%
sInDEL	SRIM5523	Hypothetical protein	Uncharacterized protein	Streptomyces venezuelae ATCC 10712	41%
sInDEL	SRIM6098	tRNA delta(2)-isopentenylpyrophosphate transferase	RNA dimethylallyltransferase	Streptomyces auratus AGR0001	92%
sInDEL	SRIM7178	Hypothetical protein	Uncharacterized protein	Streptomyces griseus subsp. Griseus	78%
SNC	SRIM8650	Long-chain-fatty-acid--CoA ligase	Peptide synthetase	Streptomyces lavendulae	74%
SNC	SRIM0342	Phytoene desaturase	Squalene/phytoene dehydrogenase	Streptomyces auratus AGR0001	79%
SNC	SRIM5528	Long-chain-fatty-acid--CoA ligase	Amino acid adenylation	Frankia sp. CcI3	51%
SNC	SRIM6208	Long-chain-fatty-acid--CoA	Putative non-ribosomal peptide	Rhodococcus opacus B4	55%
SNC	SRIM6225	Long-chain-fatty-acid--CoA ligase	NRPS protein	Streptomyces auratus AGR0001	45%
SNC	SRIM6226	Long-chain-fatty-acid--CoA ligase	Peptide synthetase 2	Streptomyces filamentosus	45%
SNC	SRIM6298	aminoglycoside-O-phosphotransferase VIII	Aminoglycoside 3'-phosphotransferase	Streptomyces fradiae	64%

SNC	SRIM6596	Penicillin acylase (Penicillin amidase)	Acyl-homoserine-lactone acylase	Pseudonocardia dioxanivorans 55486	58%
SNC	SRIM8075	Long-chain-fatty-acid--CoA ligase	HrmO	Streptomyces griseoflavus	50%
SNC	SRIM3537	BarH	Amidohydrolase family protein	Streptomyces turgidiscabies Car8	93%
SNC	SRIM7328	Respiratory nitrate reductase alpha chain	Nitrate reductase A subunit alpha	Streptomyces albulus CCRC 11814	82%
SNC	SRIM3292	Possible hydrolase	Putative peptidase	Streptomyces bingchenggensis BCW-1	79%
SNC	SRIM5329	Putative hydrolase	Alpha/beta hydrolase	Streptomyces auratus AGR0001	75%
SNC	SRIM6124	Putative hydrolase	ADP-ribosylation/Crystallin J1	Streptomyces auratus AGR0001	87%
SNC	SRIM6700	Secreted protein	Putative secreted protein	Streptomyces venezuelae ATCC 10712	59%
SNC	SRIM8069	Photopexin B	Uncharacterized protein	Nocardia cyriacigeorgica GUH-2	41%
SNC	SRIM8090	Lycopene beta cyclase	Lycopene cyclase	Streptomyces auratus AGR0001	69%

Table 4.9 Percentage of transition/transversion mutations of all SNCs for SR20, SR50S and SR55

Strain	Transition				Transversion							
	A > G	C > T	G > A	T > C	A > C	A > T	C > G	C > A	G > C	G > T	T > A	T > G
SR20	20 %	26 %	14 %	11 %	0 %	0 %	7 %	3 %	6 %	4 %	1 %	6 %
SR50S	4 %	36 %	33 %	8 %	3 %	0 %	4 %	2 %	4 %	3 %	0.5 %	3 %
SR55	2 %	41 %	48 %	2.4 %	0.5 %	0.25 %	1 %	1 %	1 %	0.4 %	0.5 %	0.4 %

The transition and transversion mutations are in relation to SR1.

4.6 Genomic comparisons between *S. rimosus* SR1 and *S. rimosus* SR55

4.6.1 SR55 has lost genes associated with nitrogen metabolism

The strain SR55 of *S. rimosus*, that produced 55 g/L of OTC, was isolated after further strain improvement of SR50 was undertaken. Two further genomic deletions, distinct from those previously discussed, occurred in SR55, one of 15,864 bp and another of 41,961 bp, resulting in the loss of 55 genes. Within the 42 kbp deletion there was a lantipeptide biosynthetic cluster identified using the antiSMASH program (Medema *et al.,* 2011), as well as general mobile element genes either side of the cluster (Figure 4.9), with hypothetical and genes associated with nitrogen metabolism present within the 16 kbp region. As was found with the mutations in SR20 and the deletion profile of SR50 there is a loss of a NRPS cluster.

Of the 55 genes identified the main SEED subsystems represented were; phages, prophage, mobile elements and plasmids, nitrogen metabolism, secondary metabolism, cell wall and capsule and miscellaneous. The genes associated with nitrogen metabolism were *SRIM7325, SRIM7326*, and *SRIM7327*, encoding NarI3, NarJ3 and NarH3 respectively (Table 4.10). As mentioned in Chapter 3, *S. rimosus* SR1 has four sets of the respiratory nitrate reductase (Nar) genes, two of which are *nar3* like. During the deletion analysis of SR20 a *narG3* gene, *SRIM7562*, was deleted, with the genes encoding the other subunit still present. It should be noted that *SRIM7562* encodes a form of NarG3 that is missing the central portion of the protein, and as proposed in the previous chapter, could be a reason for the presence of two sets. Therefore the impact from the loss of these genes are lessened as it could be assumed the NarG3 subunit, encoded by *SRIM7328*, can interact with the intact NarHIJ3 subunits encoded by *SRIM7561, SRIM7560* and *SRIM7559*. The role of the Nar3 set of respiratory nitrate reductases during growth has yet to be determined, however the *nar1* gene set have been shown to function exclusively in resting spores of *S. coelicolor* A3(2) (Fischer *et al.,* 2013). It is possible that

the loss of these subunits of the Nar3 respiratory nitrate reductases is the removal redundant genes, if it is assumed that both sets of Nar3 proteins in *S. rimosus* SR1 perform the same role and function.

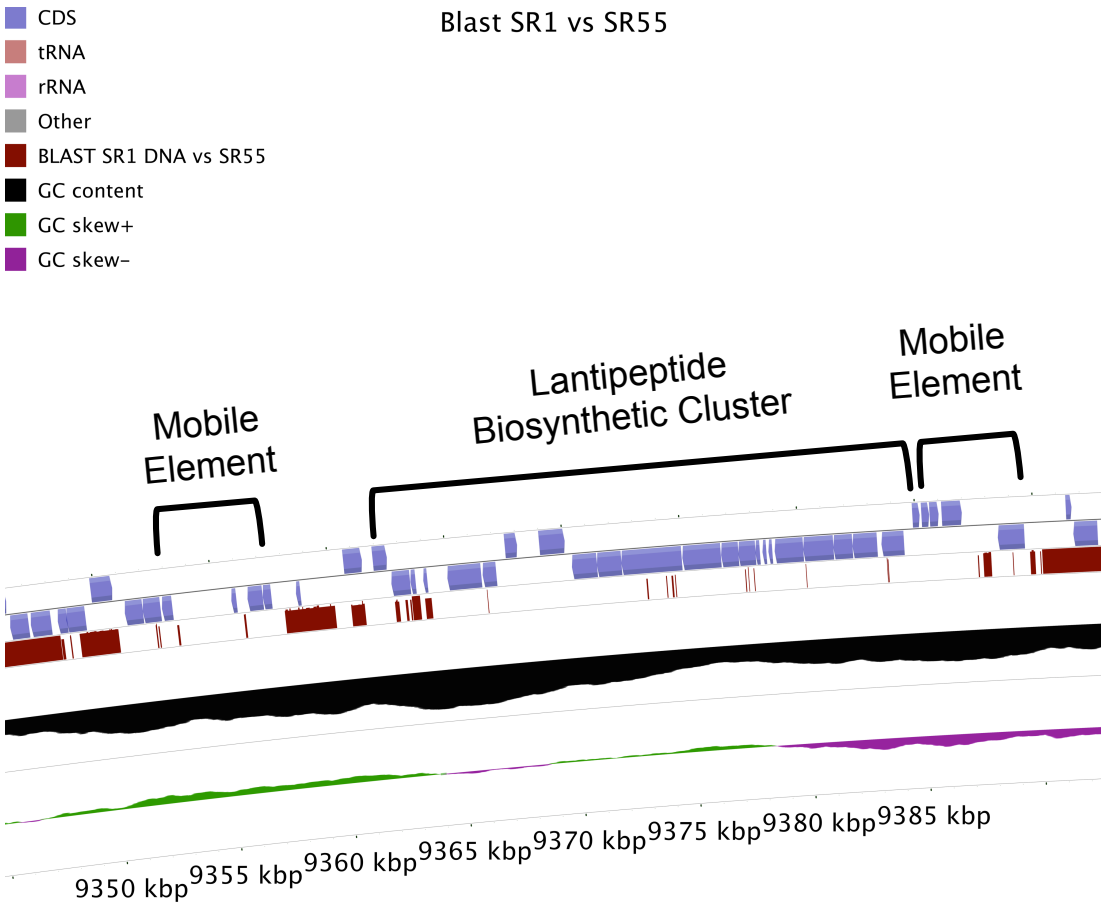


Figure 4.9 Zoomed CCT nBLAST map of the 41 kbp deletion region of SR55. As before small regions of homologue can be seen as thin red lines, however viewing SR50S sequence reads mapped to SR1 in Artemis shows no mapping in these areas.

Table 4.10 Genes within the SR55 deletion profile contributing to the main SEED subsystems.

Gene	RAST Annotation	Homologue	Species	Identity
SRIM7325	Respiratory nitrate reductase gamma chain	Uncharacterized protein	*Streptomyces roseochromogenes* subsp oscitans	79 %
SRIM7326	Respiratory nitrate reductase delta chain	Nitrate reductase delta chain NarJ3	*Streptomyces bingchenggensis* BCW-1	67 %
SRIM7327	Respiratory nitrate reductase beta chain	Nitrate reductase beta chain NarH3	*Streptomyces bingchenggensis* BCW-1	79 %
SRIM8523	Probable methyltransferase	Predicted protein	*Streptomyces sp*. C	55 %
SRIM8525	IS1647-like transposase	Transposase, IS4 family	*Streptomyces ipomoeae* 91-03	73 %
SRIM8526	Mobile element protein	Transposase, IS4 family	*Streptomyces ipomoeae* 91-03	87 %
SRIM8527	Mobile element protein	Transposase IS116/IS110/IS902 family protein	*Geodermatophilus obscurus* DSM 43160	68 %
SRIM8528	Mobile element protein	Transposase IS116/IS110/IS902 family protein	*Geodermatophilus obscurus* DSM 43160	59 %
SRIM8539	Glucose-1-phosphate thymidylyltransferase	Glucose-1-phosphate thymidylyltransferase	*Streptomyces griseoflavus* Tu 4000	68 %
SRIM8540	Glycosyltransferase involved in cell wall biogenesis	Glycosyl transferase family protein	*Nocardia brasiliensis* ATCC 700358	43 %
SRIM8541	Lanthionine biosynthesis protein LanL	Predicted protein	*Streptomyces sp*. C	60 %
SRIM8554	IS1647-like transposase	Uncharacterized protein	*Streptomyces roseochromogenes* subsp oscitans	85 %
SRIM8555	Mobile element protein	Transposase IS5	*Streptomyces albulus* CCRC 11814	74 %

The amino acid sequence encoded by each gene was subjected to pBLAST via Uniprot. Top pBLAST hits are listed

4.6.2 The presence of non-synonymous SNC within a potential regulator of the OTC biosynthetic gene cluster

A total of five sInDELs and 696 SNCs were identified between SR1 and SR55. Four of the sInDELs were intragenic, 213 SNCs synonymous, 360 non-synonymous and 126 were intergenic. The major SEED subsystems identified for these SNCs secondary metabolism, miscellaneous and with cell wall and capsule, DNA metabolism, nucleotides and nucleosides, and virulence, disease and defense. The four DIPs and nonsense/nonstart SNCs, are listed in Table 4.11, 19 were predicted to cause frame-shift mutations. From initial annotation and pBLAST homologues (Table 4.12) the majority of the mutations were within hypothetical/uncharacterized proteins.

During the strain development process Pfizer selected for 2-deoxyglucose (2-DOG) resistant isolates of *S. rimosus*, before the isolation of SR55, as these strains were no longer subject to carbon catabolite repression (CCR) (Kwakman & Postma 1994). Originally loss or mutation of *glK* was thought to be wholly responsible for the generation of 2-DOG mutants (Kwakman & Postma 1994); however, it was later established that another gene was involved, *SCO2127* (Ramos *et al.,* 2004; Guzman 2005). Glucose transport itself was also established as a means of CCR, by activating the Glk (van Wezel *et al.,* 2007). As no genes were identified within the deletion profile of SR55 that had a putative link to 2-DOG resistance, SNC/DIPs were inspected. The homologue of *glk* in *S. rimosus* is *SRIM1977*, with *SRIM1978* being the homologue of *SCO2127*, and *SRIM5909* the homologue of *glcP1*/*glcP2*; however, no mutations could be identified within these genes. SCO2127 was found to interact with BldKB and SCO2582, a lipoprotein ABC-type permease and a putative membrane metalloendopeptidase respectively (Chávez *et al.,* 2010). Again the equivalent protein encoding CDSs, *SRIM5280* and *SRIM2624*, were examined in *S. rimosus* SR55 but no mutations could be detected in either.

It appears that 2-DOG resistance was achieved in *S. rimosus* SR55 via a different mechanism.

BldKB is a membrane protein and SCO2582 is also predicted to be a membrane protein. It has been proposed that before these proteins are incorporated into the membrane SCO2127 binds to them and exert its influence on them (Chávez *et al.,* 2010). BldKB is ADP-ribosylated and it was predicted that this ADP-ribosylation took place at a conserved cysteine residue within the consensus sequence of a cleavage site in the signal peptide (Sugawara *et al.,* 2003). SRIM5280 shares the same conserved cysteine residue within the predicted signal peptide cleavage site (L(A)T(A)ACG). Sugawara *et al.,* proposed that if BldKB was ADP-ribosylated then it could not be exported to the surface, lipid-modifed, or function in transport; it may be that ADP-ribosylation ensures the BldKB remains within the cytoplasm long enough for SCO2127 to have a regulatory effect. For this reason it was hypothesised that lack of ADP-ribosylation of SRIM5280 may be the reason for 2-DOG resistance as SRIM1978 (SCO2127 homologue) could not bind and regulate the protein. From the SNC analysis of SR55 two genes predicted to encode adenine phosphoribosyltransferases, *SRIM1886* and *SRIM1205*, were identified with SNC of a tryptophan to stop codon change and a valine to methionine change, respectively. These SNC would support the hypothesis for an alternative mechanism for 2-DOG resistance.

Six genes are associated with virulence, disease and defense, of these genes *SRIM1265* was selected to examine further as it encoded a multidrug efflux pump. Pfam (Punta *et al.,* 2011) analysis places SRIM1265 in the major facilitator superfamily of proteins, with SMART (Letunic *et al.,* 2011) defining this further with SRIM1265 as a member of the 14 transmembrane segments (TMS) of 14-TMS subfamily. In addition to this SRIM1265 is also a homologue of PqrB and EpeA, sharing 71 % and 70 % identity respectively. PqrB is associated with paraquat resistance in *S. coelicolor* (Cho *et al.,* 2003), with EpeA providing multidrug resistance in *S. clavuligeris* (Rodríguez-García *et al.,* 2006), suggesting that SRIM1265 may

provide multidrug resistance. The mutation seen in SRIM1265 may improve the protein's function, allowing for the efflux of OTC, as increased production would require increased export. The position of the mutation is within a TMS domain known as motif D (Paulsen *et al.,* 1996; Rodríguez-García *et al.,* 2006) and is not at a conserved position. It is possible that this mutation alters substrate specificity by altering the configuration of the membrane, thus allowing export of OTC, however very little is known about motif D.

In addition to the mutations discussed above, one more was identified within *SRIM7546*, a gene that lies downstream of the OTC exporter encoding gene, *otrB*. *SRIM7546* is a predicted regulatory gene and was putatively associated with the OTC biosynthetic cluster due to its proximity. Initial SMART analysis identified an N-terminal transcriptional regulatory (Trans-Reg) domain and a bacterial transcriptional activator domain at the carboxy-terminal (C-terminal), features common in *Streptomyces* antibiotic regulatory proteins (SARP). Due to the SNC present in the gene, and the potential of playing a role in the regulation of the OTC cluster *SRIM7546* was selected to investigate further (Chapter 7), in order to elucidate its role and the effect of the SNC.

Table 4.11 CDS in which frame-shift, stop or non-stop condon changes were predicted as well as SNCs contributing to the main SEED subsystems.

Mutation Type	Reference Position	Gene	Coding Region Change	Amino Acid Change
SNC	852060	*SRIM0728*	n.1T > C	a.Met1 > Thr
SNC	852060	*SRIM0729*	n.1047T > C	a.350* > R
sInDEL	1171866	*SRIM1011*	n.507G > -C	a.171Pro > Fs
SNC	1175771	*SRIM1015*	n.923G > A	a.308Trp > *
SNC	2106838	*SRIM1886*	n.251C > T	a.84Trp > *
SNC	4208015	*SRIM3906*	n.727G > A	a.243Trp > *
SNC	4326517	*SRIM4005*	n.35G > A	a.12Trp > *
SNC	4349529	*SRIM4028*	n.40G > A	a.14Trp > *
SNC	5055532	*SRIM4695*	n.260C > T	a.87Trp > *
SNC	5159468	*SRIM4802*	n.737C > T	a.246Trp > *
SNC	5381705	*SRIM5009*	n.1127G > A	a.376Trp > *
sInDEL	5924804	*SRIM5524*	n.633C > -G	a.214Asp > Fs
SNC	6260044	*SRIM5803*	n.170G > A	a.57Trp > *
SNC	6455170	*SRIM5991*	n.360C > T	a.121Gln > *
SNC	7690190	*SRIM7031*	n.278G > A	a.93Trp > *
sInDEL	7829087	*SRIM7168*	n.22G > -C	a.8Ala > Fs
sInDEL	8139290	*SRIM7476*	n.1369C > -G	a.457Arg > Fs
SNC	8565017	*SRIM7755*	n.11750G > A	a.3917Trp > *
SNC	8816496	*SRIM8021*	n.2499C > T	a.834Gln > *
SNC	3573713	*SRIM3267*	n.982C > T	a.328Ser > Phe
SNC	4872938	*SRIM4542*	n.1692G > A	a.565Ala > Thr
SNC	4886535	*SRIM4543*	n.6012 C > T	a.2005His > Tyr
SNC	4887055	*SRIM4543*	n.6532G > A	a.2178Gly > Asp
SNC	4895234	*SRIM4544*	n.1533G > A	a.512Gly > Arg
SNC	6730469	*SRIM6206*	n.615C > T	a.206Leu > Phe

SNC	7926586	*SRIM7257*	n.321G > A	a.108Glu > Lys
SNC	8566057	*SRIM7755*	n.12790C > T	a.4264Ala > Val
SNC	8877612	*SRIM8075*	n.8039G > C	a.2680Glu > Asp
SNC	9254071	*SRIM8445*	n.403G > A	a.135Gly > Asp
SNC	9325969	*SRIM8505*	n.4914C > T	a.1639Leu > Phe
SNC	9330168	*SRIM8505*	n.715C > T	a.239Ala > Val
SNC	9461163	*SRIM8620*	n.649C > T	a.217Thr > Met
SNC	1020663	*SRIM0883*	n.886C > T	a.296Ala > Val
SNC	1169056	*SRIM1008*	n.558C > T	a.187Pro > Ser
SNC	1662249	*SRIM1495*	n.699C > T	a.234Pro > Ser
SNC	2580519	*SRIM2376*	n.1351G > A	a.451Gly > Asp
SNC	3030651	*SRIM2785*	n.178G > A	a.60Ala > Val
SNC	4393547	*SRIM4074*	n.57G > A	a.20Ala > Thr
SNC	4568596	*SRIM4239*	n.118C > T	a.40Ala > Val
SNC	5036405	*SRIM4676*	n.651C > A	a.218Leu > Met
SNC	5393938	*SRIM5022*	n.64A > G	a.22Lys > Arg
SNC	5548452	*SRIM5176*	n.256G > A	a.86Thr > Ile
SNC	6070973	*SRIM5636*	n.36A > G and n.37A > C	a.13Lys > Ala
SNC	6106609	*SRIM5670*	n.43C > T	a.15Ala > Val
SNC	6352482	*SRIM5903*	n.869G > A	a.290Met > Ile
SNC	6896575	*SRIM6334*	n.1032G > A	a.345Glu > Lys
SNC	7453470	*SRIM6832*	n.93T > C	a.32Tyr > His
SNC	7698887	*SRIM7038*	n.43C > T	a.15Thr > Ile
SNC	7836642	*SRIM7173*	n.411C > T	a.138Arg > Cys
SNC	8048246	*SRIM7373*	n.636C > T	a.213Pro > Ser
SNC	9452656	*SRIM8611*	n.490G > A	a.164Gly > Asp
SNC	884763	*SRIM0760*	n.454G > A	a.152Gly > Asp
SNC	916781	*SRIM0792*	n.1240G > A	a.414Gly > Asp
SNC	1322159	*SRIM1171*	n.811C > T	a.271Thr > Ile

SNC	1323519	*SRIM1172*	n.2854G > A	a.952Gly > Asp
SNC	1326890	*SRIM1173*	n.657C > T	a.220Arg > Cys
SNC	1359427	*SRIM1205*	n.183G > A	a.62Val > Met
SNC	5484240	*SRIM5115*	n.139C > T	a.47Ala > Val
SNC	5748227	*SRIM5359*	n.858G > A	a.287Val > Ile
SNC	6400069	*SRIM5944*	n.936C > T	a.313Leu > Phe
SNC	9188924	*SRIM8371*	n.510G > A	a.171Glu > Lys
SNC	1376698	*SRIM1222*	n.1173G > A	a.392Val > Met
SNC	3324855	*SRIM3041*	n.421C > T	a.141Ala > Val
SNC	3336497	*SRIM3050*	n.802G > A	a.268Ala > Val
SNC	5098284	*SRIM4744*	n.129G > A	a.44Ala > Thr
SNC	5195345	*SRIM4837*	n.301G > A	a.101Gly > Asp
SNC	5209411	*SRIM4848*	n.541G > A	a.181Thr > Ile
SNC	5593437	*SRIM5219*	n.906C > T	a.303Arg > Trp
SNC	5642106	*SRIM5260*	n.2089G > A	a.697Gly > Asp
SNC	6466296	*SRIM6000*	n.1278G > A	a.427Val > Met
SNC	7988611	*SRIM7311*	n.400G > A	a.134Ala > Val
SNC	8840906	*SRIM8045*	n.421C > T	a.141Pro > Leu
SNC	281595	*SRIM0235*	n.154C > T	a.52Pro > Leu
SNC	904268	*SRIM0783*	n.780G > A	a.261Val > Met
SNC	1403796	*SRIM1253*	n.496G > A	a.166Cys > Tyr
SNC	1416721	*SRIM1265*	n.135G > A	a.46Gly > Ser
SNC	2891717	*SRIM2665*	n.1015C > T	a.339Thr > Ile
SNC	9442471	*SRIM8602*	n.186G > A	a.63Ala > Thr
SNC	1365065	*SRIM1210*	n.228G > A	a.77Ala > Thr
SNC	1965571	*SRIM1783*	n.1912G > A	a.638Gly > Asp
SNC	1965824	*SRIM1783*	n.1659G > A	a.554Asp > Asn
SNC	1979227	*SRIM1792*	n.393C > T	a.132Asp > Asn

SNC	2014899	*SRIM1822*	n.889C > T	a.297Ala > Val
SNC	4177362	*SRIM3879*	n.178G > C	a.60Arg > Pro
SNC	4432568	*SRIM4114*	n.2344G > A	a.782Gly > Asp
SNC	5081977	*SRIM4725*	n.1572G > A	a.525Val > Ile
SNC	6549391	*SRIM6070*	n.376G > A	a.126Gly > Asp
SNC	6649493	*SRIM6151*	n.706C > T	a.236Gly > Asp
SNC	7247033	*SRIM6638*	n.493C > T	a.165Pro > Leu
SNC	8049391	*SRIM7374*	n.40C > T	a.14Ser > Phe

n = nucleotide and a = amino acid, numbers following represent the nucleotide number of the gene or amino acid number within the protein.

Table 4.12 Top pBLAST hits for proteins SNCs or sInDEL mutations within the main SEED subsystem.

Mutation Type	Protein	RAST Annotation	Homologue	Species	Identity
SNC	SRIM0728	Inositol transport system ATP-binding protein	Inositol transport system ATP-binding protein	*Streptomyces sp.* PAMC26508	86 %
SNC	SRIM0729	Inositol transport system permease protein	Ribose ABC transporter permease	*Streptomyces albulus* CCRC 11814	90 %
sInDEL	SRIM1011	Transcriptional regulator, DeoR family	Uncharacterised protein	*Streptomyces afghaniensis* 772	86 %
SNC	SRIM1015	Hypothetical protein	NADH-ubiquinone/plastoquinone oxidoreductase	*Streptomyces auratus* AGR0001	70 %
SNC	SRIM1886	Adenine phosphoribosyltransferase	Phosphoribosyltransferase	*Clostridium cellulolyticum* ATCC 35319	53 %
SNC	SRIM3906	Putative intergral membrane protein	Uncharacterised protein	*Streptomyces albulus* CCRC 11814	64 %
SNC	SRIM4005	Hypothetical protein	Uncharacterised protein	*Streptomyces scabies* 87.22	36 %
SNC	SRIM4028	Hypothetical protein	Uncharacterised protein	*Streptomyces venezuelae* ATCC 10712	75 %
SNC	SRIM4695	DNA-binding protein	Uncharacterised protein	*Streptomycetaceae bacterium* MP113-05	65 %
SNC	SRIM4802	Putative oxidoreductase	Putative NADPH quinone oxidoreductase protein	*Streptomyces himastatinicus* ATCC 53653	76 %
SNC	SRIM5009	Glycine betaine ABC transport system	Glycine/betaine ABC transporter permease	*Streptomyces albulus* CCRC 11814	85 %
sInDEL	SRIM5524	Putative secreted protein	Uncharacterised protein	*Streptomyces roseochromogenes subsp oscitans*	74 %
SNC	SRIM5803	Hypothetical protein	Two-component system sensor kinase	*Streptomyces sp.* e14	48 %
SNC	SRIM5991	Hypothetical protein	Uncharacterised protein	*Streptomyces coelicoflavus* ZG0656	65 %
SNC	SRIM7031	Hypothetical protein	Uncharacterised protein	*Streptomyces ipomoeae* 91-03	54 %
sInDEL	SRIM7168	Hypothetical protein	Putative RNA-binding protein	*Amycolatopsis decaplanina* DSM 44594	55 %

sInDEL	SRIM7476	Hypothetical protein	Uncharacterised protein	*Amycolatopsis mediterranei* RB	35 %
SNC	SRIM7755	Long-chain-fatty-acid--CoA ligase	Non-ribosomal peptide synthetase	*Amycolatopsis mediterranei* RB	48 %
SNC	SRIM8021	Beta-galactosidase	Hydrolase	*Streptomyces pristinaespiralis* ATCC 25486	69 %
SNC	SRIM3267	Penicillin amidase precursor	Penicillin acylase	*Streptomyces auratus* AGR0001	80 %
SNC	SRIM4542	Long-chain-fatty-acid--CoA ligase	Linear gramicidin synthetase LgrD	*Streptomyces griseoflavus* Tu4000	48 %
SNC	SRIM4543	Long-chain-fatty-acid--CoA ligase	Amino acid adenylation domain protein	*Streptomyces violaceusniger* Tu 4113	47 %
SNC	SRIM4544	Long-chain-fatty-acid--CoA ligase	Putative NRPS	*Streptomyces sp*. W007	53 %
SNC	SRIM6206	Long-chain-fatty-acid--CoA ligase	Putative Linear gramicidin synthase subunit C	*Streptomyces aurantiacus* JA 4570	55 %
SNC	SRIM7257	Aminoglycoside 3'-phosphotransferase	Streptomycin 3''-kinase	*Streptomyces himastatinicus* ATCC 53653	73 %
SNC	SRIM7755	Long-chain-fatty-acid--CoA ligase	Non-ribosomal peptide synthetase	*Amycolatopsis mediterranei* RB	48 %
SNC	SRIM8075	Long-chain-fatty-acid--CoA ligase	HrmO	*Streptomyces griseoflavus*	50 %
SNC	SRIM8445	Protein-L-isoaspartate O-methyltransferase	Protein-L-isoaspartate	*Streptomyces griseus* XylebKG-1	43 %
SNC	SRIM8505	Long-chain-fatty-acid--CoA ligase	Non-ribosomal peptide synthetase	*Streptomyces clavuligerus* ATCC 27064	65 %
SNC	SRIM8620	Long-chain-fatty-acid--CoA ligase	Amino acid adenylation domain-containing protein	*Streptomyces sp*. HPH0547	57 %
SNC	SRIM0883	O-methyltransferase, family 2	Hydroxyneurosporene-O-methyltransferase	*Streptosporangium roseum* ATCC 12428	59 %
SNC	SRIM1008	Leucine-rich repeat (LRR) protein-like protein	Putative Teneurin-2	*Streptomyces aurantiacus* JA 4570	65 %
SNC	SRIM1495	L-carnitine dehydratase	Dehydratase	*Streptomyces roseosporus* NRRL 15998	81 %
SNC	SRIM2376	Putative cytochrome P450 hydroxylase	Cytochrome p450 superfamily protein	*Streptomyces auratus* AGR0001	77 %
SNC	SRIM2785	Sugar phosphate isomerases/epimerase	Xylose isomerase	*Streptomyces albulus* CCRC 11814	83 %
SNC	SRIM4074	Phenylacetate-CoA oxygenase, PaaG subunit	Phenylacetate-CoA oxygenase subunit PaaA	*Streptomyces auratus* AGR0001	90 %

SNC	SRIM4239	Dyp-type peroxidase family protein	Dyp-type peroxidase	*Streptomyces albulus* CCRC 11814	78 %
SNC	SRIM4676	Methyltransferase	Uncharacterised protein	*Streptomyces auratus* AGR0001	82 %
SNC	SRIM5022	Putative ClpX protein	Probable ATP-dependent Clp protease ATP-binding subunit	*Rhodococcus sp.* RHA1	75 %
SNC	SRIM5176	Putative D-aminoacylase	Amidohydrolase	*Streptomyces auratus* AGR0001	85 %
SNC	SRIM5636	Putative ATP/GTP-binding protein	ATP/GTP-binding protein	*Streptomyces auratus* AGR0001	86 %
SNC	SRIM5670	Cholesterol esterase	Uncharacterised protein	*Streptomyces auratus* AGR0001	80 %
SNC	SRIM5903	Allergen V5/Tpx-1 related	Uncharacterised protein	*Streptomyces auratus* AGR0001	59 %
SNC	SRIM6334	Putative monooxygenase	Putative monooxygenase	*Streptomyces avermitilis* ATCC 31267	86 %
SNC	SRIM6832	Putative Rif11 homolog	Uncharacterised protein	*Streptomyces albulus* CCRC 11814	73 %
SNC	SRIM7038	UbiA prenyltransferase	UbiA prenyltransferase	*Chloroflexus aurantiacus* ATCC 29364	44 %
SNC	SRIM7173	Major facilitator superfamily MFS_1	Major facilitator superfamily MFS_1	*Catenulispora acidiphila* DSM 44928	36 %
SNC	SRIM7373	Putative secreted protein	Putative secreted protein	*Streptomyces lividans* 1326	76 %
SNC	SRIM8611	Putative acetyltransferase	Putative ribosomal N-acetyltransferase YdaF	*Streptomyces aurantiacus* JA 4570	74 %
SNC	SRIM0760	Xanthine-uracil permease	Uracil permease	*Streptomyces albulus* CCRC 11814	89 %
SNC	SRIM0792	Xanthine dehydrogenase	Xanthine dehydrogenase	*Streptomyces auratus* AGR0001	92 %
SNC	SRIM1171	Dihydroorotate dehydrogenase	Dihydroorotate dehydrogenase (quinone)	*Streptomyces albulus* CCRC 11814	91 %
SNC	SRIM1172	Carbamoyl-phosphate synthase large chain	Carbamoyl-phosphate synthase large chain	*Streptomyces albulus* CCRC 11814	94 %
SNC	SRIM1173	Carbamoyl-phosphate synthase small chain	Carbamoyl-phosphate synthase small chain	*Streptomyces bottropensis* ATCC 25435	89 %
SNC	SRIM1205	Adenine phosphoribosyltransferase	Adenine phosphoribosyltransferase	*Streptomyces auratus* AGR0001	91 %
SNC	SRIM5115	Purine nucleoside phosphorylase	Purine nucleoside phosphorylase	*Streptomyces auratus* AGR0001	92 %

SNC	SRIM5359	NADH pyrophosphatase	Uncharacterised protein	*Streptomyces auratus* AGR0001	94 %
SNC	SRIM5944	Rossmann fold nucleotide-binding protein	DNA protecting protein DprA	*Streptomyces auratus* AGR0001	76 %
SNC	SRIM8371	Putative guanylate kinase	Putative guanylate kinase	*Streptomyces rochei*	78 %
SNC	SRIM1222	Putative membrane protein	Membrane protein	*Streptomyces albulus* CCRC 11814	77 %
SNC	SRIM3041	LpqB	Lipoprotein LpqB	*Streptomyces auratus* AGR0001	76 %
SNC	SRIM3050	Putative transmembrane protein	Uncharacterised protein	*Streptomyces auratus* AGR0001	84 %
SNC	SRIM4744	Putative intergral membrane protein	Uncharacterised protein	*Streptomyces albulus* CCRC 11814	86 %
SNC	SRIM4837	OrfL2	Uncharacterised protein	*Actinoplanes missouriensis* ATCC 14538	69 %
SNC	SRIM4848	Periplasmic binding protein	Uncharacterised protein	*Streptomyces albulus* CCRC 11814	69 %
SNC	SRIM5219	Putative intergral membrane protein	Integral membrane protein	*Streptomyces auratus* AGR0001	64 %
SNC	SRIM5260	Putative lipoprotein	Uncharacterised protein	*Streptomyces albulus* CCRC 11814	57 %
SNC	SRIM6000	Membrane-associated zinc metalloprotease	Peptidase M50	*Streptomyces auratus* AGR0001	91 %
SNC	SRIM7311	Multimodular transpeptidase-transglycosylase	Putative penicillin-binding protein	*Streptomyces mobaraensis* NBRC 13819	77 %
SNC	SRIM8045	Apolipoprotein N-acyltransferase	Apolipoprotein N-acyltransferase	*Streptomyces auratus* AGR0001	81 %
SNC	SRIM0235	Streptomycin 6-kinase	Hydroxyurea phosphotransferase	*Saccharopolyspora erythraea* NRRL 23338	62 %
SNC	SRIM0783	Enhanced intracellular survival protein	Uncharacterised protein	*Streptomyces auratus* AGR0001	65 %
SNC	SRIM1253	Cobalt-precorrin-6y C5-methyltransferase	Precorrin-6Y C5,15-methyltransferase	*Streptomyces auratus* AGR0001	88 %
SNC	SRIM1265	Multidrug resistance efflux pump	MFS transporter	*Streptomyces albulus* CCRC 11814	83 %
SNC	SRIM2665	Enhanced intracellular survival protein	Uncharacterised protein	*Streptomyces albulus* CCRC 11814	70 %
SNC	SRIM8602	Na+-driven multidrug efflux pump	Putative efflux protein	*Streptomyces avermitilis* ATCC 31267	75 %

SNC	SRIM1210	Holliday junction DNA helicase RuvA	Holliday junction ATP-dependent DNA helicase RuvA	*Streptomyces auratus* AGR0001	91 %
SNC	SRIM1783	Excinuclease ABC subunit A	UvrABC system protein A	*Streptomyces albulus* CCRC 11814	91 %
SNC	SRIM1792	Excinuclease ABC subunit B	UvrABC system protein B	*Streptomyces auratus* AGR0001	91 %
SNC	SRIM1822	DNA polymerase I	DNA polymerase I	*Streptomyces auratus* AGR0001	88 %
SNC	SRIM3879	Type I restriction-modification system, restriction subunit R	Type I site-specific deoxyribonuclease, HsdR family	*Streptomyces sp*. SirexAA-E	45 %
SNC	SRIM4114	DNA gyrase subunit A	DNA gyrase subunit A	*Streptomyces auratus* AGR0001	95 %
SNC	SRIM4725	Excinuclease ABC subunit A, paralogue	Putative UvrABC system protein A	*Streptomyces afghaniensis* 772	87 %
SNC	SRIM6070	LigA	Putative Uncharacterised protein	*Aureococcus anophagefferens*	37 %
SNC	SRIM6151	Topoisomerase IV subunit A	DNA topoisomerase IV subunit A	*Streptomyces auratus* AGR0001	91 %
SNC	SRIM6638	Alkylated DNA repair protein AlkB	2OG-Fe(II) oxygenase	*Streptomyces auratus* AGR0001	79 %
SNC	SRIM7374	Putative DNA methylase	Putative DNA methylase	*Streptomyces scabies* 87.22	93 %

4.7 Discussion

By analysing the genetic differences between the original soil isolate, *S. rimosus* SR1, and the strain improvement lineage, SR20, SR50S and SR55, an informed picture for the process of strain improvement can be developed, rather than the black box of unknown mechanisms. A total of 484 genes were deleted from the *S. rimosus* chromosome by the time SR55 was isolated, the majority of which came from only two regions over the strain improvement period. *Streptomyces* genomes are known for having a central core region comprised of essential genes and variable regions that contain non-essential genes, such as secondary metabolite clusters, at the chromosomal arms (Friend & Hopwood 1971; Bentley *et al.,* 2002). It was within the variable region of the *S. rimosus* chromosome in which the deletions took place (Figure 4.10), and this is understandable as it is likely that any deletions within the core region are fatal. Another interesting feature of the deletions was the presence of at least one gene associated with genetic mobility i.e transposases, insertion sequences. A transposase of the IS4 family was at a central position within the first major deletion of SR20 (277 kbp region, Figure 4.3), and these types are known for generating and short flanking DNA repeats on insertion, ending with imperfect inverted repeats (Siguier *et al.,* 2014). Within the 69/142 kbp region (Figure 4.8) of SR50 there were IS21 and IS481 type transposases. The transpoases flanking the lantipeptide biosynthetic cluster (Figure 4.9) in SR55 are IS110 on the left hand side and IS4 on the right, with IS1647 transpoases flanking the cluster. Apart from IS110, all the afore mentioned IS type transposases are members of the to IS4 family of transpoases (Siguier *et al.,* 2014). It makes sense that a high proportion of genetic mobility genes would be present within the deletions, given that the chromosomal arms are variable regions in general and the presence of these elements could explain the labile nature of these chromosomal regions. Having genetic plasticity is beneficial to a bacterium as it allows for the integration of advantageous

genes, such as secondary metabolite clusters, but also, as in this case, the removal of extraneous DNA.

4.7.1 Genetic differences between SR1 and SR20

It is likely the most significant contribution to increased OTC productivity in SR20 is due to the loss of PKS gene clusters, rimocidin and mnp. It is still interesting to note the mutations occurred in genes associated with starvation and nutrient depletion, *SRIM1204* and *SRIM7514* (Saito 2002; Horinouchi *et al.,* 2001; Gropp *et al.,* 2001; Saito *et al.,* 2006; Gomez-Escribano *et al.,* 2008). RelA is a (p)ppGpp synthetase and is required for the production of (p)ppGpp, which has been shown to play a role in antibiotic production (Hesketh *et al.,* 2007; Sun *et al.,* 2001). As mentioned previously deletion of eshA results in lower levels of ppGpp and this can be countered by the introduction of a truncated form of RelA (Saito *et al.,* 2006). It is interesting that mutations occur in the (p)ppGpp synthetase and a CDS whose product regulates the levels of (p)ppGpp. It is possible that functional redundancy of SRIM1204 and SRIM7514 negates the effects of polymorphisms in these CDSs and therefore either prevent any negative impact of the mutation or provide a positive influence in regards to OTC production.

The loss of genes associated with arsenic resistance is interesting as arsenate is an analogue of phosphate (Elias *et al.,* 2012; Sánchez-Riego *et al.,* 2014), which is known to repress antibiotic production (Chapter 1.4.5). Arsenic is present in two biological forms as arsenate (AsO_4^{3-}) and arsenite (AsO_3^{3-}), with arsenate able to enter cells using phosphate transporters and arsenite using glycerol uptake facilitator protein (Glfp) (Meng *et al.,* 2004). The arsenic resistance genes deleted in SR20 were *SRIM7589*, *SRIM7590* and SRIM7591, with the former two encoding homologues of ACR3 like proteins and the latter encoding a regulator of the operon. ACR3 like proteins are members of the BART (bile/arsenite/riboflavin transporter) family and confer arsenite resistance (Fu *et al.,* 2009). However 6 genes are still

present within SR20 that are associated with arsenic resistance, 3 specifically linked to arsenate resistance based on bioinformatic analysis. Isolation of *Streptomyces* mutants resistant to arsenate were found to be relieved from phosphate repression of antibiotic production (Hanel *et al.,* 1989). However there seems to be no correlation between natural arsenate or arsenite resistance and loss of repression, suggesting the arsenate mutants occur by a different means. It is possible that mutations in phosphate transport pathways lead to arsenate resistance, which could also be related to loss of antibiotic repression due to the presence of phosphate. Further to this all the tested *Streptomyces* that were sensitive to arsenate were also sensitive to phosphate control of antibiotic production (Hanel *et al.,* 1989). As arsenite is not an analogue of phosphate and enters cells via a different mechanism, it is unlikely that loss of the arsenite resistance genes leads to loss of phosphate repression of antibiotic production. However it is possible that phosphate control of antibiotic production is modulated due to the loss of these genes. In order to determine this OTC production by SR20 would need to be measured against increasing phosphate concentrations and compared to SR1 OTC production under the same conditions.

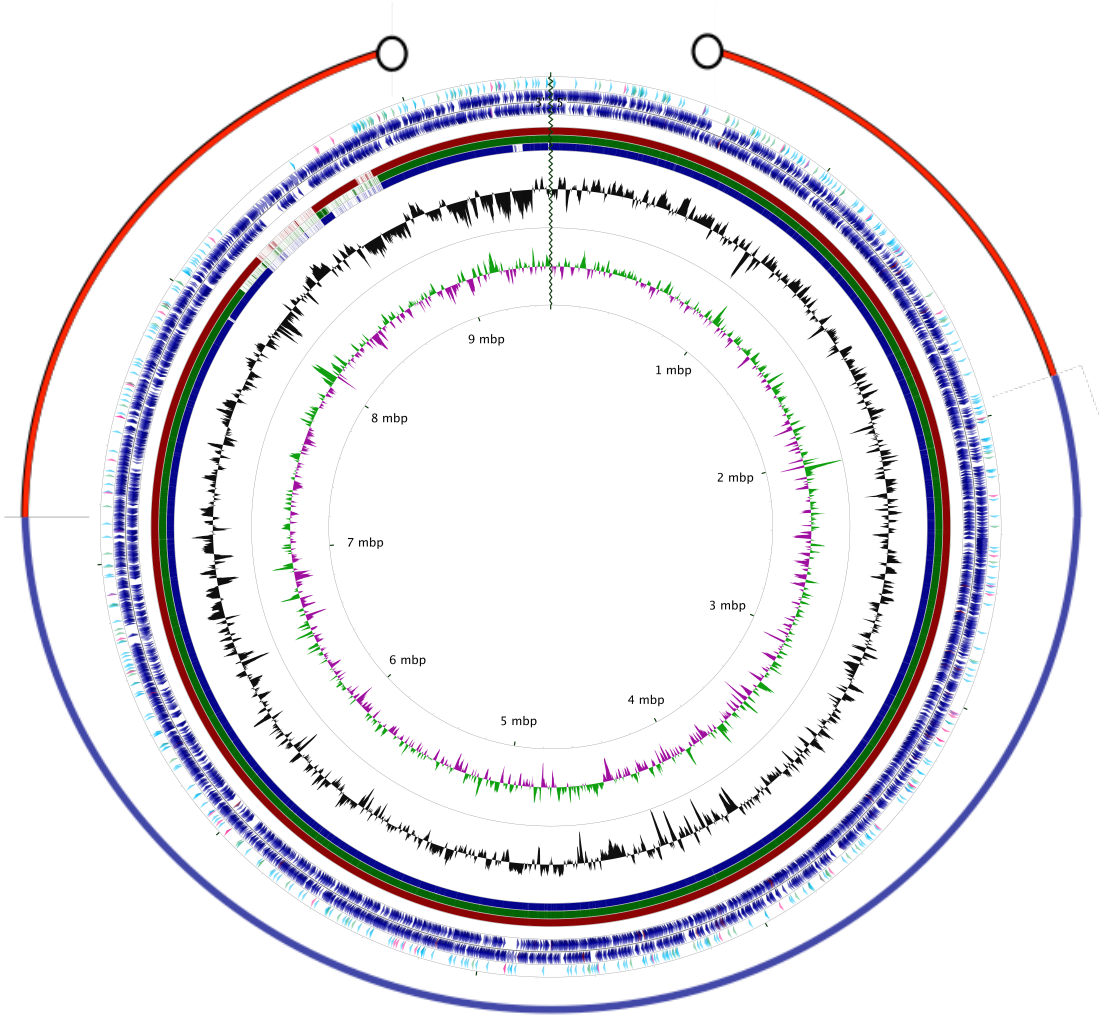


Figure 4.10 *S. rimosus* SR1 chromosome as a circular representation with SR20, SR50S and SR55 sequence mapped. The inner circle represents chromosome position mega base pairs. The outer ring represents the core region in blue and the variable region in red. The terminal proteins are highlighted by black circles at the end of each chromosomal arm.

4.7.2 Genetic differences between SR1 and SR50/50S

The OTC productivity difference between SR20 and SR50 was significant with an increase of between 20 and 30 g/L, however unlike SR20 no single mutation was predicted to be the significant contributor. A greater proportion of the deleted CDSs belong to a putative NRPS biosynthetic cluster, however the presence of an orphaned sensor kinase reminiscent of CutS was intriguing. Around 60 % of sensor kinases (SK) and response regulators (RR) of *S. coelicolor* A3(2) lie adjacent to one another and are expected to produce two component regulatory systems, this results in around 40 % orphaned SK and RR (Yepes *et al.,* 2011). Similarly for *S. rimosus* SR1 there are 43 % orphaned SK and RR, as only 84 SK and RR genes lie adjacent. Studies have shown that orphaned SK or RR can interact with unrelated RR or SK respectively (Hutchings 2004; Wei *et al.,* 2007; Wang *et al.,* 2009c). As *SRIM7808* was orphaned from a RR it was proposed that it may encode a SK that regulates a noncognate RR. Wang *et al.,* demonstrated that an orphaned RR was under the control of SK already paired with a RR and that deletion of both the orphaned RR and the paired RR resulted in the same phenotype (Wang *et al.,* 2009c). Due to the level of similarity between SRIM7808 and SRIM6174 it seems likely that SRIM7808 can phosphorylate SRIM6173, providing an extra level of regulation over SRIM6173 and its regulon.

The ECF (extracytoplasmic function) subfamily of RNA polymerase sigma factors are associated with the regulation of membrane linked proteins, and because of this are usually activated by external stimuli (Brooks & Buchanan 2008). This is the case for SigQ in *S. coelicolor* A3(2) in which it exerts control over the two component system AfsQ1-Q2 under specific nitrogen conditions (Shu *et al.,* 2008). The identification of a *sigQ* homologue, *SRIM5103*, close to *SRIM5101* and *SRIM5102*, homologues of *afsQ1* and *afsQ2* respectively, suggested that a similar mechanism of control may be active for those genes. A frameshift mutation was present in *SRIM5103* which is likely to result in a *sigQ* deletion phenotype for *S.*

rimosus SR50S when grown on minimal medium supplemented with glutamate as the sole nitrogen source. It is possible that this mutation may also have resulted in a similar phenotype when grown on the media used at Pfizer, which seems likely to have resulted in increased OTC production.

4.7.3 Genetic differences between SR1 and SR55

The final strain of the strain improvement lineage available was SR55, and it produced an extra 5 g/L of OTC compared to SR50. A total of 55 genes were attributed to the SR55 deletion profile and the majority of the genes were part of a predicted lantipeptide cluster. The deletion profile of SR55 was not significant when compared to the number of SNCs. There was a 6-fold and 3-fold increase in the number of non-synonymous SNCs associated with SR55 compared to SR20 and SR50S respectively. For this reason it is likely that the most significant genetic difference that lead to increased OTC production for SR55 is due to non-synonymous SNCs.

As SR55 was known to be 2-DOG resistant the task of establishing if the mechanism behind its resistance was the same as those previously published was undertaken. So far three proteins have been shown to play a role in CCR, Glk, SCO2127 and the glucose transporter GlcP1 (Kwakman & Postma 1994; Guzman 2005; van Wezel *et al.,* 2004; Swiatek *et al.,* 2013). The homologues of *glk*, *SCO2127* and *glcP1* were identified in the *S. rimosus* chromosome to pinpoint any mutations that might have occurred that could explain 2-DOG resistance in SR55, however no mutations were present. SCO6008, a repressors, open reading frames, and kinases (ROK) family regulator of *S. coelicolor* A3(2), is known to influence CCR, growth and antibiotic production (Swiatek *et al.,* 2013). One gene affected by SCO6008 (Rok7B7), was *glcP1* which was upregulated in a *rok7B7* deletion mutant. The homologue of *rok7B7* in SR55 was *SRIM6404* and once again no SNC/DIPs were present in the gene. If 2-DOG resistance in SR55 is controlled by SRIM6404 it must be due to an as yet unknown regulation of the *SRIM6404*.

An alternative explanation for 2-DOG resistance in SR55 was offered via the SCO2127 pathway. It has been well established the SCO2127 plays a role in CCR and 2-DOG resistance (Guzman 2005; Guzmán *et al.,* 2005; Chávez *et al.,* 2010). Protein binding studies demonstrated that SCO2127 binds to two proteins, BldKB (SCO5113) and SCO5282, and it is possible that through this interaction CCR is activated (Chávez *et al.,* 2010). Preventing this interaction would potentially prevent CCR and could be an alternative route for 2-DOG resistance. As previously mentioned this was the alternative mechanism proposed for SR55 2-DOG resistance. As it was shown that BldKB was ADP-ribosylated, then perhaps SRIM5280 can also be ADP-ribosylated and it is due to this modification that SRIM5280 is present for regulation by SRIM1978. Supporting this theory are polymorphisms in two genes that encode adenine phosphoribosyltransferases. Modification of the ADP-ribosylation state of SRIM5280 may be a preferred method for producing a 2-DOG resistance strain as it may avoid any detrimental affects associated with a *glk, glcP* or *SCO2127*.

Taken together these observations serve as a starting point in developing a high producing antibiotic strain, the next step is to ascertain which of the above mutations and/or deletions contribute to greater antibiotic productivity. The next chapter will aim to address the question of the influence other secondary metabolite clusters have on OTC production and if disruption can increase OTC productivity.

Chapter 5

5 Characterization of the rimocidin biosynthetic gene cluster from *S. rimosus*

5.1 Introduction

S. rimosus SR1 (G7) produces the antifungal rimocidin, as well as OTC (Phornphisutthimas *et al.,* 2010). Rimocidin is a Type-1 polyketide antifungal agent that was first documented in 1951 (Davisson *et al.,* 1951) and is active against *Saccharomyces cerevisiae*, as well as a range of other pathogenic and non-pathogenic fungi (Ball *et al.,* 1957; Molzahn & Woods 1972). Although rimocidin is active against a broad range of fungi its toxicity means that it is not used clinically and, as a result, it is not as well studied as other macrolide antifungal agents; although, some work has been undertaken to characterize the biosynthetic gene cluster from *Streptomyces diastaticus* (Seco *et al.,* 2004; Kong *et al.,* 2013; Heia *et al.,* 2011). Sequencing the genome of *S. rimosus* SR1 provided the opportunity to elucidate the biosynthetic cluster.

The synthesis of rimocidin is achieved by a modular polyketide synthase (PKS) containing a loading module able to accept starter units of either acetate or butyrate (Heia *et al.,* 2011), and 13 elongation modules. The elongation modules can either incorporate acetate, propionate, and butyrate as extender units depending on the acyltransferase domain present (Seco *et al.,* 2004). OTC biosynthesis uses acetate, in the form of malonyl-CoA, to form the extender units (Pickens & Tang 2010). Acetyl-CoA for use in the production of polyketides is supplied from the catabolism of glucose via glycolysis. Acetyl-CoA can then be converted to malonyl-CoA by acetyl-CoA carboxylase (Ryu *et al.,* 2006). Alternatively acetyl-CoA is directed into tricarboxylic acid cycle (TCA) as a precursor (Hallam *et al.,* 2006; Akopiants *et al.,* 2005). This presents a limitation in precursor supply as well as

providing only a single source for these resources that can be used in the production of polyketides (Coze *et al.,* 2013). Due to the absence of the rimocidin as a secondary metabolite extract in the later production strains, it was hypothesised that the use of acetyl-CoA in the biosynthesis of both could lead to resource competition. Therefore the absence of rimocidin production could allow for carbon to be directed solely to OTC biosynthesis, as seen with nystatin production in *S. ahygroscopicus*, in which disruption of a resource competing PKS pathway lead to increased production of nystatin (Ren *et al.,* 2013).

In order to test this hypothesis it was first necessary to generate a rimocidin knockout strain of *S. rimosus* SR1. Deletion of the complete 95.7 bp biosynthetic cluster was not a practical option and a different method was employed. It is common for gene knockouts to be produced in *Streptomyces* species using homologous recombination to produce a double crossover, as single crossover events may lead to polar effects on the expression of adjacent genes (Otten *et al.,* 2000; Kawamoto *et al.,* 2001). This method involves amplifying regions of DNA flanking the gene of interest and then introducing these PCR products into a suicide plasmid either side of a selectable marker. This plasmid is then introduced into the strain selected for mutation and gene knockout strains are isolated using the selectable marker (Figure 5.1). Identification of a double crossover event is determined by the mutant strain being resistant to one of the selectable markers on the original suicide plasmid, but sensitive to the other; if the mutant strain is resistant to both then a single crossover event has occurred (Figure 5.1). This method was used to generate rimocidin knockouts by disrupting specific areas of the biosynthetic cluster.

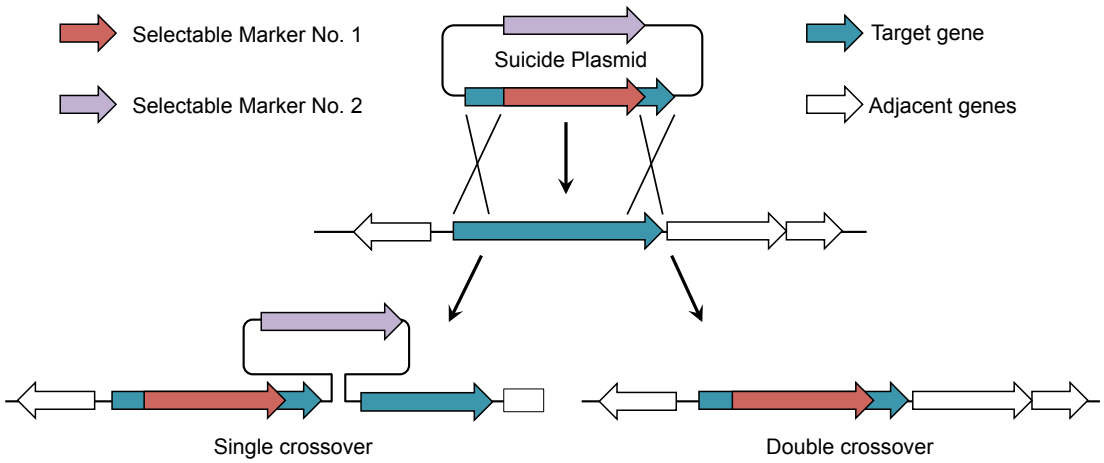
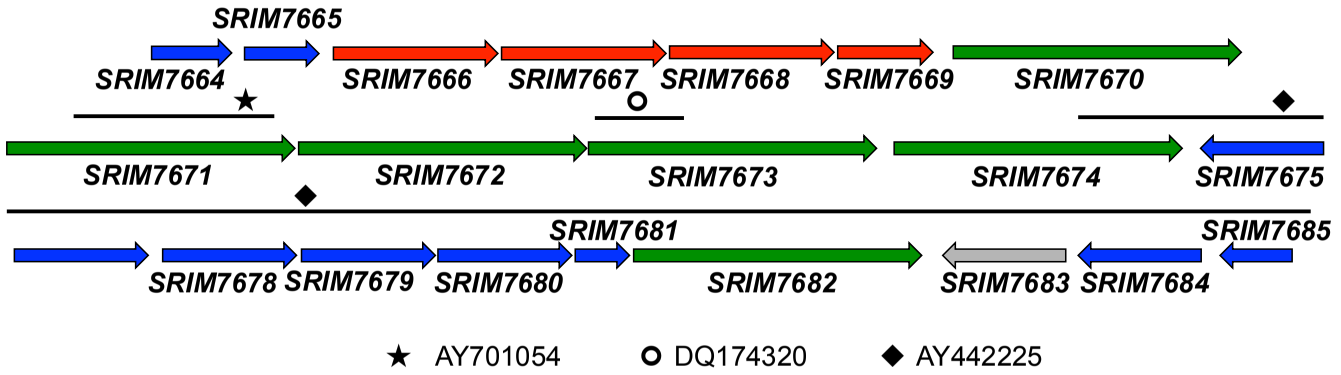


Figure 5.1 Schematic for the generation of double and single crossovers. After cloning the flanking regions of the target gene into a suicide plasmid with two selectable markers, the plasmid is introduced into the strain. Homologous recombination can result in either a single crossover, resulting in incorporation of the all of the suicide plasmid, or double crossover, which incorporates only selectable marker No. 1, disrupting the target gene. If a single crossover event occurs then colonies of the mutant strain will arise that are resistant to selectable marker No. 1 and No. 2. However in the event of a double crossover mutant strains will only be resistant to selectable marker No. 1, and sensitive to selectable marker No. 2, providing a screening method for determining the type of crossover event that has occurred.

5.2 Elucidation of the rimocidin biosynthetic cluster

The initial steps in the elucidation of the *rim* cluster required identification of the contigs in which the cluster was present. This process was made easier by the presence of three sequences associated with the *rim* cluster available in Genbank, two from *S. rimosus* (Accession numbers DQ174320 and AY701054) and one from *S. diastaticus* (Accession number AY442225) (Figure 5.2 A). The *S. diastaticus* sequence was 19 kbp in size and was a contiguous region of the cluster containing the genes *rimA*-*rimH*, as well as ORF1, *rimJ* and *rimK* (Figure 5.2, A and B). This sequence was used to find the first contig containing the *rim* cluster, and unfortunately the complete cluster was not present in this contig, therefore it was necessary to use the other two sequences to identify other potential contigs. This was achieved by nBLAST alignments and four additional contigs were identified. Once all the contigs containing the cluster were identified the next step was to order them correctly; this was achieved by analysing the genes encoded by the contiguous sequence. The corresponding protein was examined and the module/s documented through identification of the enzymatic domains (Figure 5.2) present within each protein. In order to correctly categorise each domain, it was necessary to identify catalytic amino acids and motifs (Keatinge-Clay 2008; Kwan *et al.,* 2008; Haydock *et al.,* 1995; Bisang *et al.,* 1999; Banskota *et al.,* 2006; Bonnett *et al.,* 2013). These enzymatic domains were ketoreductases (KR), dehydratases (DH), acyl-carrier proteins (ACP), ketosynthases (KS), enoylreductases (ER), thioesterases (TH), acyltransferases (AT) and carboxylic acid-CoA ligases (CoL). Using the predicted biosynthetic pathway for rimocidin production (Seco *et al.,* 2004; Petković *et al.,* 2006), in combination with the newly identified enzymatic domains, the contigs were reordered and the *rim* biosynthetic cluster was formed (Figure 5.2).

A



B

Gene	Protein	Function
SRIM7664	RimM	Sugar Dehydratase
SRIM7665	RimL	Thioesterase
SRIM7666	RimRI	Transcriptional Regulator
SRIM7667	RimRII	Transcriptional Regulator
SRIM7668	RimRIII	Transcriptional Regulator
SRIM7669	RimRIV	Transcriptional Regulator
SRIM7670	RimS2	Module 5 – AT of 7
SRIM7671	RimS3	Module 7 KR – 10
SRIM7672	RimS4	Module 11
SRIM7673	RimS5	Module 12 – 13
SRIM7674	RimB (RimS1)	Module 1 – 4

Gene	Protein	Function
SRIM7675	RimC	Tyrosine Phosphatase
SRIM7677	RimD	Oxidase
SRIM7678	RimE	Glycosyl Transferase
SRIM7679	RimF	Aminotransferase
SRIM7680	RimG	cP450 Monooxygenase
SRIM7681	RimH	Ferredoxin
SRIM7682	RimA (RimS0)	Loading Module
SRIM7683	ORF1	Transposase
SRIM7684	RimJ	Crotonyl-CoA Reductase
SRIM7685	RimK	Acetyltransferase

Figure 5.2 The Rimocidin biosynthetic cluster. **A.** Putative arrangement of the *rim* cluster. Previously discovered sequences on Genbank are highlighted and labelled with each accession number. **B.** Proteins and polypeptides encoded, with putative function. Red indicates regulators, green for polyketide synthase and blue for modifying/accessory proteins.

Once the *rim* biosynthetic cluster was elucidated the predicted model for rimocidin production could be reinforced and refined (Figure 5.3). Previous predictions based on the final rimocidin chemical structure indicated that the AT of both module 7 and 13 would have different substrate specificity from that of the ATs associated with the other modules (Seco *et al.,* 2004). The amino acid motif of ATs with acetate specificity have been well characterised (Haydock *et al.,* 1995; Yadav *et al.,* 2003; Liew *et al.,* 2012) and a marked difference was seen upon inspection of the amino acid motifs of the AT domains within module 7 and 13. Further analysis of these motifs further revealed homology to ATs that incorporate propionate (Haydock *et al.,* 1995), in the case of module 7, or butyrate for module 13, coinciding with the prediction made by Seco *et al.,* 2004. However, upon examination of module 12, DH, ER and KR domains were identified (Figure 5.3), contrary to earlier predictions (Seco *et al.,* 2004). The presence of these domains also disagree with the chemical structure of rimocidin, as these enzymatic groups would function together to remove a carbonyl group from the structure, which will be explained in more detail below (Figure 5.4 B). It is not uncommon for a PKS to contain enzymatic domains within modules that are inactive due to substitution of amino acids that are essential for enzyme activity (Rawlings 2001a; Rawlings 2001b; Aparicio *et al.,* 2000). As a result, the domains within module 12 were examined to determine if the KR, DH and ER domains were active.

KRs contain both structural and catalytic subdomains (Anand & Mohanty 2012) and it is well documented that the active site residues are within the catalytic subdomain, and consist of a lysine, serine and tyrosine (Keatinge-Clay 2007; Bonnett *et al.,* 2013; Zhou *et al.,* 2012a; Zheng *et al.,* 2012). KR's also contain a NADP+ binding site motif GXGXXGXXXA (Figure 5.4 A) (Chen *et al.,* 2003; Bonnett *et al.,* 2013). In addition, PKS modules that contain both a KR and an ER, have the structural and catalytic subdomains of the KR separated by the ER domain (Keatinge-Clay & Stroud 2006; Anand *et al.,* 2010), which is observed in both module 11 and 12.

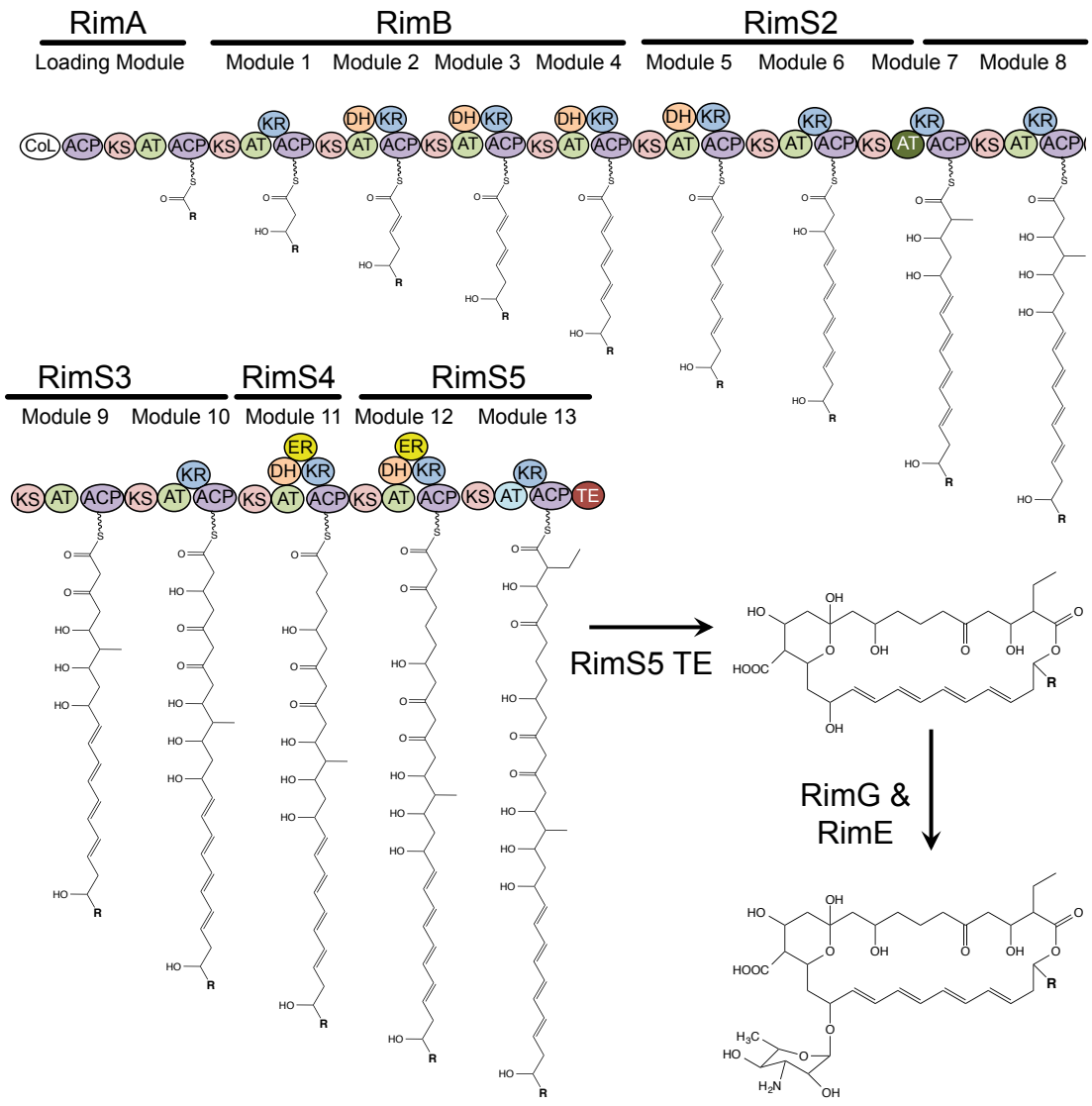


Figure 5.3 Model of rimocidin biosynthesis. The proteins and their associated modules are labelled. Enzymatic domains present in each module are marked with the following CoL = carboxylic acid-CoA ligase, ACP = acyl carrier protein, KS = ketosynthase, AT = acyl transferase, KR = ketoreductase, DH = dehydratase, ER = enoyl reductase, TE = thioesterase. The ATs of module 7 and 13 incorporate propionate and butyrate respectively, rather than acetate. R = CH_2–CH_2–CH_3

In addition to this the active site amino acids within the catalytic subdomain are present in module 12 (Figure 5.4 A), and the NADP+ binding site motif was also identified. This lead to the prediction that the KR observed is genuine and potentially active.

DHs have two important conserved domains HXXXGXXXP and DXXXQ/H (Figure 5.4 A) that contain the active site residues of a histidine and aspartic acid, respectively (Gay *et al.,* 2013; Guo *et al.,* 2010; Akey *et al.,* 2010); mutation of the histidine residue can abolish DH activity (Valenzano *et al.,* 2010). Module 11 and 12 both contain these motifs and the associated amino acids (Figure 5.4 A), suggesting the DH within module 12 is functional. Lastly the ER of module 12 was examined to identify amino acids residues indicative of an active ER domain. The definitive amino acids responsible for the catalytic activity of an ER domain have yet to be identified (Zheng *et al.,* 2012), though some residues have been selected as playing a significant catalytic role. Despite this, mutation of these residues does not eliminate ER activity (Kwan & Leadlay 2010; Ames *et al.,* 2012). Some of these residues can also be used to determine stereospecificity or to identify NADPH binding sites (Kwan *et al.,* 2008; Kwan & Leadlay 2010), and within module 12, as with module 11, these residues were identified (Figure 5.4 A). The NADPH binding motif HAXAGGVGMA is present in module 12, along with residues associated with stereospecificity indicating an active ER domain which is predicted to confer an S configuration (Figure 5.4 B) of the adjacent methyl group due to the presence of a conserved tyrosine (Kwan *et al.,* 2008). Due to these observations it is predicted that the KR, DH and ER domains of module 12 are present and active.

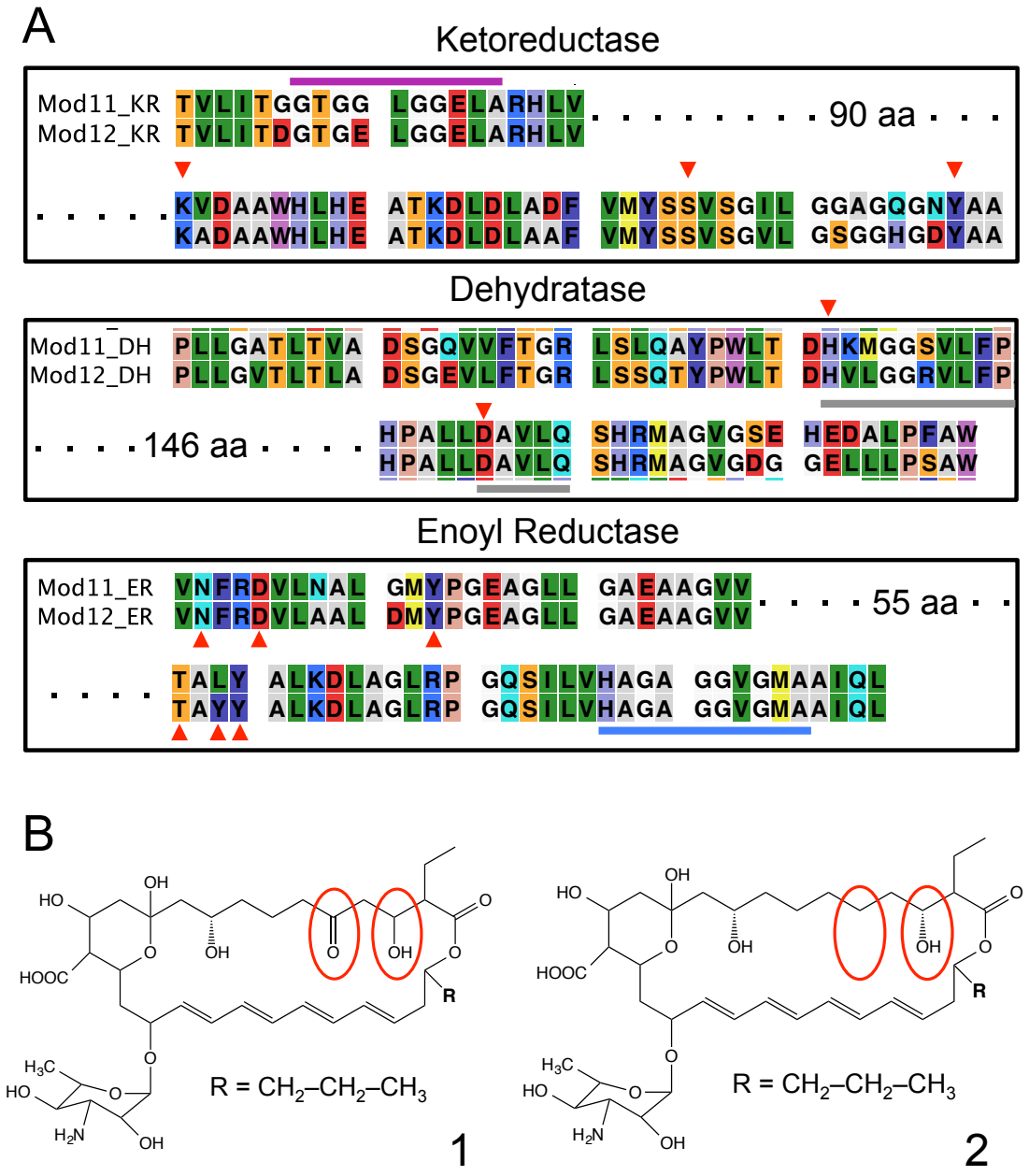


Figure 5.4 Catalytic domains of Module 12. **A.** Alignments of the KR, DH and ER from Module 12 with Module 11. Catalytic amino acids are marked with red triangles. The NADP+ binding motif of KR (GXGXXGXXXA) is indicated by the purple bar. Motifs of DH (HXXXGXXXXP and DXXXQ/H) are indicated by the grey bar. The NADPH binding site (HAXAGGVGMA) of the ER is highlighted by the blue bar. **B.** The red circles indicate the carbon groups that are changed by active KR, ER and DH within Module 12.

In addition to the extra domains detected within module 12, four transcriptional regulators, *SRIM7666, SRIM7667, SRIM7668* and *SRIM7669*, were discovered upstream of *SRIM7670* and it is proposed that the proteins encoded by these genes are involved in the regulation of the *rim* cluster. The product of *SRIM7666*, RimRI, shares 53% identity with NysRI, a proposed regulator of the nystatin biosynthetic cluster (Sekurova *et al.,* 2004). In addition, RimRII, the product of *SRIM7667*, shares 47% identity with NysRII, another proposed regulator of nystatin biosynthesis. Both RimRI and RimRII contain LuxR-type helix turn helix (HTH) domains at the carboxy-terminus (C-terminal) and a putative amino-terminal (N-terminal) ATP-binding domain (Figure 5.5). These features place RimRI and RimRII into the LAL (LuxR) family of transcriptional regulators (Kuščer *et al.,* 2007; Wilson *et al.,* 2001). The product of SRIM7668, RimRIII, has 72% identity with TtmRIII, a putative regulator of tetramycin biosynthesis (Ren *et al.,* 2013) and contains a putative N-terminal ATP-binding domain and a LuxR-type HTH domain at the protein's C-terminal (Figure 5.5); placing RimRIII into the LAL (LuxR) family of transcriptional regulator. RimRIV, the product of SRIM7669, shares 78% identity with TtmRVI, another potential regulator of tetramycin biosynthesis. Bioinformatic analysis of RimRIV highlighted the presence of a N-terminal PAS domain (Ponting & Aravind 1997; Hefti *et al.,* 2004) and a C-terminal LuxR-type HTH domain (Figure 5.5) (Zhang *et al.,* 2002; Ducros *et al.,* 2001). It was due to the higher homology to other type I PKS regulators and the identification of N-terminal and C-terminal domains that RimRIII and RimIV were selected for further analysis.

Inspection of the N-terminal region of the RimIII allows for the identification of putative Walker A and Walker B motifs (Walker *et al.,* 1982) (Figure 5.6 A and B) reminiscent of other regulators of polyene macrolide antibiotic biosynthesis (Wilson *et al.,* 2001; Sekurova *et al.,* 2004; Knirschová *et al.,* 2007). Amino acid alignments show conservation between the Walker A and Walker B motifs between transcriptional regulators of Type I PKS (Figure 5.6).

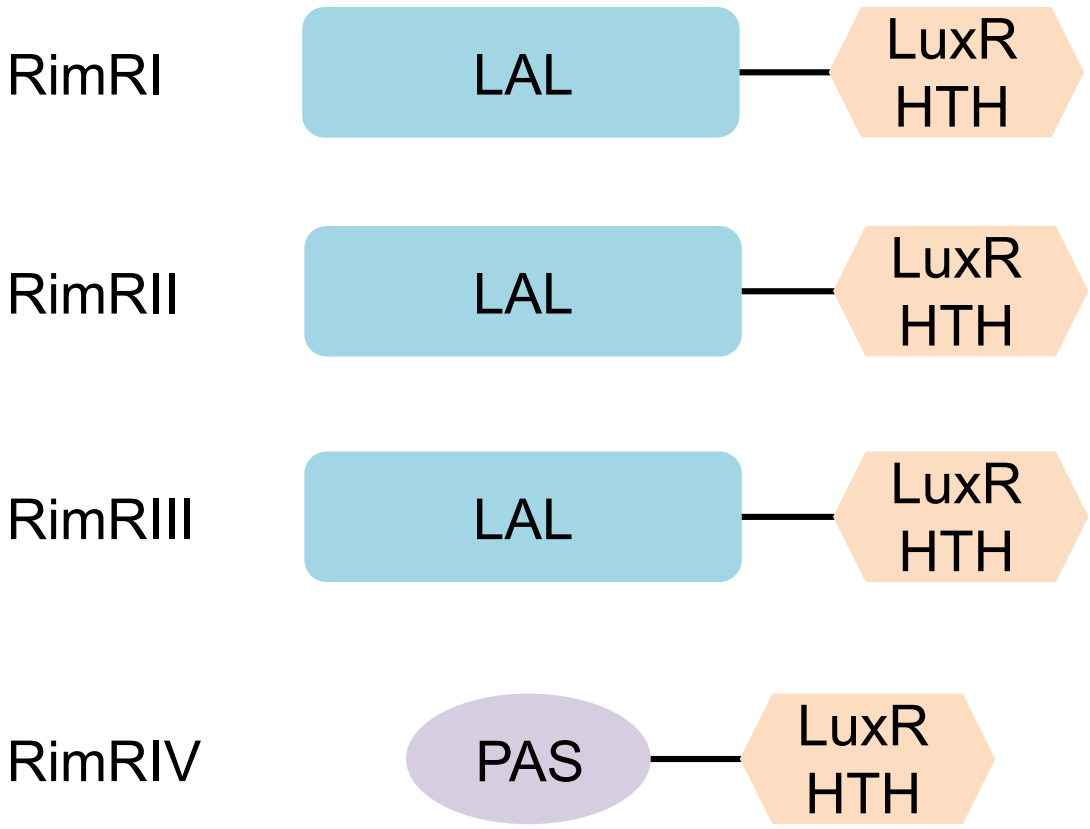


Figure 5.5 Domain architecture of RimRI, RimRII, RimRIII and RimRIV. SMART and Pfam (Letunic *et al.,* 2011; Punta *et al.,* 2011) analysis was used to determine the domain architecture of each protein. Pfam analysis identified an N-terminal AAA (ATPases Associated with diverse cellular Activities) domain with Pfam accession number PF13191, for RimRI, RimRII and RimRIII. An N-terminal PAS domain was identified by SMART analysis with accession number SM000091. SMART and Pfam analysis identified a HTH domain of the LuxR family, with SMART accession number SM000421 and Pfam accession number PF00196. All Pfam and SMART domain analysis results had E-values less than 2.00E-07, with the exception of RimRIV PAS domain result, which had an E-value of 1290.

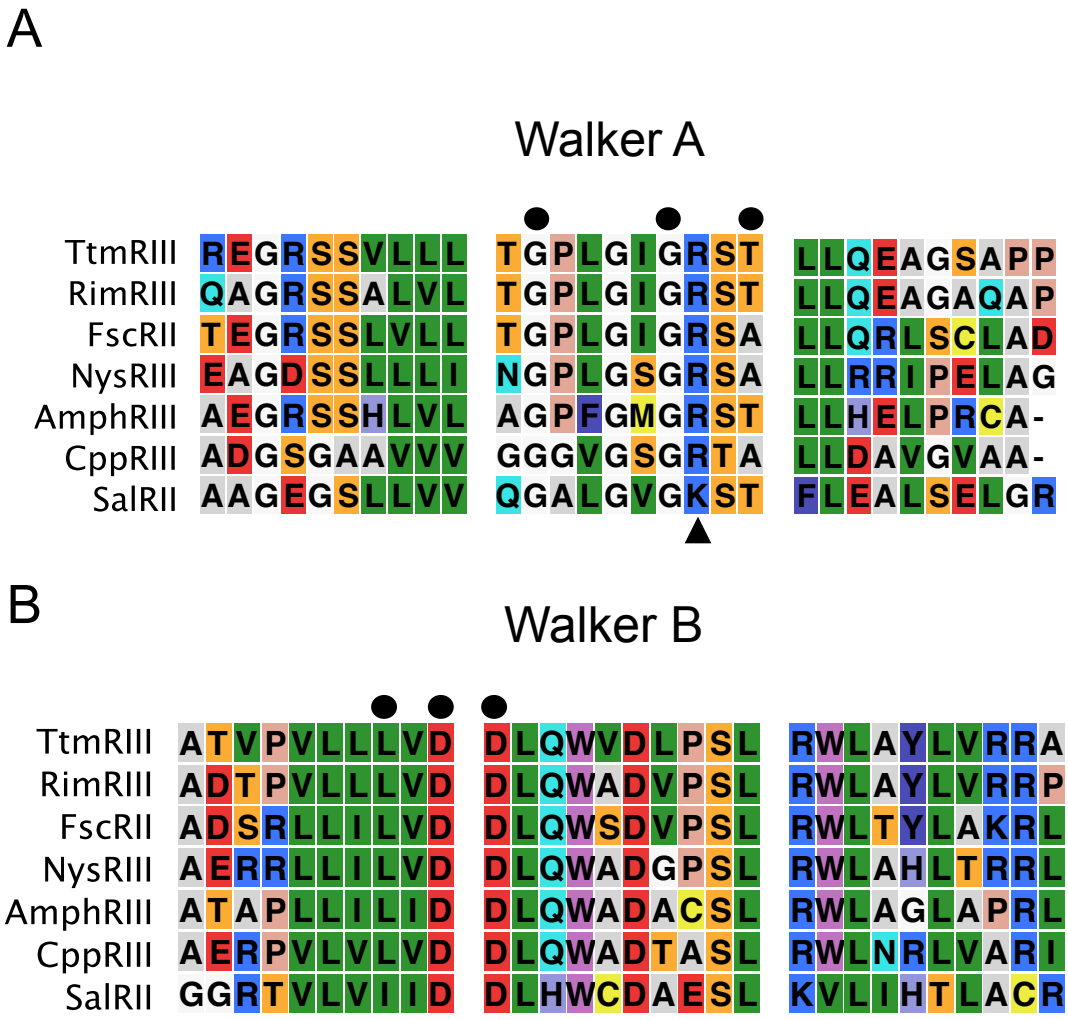


Figure 5.6 Alignment of RimRIII with other regulators of polyene biosynthesis. AmphRIII, amphotericin *S. nodosus;* FscRII, candicidin *S. sp* FR-008; NysRIII *S. noursei*, nystatin; TtmRIII, tetramycin, *S. ahygrocopicus;* CppRIII, NPP, *Pseudonocardia autotrophica;* SalRII, Salinomycin, *S. albus*. **A.** The Walker A motif. The classic Walker A motif is GXXXXGK, however the putative Walker A motif in the shown polyene regulators have the conserved lysine substituted with arginine (black triangle), with the exception of SalRII. **B.** The Walker B motif. Black circles denote amino acid residues associated with NTP-binding.

The PAS and LuxR-type HTH domains of RimRIV also feature in PimM (Figure 5.7), one of two positive regulators of pimaricin biosynthesis (Anton *et al.,* 2007; Santos-Aberturas *et al.,* 2012). Other studies have shown that this is a common feature of polyene macrolide antibiotic regulators (Carmody *et al.,* 2004; Sekurova *et al.,* 2004; Kong *et al.,* 2013; Santos-Aberturas, Payero *et al.,* 2011a). Amino acid sequence alignments between RimRIV, PimM and other known polyene macrolide regulators were undertaken. As observed previously there was homology between the regulators (Figure 5.6 and 5.7), reinforcing the hypothesis that *SRIM7668* and *SRIM7669* are regulatory genes of the *rim* biosynthetic cluster. Clearly, it is not possible to exclude the possibility that other regulatory genes exist that lie outside the cluster.

Two more genes were identified upstream of the four putative regulators, *SRIM7665* and *SRIM7664* with the products, RimL and RimM, sharing 91% identity with PimI and 96% identity with PimJ, respectively. Although the role that PimI and PimJ play in pimaricin synthesis is not known, the promotors, and those of homologues from *S. chattanoogensis*, have been mapped (Du *et al.,* 2011; Santos-Aberturas, Vicente, *et al.,* 2011b). The results for *pimJ* and *pimI* differed from those of *scnJ* and *scnI*, as these genes are co-transcribed rather than being transcribed separately. Although this does not shed light on the means by which *SRIM7664* and *SRIM7665* are transcribed, these genes do share synteny with both *pimJ* and *pimI*, and *scnJ* and *scnI*. RimL and RimM most likely play the same role as that of PimJ and PimI, considering the both homology and shared synteny.

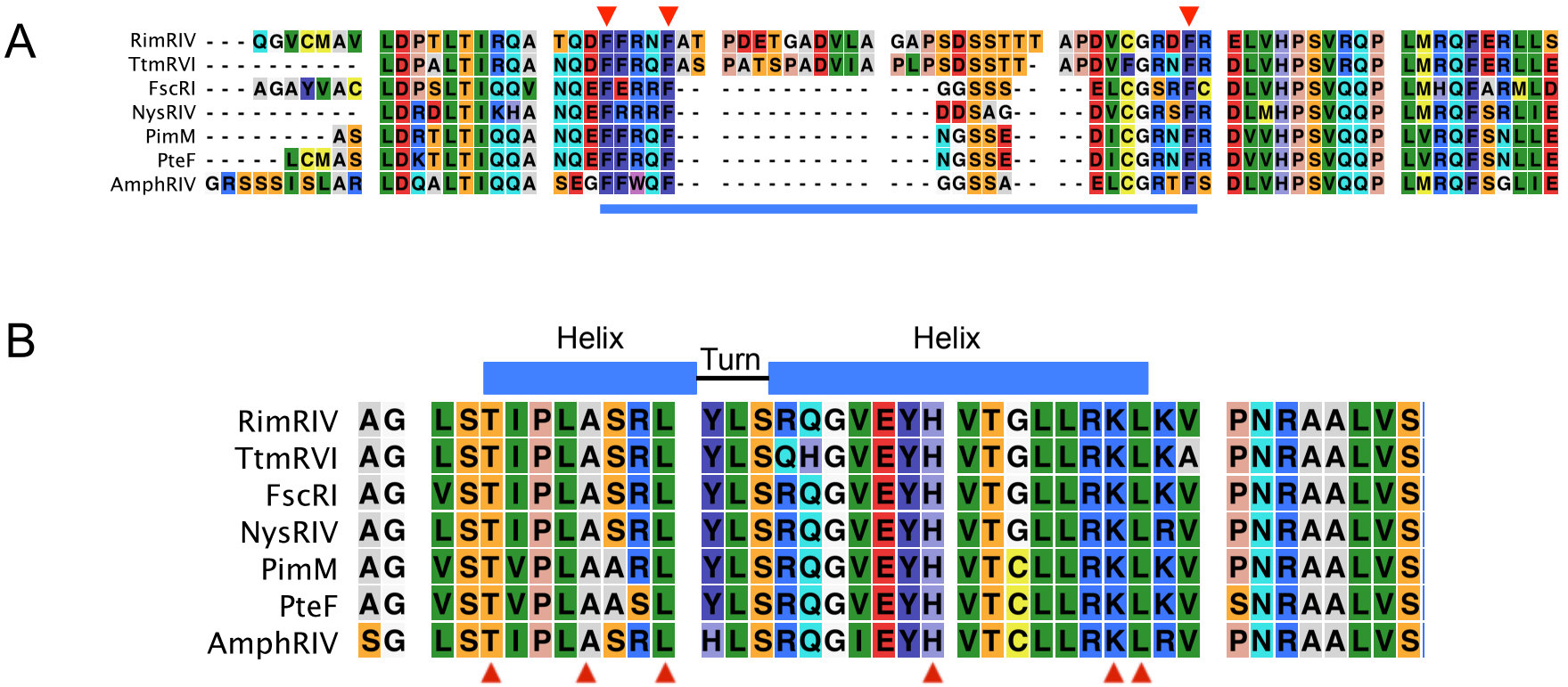


Figure 5.7 Alignment of RimRIV with other regulators of polyene biosynthesis. PimM, pimarcicin *S. natalensis;* PteF, filipin *S. avermitilis*; AmphRIV, amphotericin *S. nodosus;* FscRI, candicidin *S. sp* FR-008; NysRIV *S. noursei*, nystatin; TtmRVI, tetramycin, *S. ahygrocopicus* **A.** The PAS-domain, the blue bar indicates a potential ligand binding area. Catalytic amino acids are indicated by red triangles. **B.** The Helix-turn-Helix domain. Catalytic amino acids are once again indicated with red triangles.

5.3 Confirmation of *rim* deletion in SR20, SR50S and SR55

S. rimosus SR20 (M4018), SR50S (15883S) and SR55 (23383) do not produce rimocidin and it was predicted this was due to the 277 Kbp deletion present within SR20. Therefore, it was decided to use PCR to determine if the absence of rimocidin production was due to gene loss rather than gene mutation and also to confirm the *in silico* prediction. At the time of devising this experiment only two sequences of the *rim* biosynthetic cluster from *S. rimosus* were present on the Genbank database (Accession numbers DQ174320 and AY701054). Consequently these sequences were used to design the PCR primers for the experiment. A 2693 bp region of DQ174320 was designated Rim01, with two regions of AY701054, one being 1932 bp and the second of 1881 bp in size, named Rim02 and Rim03 respectively. Once the *rim* cluster was elucidated the sequences DQ174320 and AY701054 could be aligned (Figure 5.8 A), which allowed the corresponding genes to be identified as *SRIM7673* and *SRIM7671* respectively.

PCR amplification of the Rim01 region was achieved using the primer pair Rim01FR, Rim02 was PCR amplified using the Rim02FR primers and primers Rim03FR were used for Rim03 amplification (Figure 5.8 A and C). Genomic DNA from *S. rimosus* SR1, SR20, SR50S and SR55 was used as the template for PCR, with genomic DNA from *S. coelicolor* M145 used as a negative control. An amplification product was detected for *S. rimosus* SR1 (Figure 5.8 B) for each region and at the expected size, however no amplification products were detected for the other strains, other than non-specific products for SR20 and SR50S associated with the Rim03 primers (Figure 5.8 B). Therefore, it was concluded that the abolition of rimocidin production was indeed due to the absence of the *rim* cluster and not mutation of the genes within, confirming the *in silico* observation made.

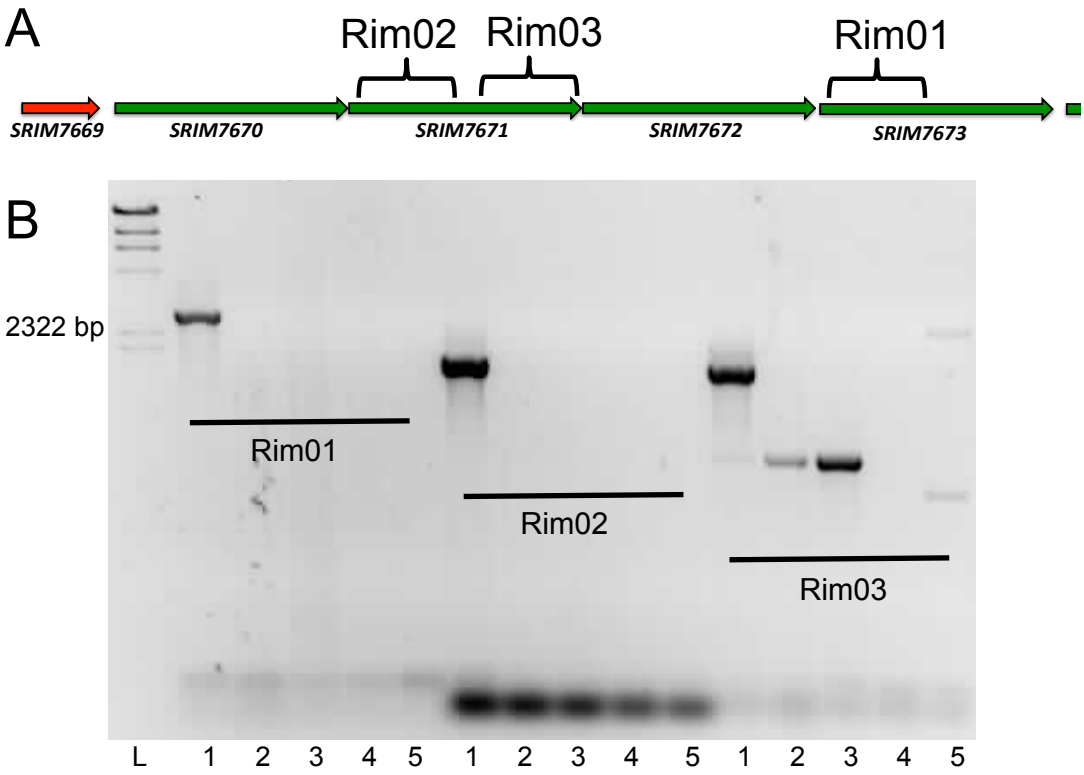


Figure 5.8 Confirmation of *rim* deletion in SR20, SR50S and SR55. A. Relative position of Rim01, Rim02 and Rim03 in relation to the *rim* biosynthetic cluster **B.** PCR confirmation of *rim* loss in strains SR20, SR50S and SR55. Electrophoresis gel of PCR samples. Lane L contains the λ*Hind*III ladder, lane 1 is SR1, lane 2 is SR20, lane 3 is SR50S, lane 4 is SR55 and lane 5 is *Streptomyces coelicolor* M145 used as a negative control. Expected product sizes for each target are: 2693 bp for Rim01, 1932 bp for Rim02 and 1881 bp for Rim03.

5.4 Generation of a rimocidin non-producing strain of *S. rimosus* SR1

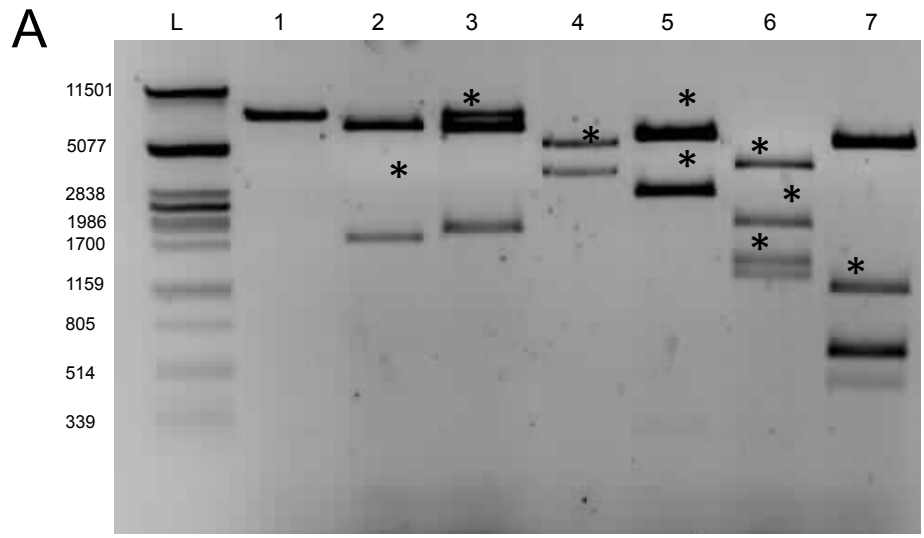
In order to generate a rimocidin knockout of *S. rimosus* SR1, via the homologous recombination method (Figure 5.1), a plasmid containing DNA amplified from the region of interest was generated. Once again the two sequences present on Genbank were used to carry out this experiment. Consequently this prevented the selection of an optimum area within the *rim* cluster to be targeted for deletion. When attempting homologous recombination, the longer the length of homologous DNA present on the plasmid, the greater the chance of recombination. To this end sequence AY701054 was chosen, as its total length was 8.3 Kbp. The sequence AY701054 corresponded with the gene *SRIM7671*, which encodes the protein RimS3 containing the modules 7-10 (Figure 5.2). Although this region was located halfway through the biosynthetic pathway (Figure 5.3), rimocidin production would still be eliminated; however a precursor of rimocidin would be produced instead.

In order to produce the knockout plasmid a parental plasmid with an antibiotic marker that could be used in *S. rimosus* was needed, for this reason pGB1 was selected as it contained a thiostrepton resistance marker. To this end pGB23, a pGB1-derived plasmid containing PCR amplified products from the gene *SRIM7671*, flanking the thiostrepton resistance (tsr) gene within the parental plasmid was constructed (Figure 5.9, C). Two regions from *SRIM7671*, used previously for deletion confirmation, Rim02 and Rim03 (Figure 5.8), were used for this purpose. The start of Rim02 lay at 5,416 bp downstream of the beginning of *SRIM7671's* coding sequence and with Rim03 positioned 9,887 bp from the start of *SRIM7671's* coding sequence (Figure 5.8 A). The two regions from *SRIM7671* were PCR amplified. PCR amplification of Rim02 was achieved using the primer pair Rim02FR, which had EcoRI and KpnI restriction sites incorporated into 5' end of both the forward and reverse primers, respectively. This PCR product was then purified and digested using EcoRI and KpnI and ligated into EcoRI and KpnI digested pGB1, generating pGB02. After construction of pGB02,

Rim03 was amplified as before and PstI and HindIII restriction sites were incorporated into the forward and reverse primers respectively. The Rim03 PCR product was purified then digested with PstI and HindIII and ligated with PstI and HindIII digested pGB02, generating pGB23 (Figure 5.9, C). This resulting plasmid was then subjected to restriction digest (Figure 5.9) in order to confirm that the pGB23 plasmid had been constructed.

After the plasmid pGB23 was constructed it was necessary to introduce this plasmid into *S. rimosus* SR1. In order to ensure protection against the methyl-dependent restriction system present within *S. rimosus* SR1, GM2929, a dam methylation minus strain of *E. coli* (Messer *et al.,* 1985) was transformed with pGB23. After passage through GM2929 pGB23 was introduced into *S. rimosus* SR1 via electroporation. Tryptone soya agar (TSA) plates containing thiostrepton were used for selection after electroporation, and any colonies present after 48 hours of growth were subsequently patched onto TSA plates containing either gentamicin or thiostrepton. Any patch of growth that was observed only on the thiostrepton plate was deemed to represent a successful double crossover event. Genomic DNA was isolated from putative double crossovers of *S. rimosus* SR1 to allow for initial PCR confirmation of pGB23 integration into the chromosome. One of these potential double crossovers, subsequently named *S. rimosus* ACM23, gave a positive result from the initial PCR check and was subjected to further scrutiny to confirm its *rim* knockout status.

Although integration of the pGB23 plasmid into the chromosome of *S. rimosus* ACM23 was indicated through the tsr^r gen^s phenotype coupled with fact that pGB23 is a sucide vector. Despite this, it was necessary to determine the precise location of the plasmid with the *S. rimosus* chromosome. A Prediction was made of the location of the integrated copy of pGB23 within the chromosome at which integration of pGB23 via homologous recombination of the two flanking regions resulting in an allelic replacement of part of the *rim* cluster with *tsr*. PCR was selected as the method of verification and a pair of primers were designed that lay 1156 bp upstream of the Rim02 region (Figure 5.8 A) and 233 bp downstream.



B

Lane	Restriction Enzyme	Expected fragment size (bp)
1	EcoRI	8630
2	HindIII & PstI	6747, 1883*
3	EcoRI & KpnI	6688*, 1942
4	NdeI	5224, 3406*
5	NcoI	5993*, 2863*, 174
6	NdeI & NcoI	3621, 1972*, 1603*, 1260*, 174
7	SalI	4993, 1351, 625*, 615 x2, 431

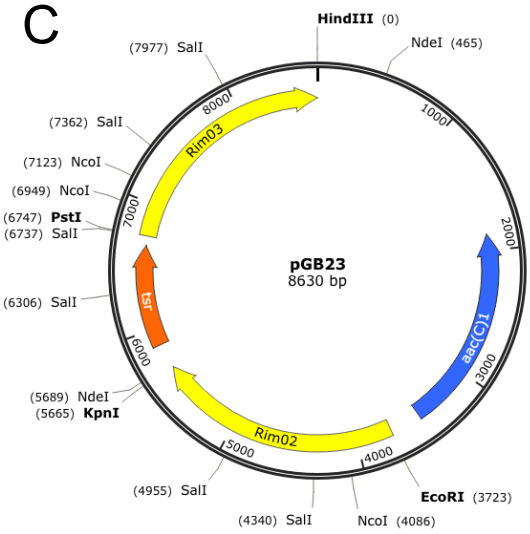


Figure 5.9 Confirmation of pGB23 by restriction digest. Asterisks highlight those bands predicted to only be produced for pGB23. **A.** Electrophoresis gel of pGB23 digested by select restriction enzymes. Lane L contains the λ*Pst*III ladder. Each lane contains pGB23 plasmid DNA digested with the following restriction enzymes: Lane 1 EcoRI, Lane 2 HindIII and PstI, Lane 3 EcoRI and KpnI, Lane 4 NdeI, Lane 5 NcoI, Lane 6 NdeI and NcoI, Lane 7 SalI. **B.** Expected band sizes for each restriction digest. **C.** Map of pGB23 showing relative position of restriction sites for each listed restriction enzyme.

Another set of primers were designed that bind sites 455 bp upstream of the Rim03 region (Figure 5.8 A) and 1299 bp downstream. These primer pairs are represented by set A and B (Figure 5.10 B) respectively, and were used to confirm an intact copy of *SRIM7671* in the parental strain of *S. rimosus*. Therefore a negative result from a PCR carried out with these primer sets would indicate disruption of *SRIM7671*. Additional primer sets were produced and the relative position of these primers in relation to the *SRIM7671* gene, and the *tsr* disruption are shown (Figure 5.10 B). Primer set E and F (Figure 5.10 B) could be used to confirm the disruption of *SRIM7671*, as a positive result would only occur in the mutant and not the wild type strain. Primer sets C and D (Figure 5.10 B) were used as a positive control for the *tsr* associated primers. Genomic or plasmid DNA was used as the template for amplification and the samples were ran on an electrophoresis gel (Figure 5.10 A). Negative results were given when using primer set A (Figure 5.10 B) with genomic DNA from ACM23 and the plasmid pGB23, however a positive result was seen with SR1 (Figure 5.10 A, indicated by the black arrow). Although there are also non-specific products, which can be observed for ACM23, the predicted band size of 3331 bp is only present for SR1. Similar results were seen with primer set B except there are no products seen for ACM23 and less non-specific products for SR1 (Figure 5.10 A). Taken together these data demonstrate that an allelic replacement of *SRIM7671* with *tsr* took place after the introduction of the suicide vector, pGB23 into *S. rimosus* SR1. Unexpectedly, primer set E (Figure 5.10 B) produced a product for both *S. rimosus* SR1 and *S. rimosus* ACM23. Despite this, the product was not 3629 bp, the expected PCR product size, and may be due to non-specific binding of the PCR primers. BLASTn alignments revealed sequence homology between primer set E and the gene *SRIM7674*. If primer set E annealed to this region during the PCR process, then a product of 2343 bp would be produced.

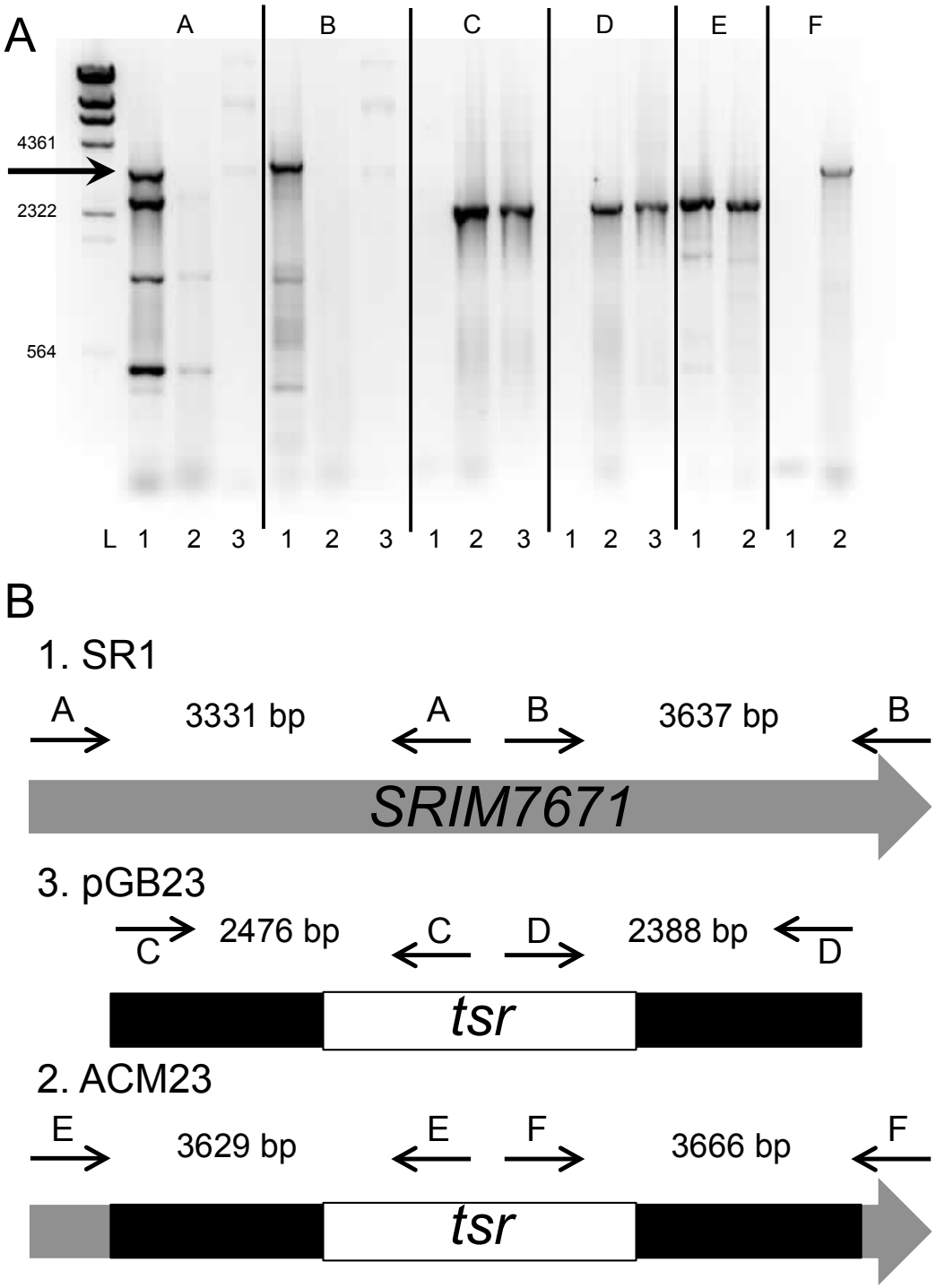


Figure 5.10 PCR confirmation of *rim* knockout in SR1 and generation of strain ACM23. **A.** Electrophoresis gel of PCR confirmation of *SRIM7671* disruption in ACM23. Letters A-F represent the primer pair used. 1 = SR1 2 = ACM23 3 = pGB23 **B.** Relative position of the primers in relation to the *SRIM7671* gene either wild type or disrupted and within the pGB23 plasmid.

It has been demonstrated that amplification bias can occur with smaller PCR products, suggesting that this smaller, unexpected product was preferentially amplified over the larger, expected product (Suzuki & Giovannoni 1996; Polz & Cavanaugh 1998; Huber *et al.,* 2009). A positive result for ACM23 was observed with primer set F (Figure 5.10 A and B) but not for SR1 that, combined with the previous data, suggests that *SRIM7671* was disrupted, as a product would only appear if the correct integration of pGB23 into *SRIM7671* had taken place.

In addition to the PCR confirmation of *SRIM7671* deletion within *S. rimosus* ACM23, it was also necessary to carry out rimocidin bioassays to determine if this strain was no longer able to produce the antibiotic, as elimination of RimS3 activity should prevent rimocidin production. Examination of the biological activity of the rimocidin derivative was an additional benefit of the bioassay. *Saccharomyces cerevisiae* was selected as the test organism due to its sensitivity to rimocidin (Molzahn & Woods 1972). The results of the bioassay revealed no antifungal activity for the derivative and confirmed elimination of rimocidin production in ACM23 (Figure 5.11 A). In order to investigate the effect of the abolition of rimocidin production, a crude analysis of OTC productivity by ACM23 was undertaken using bioassays with *Bacillus subtilis* as the test organism. The Quasi Latin method (Bailey 1984) was employed (Figure 5.11 B) using extracts prepared from liquid cultures that were grown for 10 days. This initial analysis could not confirm greater OTC production in ACM23 when compared with SR1.

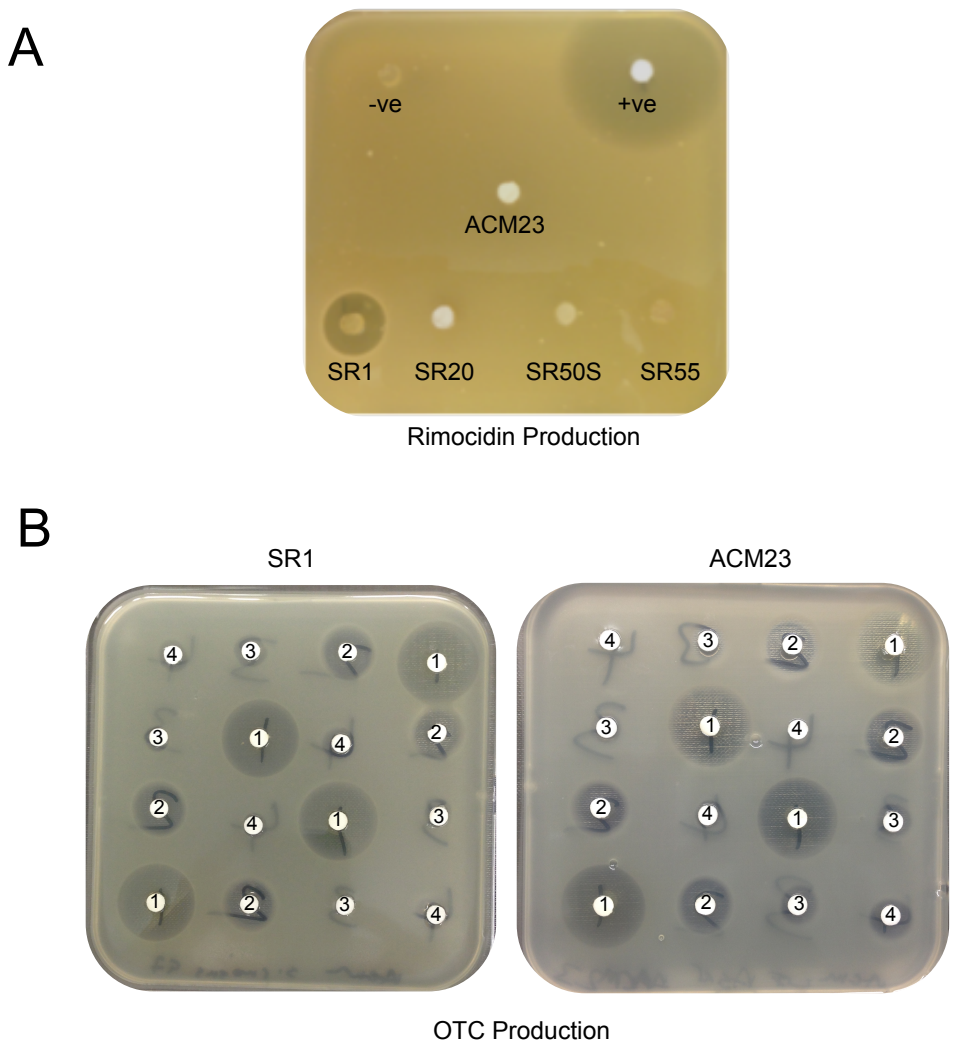


Figure 5.11 Loss of rimocidin production and analysis of OTC production by *S. rimosus* ACM23. A. Analysis of rimocidin production by each *S. rimosus* strain. *Saccharomyces cerevisiae* was used as the test organism. Each strain was grown on Emerson's agar for 10 days, after which agar plugs were transferred to the bioassay test plates. Sterile Emerson's agar was used as a negative control and Whatmann paper discs impregnated with 10 µL of 10 mg/mL of cyclohexamide was used as a positive control. **B.** Quasi Latin method comparison between SR1 and ACM23 of OTC production. *B. subtilis* was used as the test organism. 1 = 15 µg/mL of OTC, 2 = 1.5 µg/mL of OTC, 3 = undiluted extract from SR1 or ACM23, 4 = 1/10 dilution of extracts from either SR1 or ACM23.

5.5 Discussion

Sequencing the genome of *S. rimosus* SR1 allowed for the *rim* biosynthetic cluster to be fully elucidated. Seco *et al*, 2004, had already discovered 19 kbp of the *rim* biosynthetic cluster from *S. diastaticus* and here a further 76.7 kbp was characterised. In the additional 76.7 Kbp of the cluster the full sequence of *rimB* was discovered, as well as ten other genes; four putative transcriptional regulators, two modifying/accessory genes and four Type I PKS. The *rim* gene cluster shares a similar arrangement to that of the amphotericin (Carmody *et al.,* 2004), nystatin (Sekurova *et al.,* 2004) and tetramycin (Cao *et al.,* 2012) biosynthetic gene clusters, in that the PKS genes are flanked by four transcriptional regulators and modifying/accessory genes, suggesting that rimocidin production may share an analogous mechanism of regulation. It is also worthwhile noting that RimD shares 91 % identity with PimE, a cholesterol oxidase that is essential for the production of pimaricin (Mendes *et al.,* 2007a). Despite this, disruption of RimD does not abolish rimocidin production (Seco *et al.,* 2004). As only two regulators are required for pimaricin production it is possible, in the presence of additional regulators, that a signaling molecule, in this case RimD, is no longer required. It seems likely that this is the case as, with the exception of the biosynthetic clusters responsible for tetramycin and rimocidin, polyene biosynthetic clusters with at least four transcriptional regulators do not contain a homologue of PimE (Chen *et al.,* 2003; Carmody *et al.,* 2004; Sekurova *et al.,* 2004; Kong *et al.,* 2013). Sekurova *et al* examined the four regulators of the nystatin gene cluster to determine the role each played. Through the production of in-frame deletions of *nysRI, nysRII* and *nysRIII* Sekurova *et al* found that all three genes were required for positive regulation of nystatin production but were not essential, although the loss of all three genes abolished production (Sekurova *et al.,* 2004). Insertional inactivation of *nysRIV* also demonstrated a positive regulation role by NysRIV as production of nystatin was severely reduced when compared with the wild type. Serkuova *et al* also investigated the target genes of each regulator by

undertaking XylE activity assays using a *xylE* reporter gene that had putative promoter fragments from the nystatin gene cluster cloned upstream. These results demonstrated that each of the regulators play a significant role in the regulation of *nysA*, the gene that encodes the loading module, and consequently plays a crucial role in initiation of nystatin biosynthesis (Sekurova *et al.,* 2004). Given the similarity between the nystatin regulators and the four putative regulators of the *rim* gene cluster, it is proposed that these proteins play an analogous role in rimocidin production. Further to this it is predicted that *SRIM7669* plays the most important role, as it is homologous to *nysRIV*, which was shown to complement all nystatin regulatory mutants to some degree. If this is the case then *SRIM7669* mutant strains of *S. rimosus* SR1 would show little, if not complete loss of, production of rimocidin.

The extra catalytic domains detected in module 12 contradict the solved structure of rimocidin (Pandey & Rinehart 1977; Volpon & Lancelin 2002), it was therefore necessary to ascertain if these domains were active. The results demonstrated that the KR, DH and ER domains contained the amino acid residues and motifs required for activity (Figure 5.5, A). It is possible that the second form of rimocidin (Figure 5.5 B, 2) produced by the activity of module 12 could be unique to the biosynthetic cluster of *S. rimosus* SR1 or that these domains catalyse the removal of the carbonyl group at low efficiency. If the latter were true then this would explain why the second form has not been previously detected in pure samples of rimocidin. It would be expected that in crude secondary metabolite extractions from *S. rimosus* SR1 cultures, that this second form of rimocidin would be detected.

The loss of the *rim* gene cluster from *S. rimosus* SR20 might be a reason for this strain's increased OTC productivity compared with *S. rimosus* SR1, to test this hypothesis the strain *S. rimosus* ACM23, which lacks rimocidin production, was produced. The *rim* cluster knockout in *S. rimosus* ACM23 was achieved by disruption of gene *SRIM7671*, which encodes RimS3. RimS3 is involved in the PKS elongation steps of module 7-10, and therefore interruption of rimocidin at this point would mean the incorporation

of eight units of malonyl-CoA into the polyketide chain (Figure 5.4). The predicted chemical formula of the derivative produced is $C_{22}H_{36}O_5$, as the location of the *tsr* cassette within *SRIM7671* would disrupt the AT of module 9 and everything downstream of that region. This would mean modules 1-8 would function normally and it could be for this reason, that the initial analysis undertaken to establish if *S. rimosus* ACM23 produced more OTC than *S. rimosus* SR1 were inconclusive. Mass spectral analysis could allow for a more concrete answer, as well as allowing for the identification of the predicted derivative produced by the strain *S. rimosus* ACM23. Additionally, identification of a derivative would confirm the arrangement and function of the genes within the biosynthetic pathway.

In order to obtain a true representation of a rimocidin knockout and the influence this would have on OTC production a *rimA* or *SRIM7669* knockout would need to be produced. In this way no carbon would be redirected into the rimocidin pathway and therefore more acetate would be available for OTC production.

Chapter 6

6 Mass spectrometric analysis of rimocidin and related metabolites produced by *S. rimosus*

6.1 Introduction

S. rimosus SR20 (M4018) lacks a secondary metabolite cluster that encodes a Type I polyketide synthase (T1PKS) responsible for the production of rimocidin. This rimocidin cluster has not yet been fully characterised but has been studied extensively in *S. diastaticus* var. 108, where it was found that CE-108 was also made via the cluster (Seco *et al.,* 2004). The elucidation of the rimocidin genes in Chapter 5 was consistent with previous predictions and suggested that a loading module and 13 extender modules were encoded by the biosynthetic cluster. However the domain architecture of module 12 did not agree with the known structure of rimocidin (Volpon & Lancelin 2002), as this module would remove an oxygen atom, therefore incorporating 2 extra hydrogen atoms to the carbon backbone (Figure 6.1). It is possible that the ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER) domains are inactive but the genetic analysis undertaken in Chapter 5 suggests that these catalytic domains are active. As the protein encoded by module 12 would result in the production of a molecule distinct from rimocidin, we decided to analyse the secondary metabolites produced by *S. rimosus* SR1 (G7) using high resolution mass spectrometry (HRMS) as this would detect this alternative form of rimocidin (rimocidin*). As the rimocidin cluster is capable of producing CE-108 then HRMS analysis would identify not only rimocidin* but also CE-108 and the alternative form of CE-108 (CE-108*) (Figure 6.1). Due to the predicted function of *SRIM7671* and the elucidated rimocidin biosynthetic cluster, ACM23, the rimocidin negative mutant, should produce a truncated form of rimocidin. Detection of this molecule in secondary

metabolite extracts of ACM23 would support the predicted rimocidin biosynthetic cluster and the function of *SRIM7671*. For this reason HRMS was selected to analyse the secondary metabolite extractions from the 5 *S. rimosus* strains.

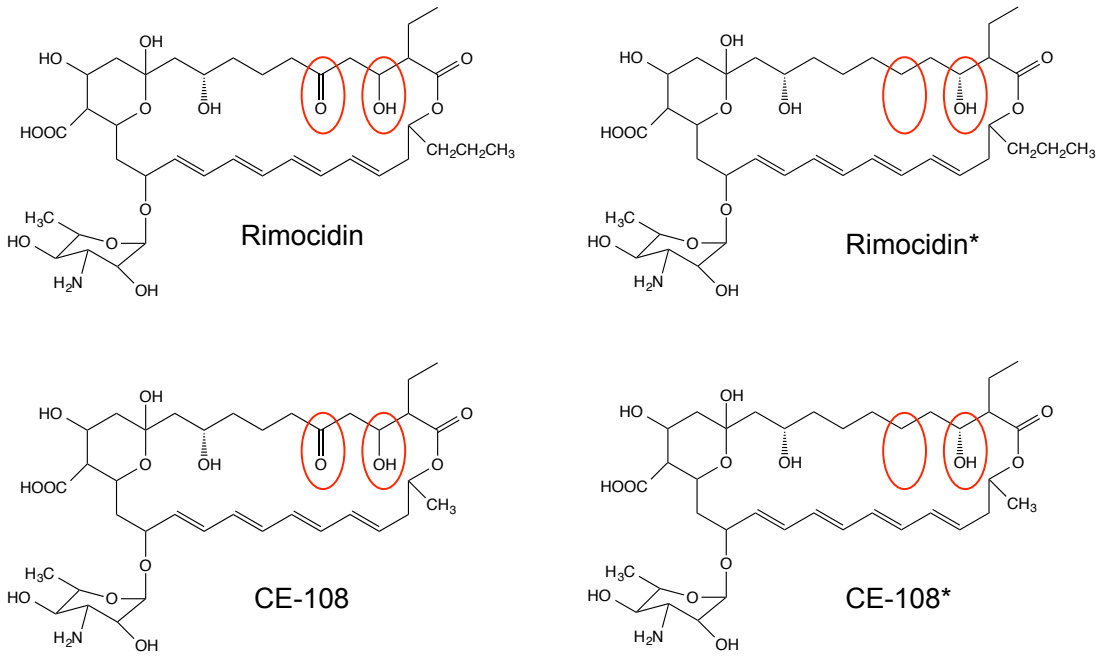


Figure 6.1 Module 12 modification of Rimocidin and CE-108. Rimocidin* and CE-108* represent the chemical structure of rimocidin and CE-108 produced by active KR, DH and ER within module 12. The red circles indicate the carbon groups that are changed by Module 12.

6.2 Preparation of samples and HRMS data for analysis

Secondary metabolite extracts were prepared for all *S. rimosus* strains and the rimocidin mutant, ACM23. Overall dry weight was determined and sufficient mass spectrometry grade methanol was used to resuspend the residue so that the samples had a concentration of 1 mg/mL. At this point the samples were subjected to positive ion HRMS, using a 10 µL injection volume for each aliquot, as well as methanol to be used as the solvent blank.

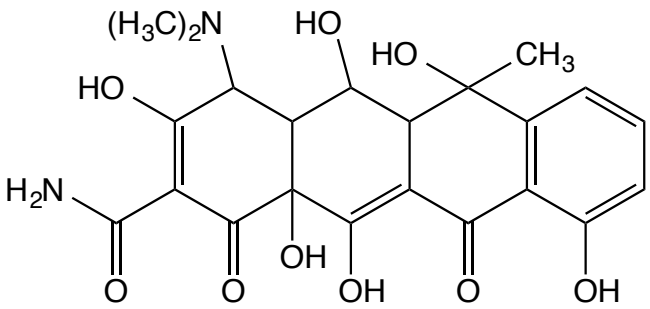
Once the raw MS data was obtained, it was processed using MZmine version 2.10 (Pluskal *et al.,* 2010) in order to generate peak lists. The data processing steps required peak detection involving mass detection and chromatogram building. After this deconvolution was carried out to remove any interfering signals (Wilmes *et al.,* 2010) to improve the identification of desired signals. This was followed by deisotoping however only peaks previously identified and assigned an ID are deisotoped by this process. Chromatogram alignment and gap filling was then carried out. At this point gap filling could be applied to the data to identify any previously undetected peaks. All peaks found within the solvent were removed as well as peaks with low intensity; this cut off was 1.0E4. In order to predict molecular formulas and minimise mis-assignment, adducts and complexes were identified. At this point formula prediction could be carried out as well as manual search for m/z and RT of interest.

6.3 Detection of oxytetracycline in *S. rimosus* extracts by HRMS

As OTC is produced by *S. rimosus* SR1, SR20, SR55 and ACM23, but not SR50S, it was selected as a control for presence and absence of a metabolite, under positive ion conditions. The chemical formula of OTC is $C_{22}H_{24}N_2O_9$, which has a predicted molecular mass of 460.1482, and therefore is to possess a m/z of 461.1482 on a positive ion survey scan chromatogram. At an m/z of 461.1554 a peak was detected that was only

present in the samples of *S. rimosus* SR1, SR20, SR55 and ACM23 (Figure 6.2), and the predicted molecular formula associated with the peak was consistent with OTC.

A



B

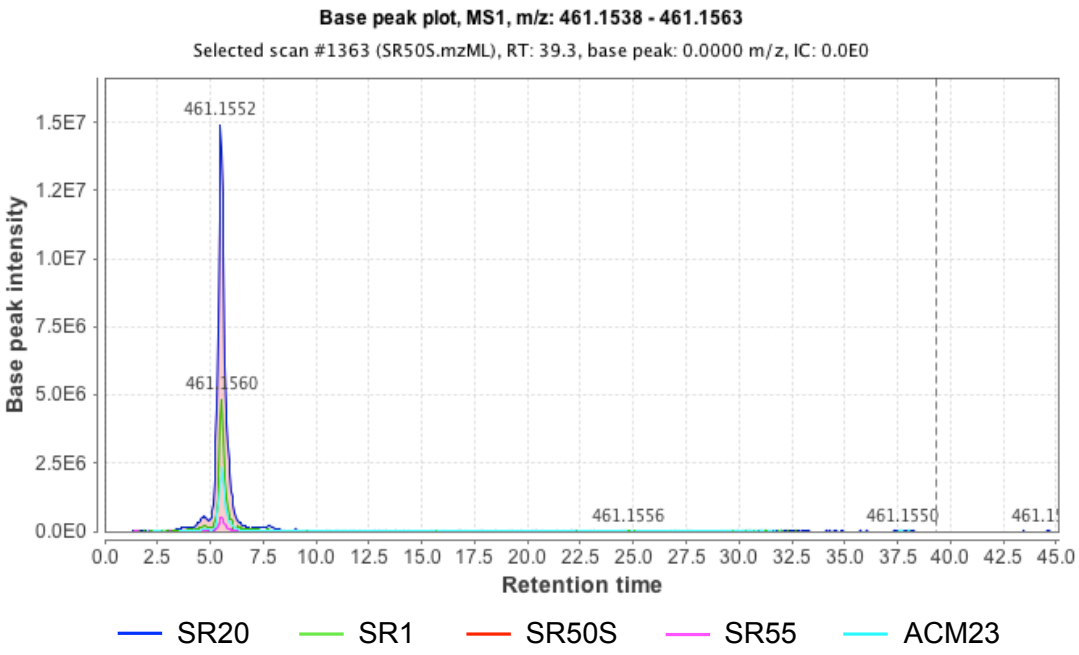
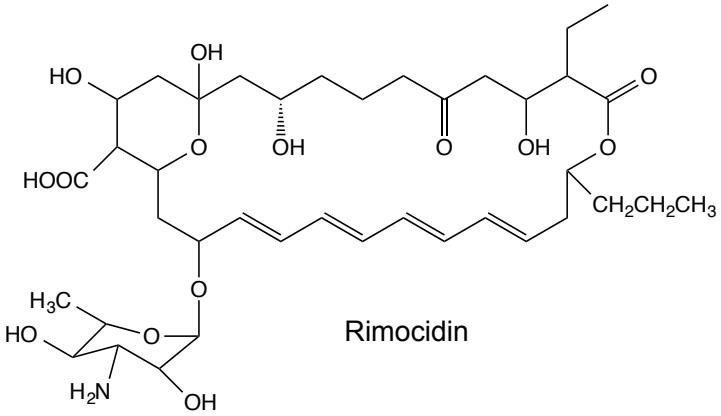


Figure 6.2 Molecular structure of oxytetracycline, and corresponding chromatogram. **A.** Molecular structure of oxytetracycline. **B.** Chromatogram of oxytetracycline analysis by positive ion MS.

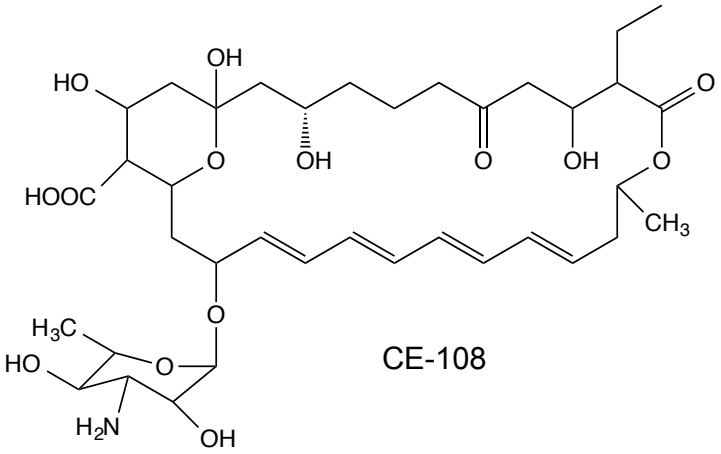
6.4 Identification of rimocidin and CE-108 by HRMS

Rimocidin, like OTC, was predicted to only be present in SR1, given the genomic data and the results of rimocidin bioassays. The molecular formula of rimocidin is $C_{39}H_{61}NO_{14}$ and the molecular mass was calculated as 767.4092. A peak at m/z 768.4166 was observed, only for SR1 (Figure 6.3 B), which coincided with that of rimocidin. The rimocidin biosynthetic pathway produces two different molecules in *S. diastaticus* var. 108, depending upon the starter unit incorporated during the initial steps of biosynthesis (Seco *et al.,* 2004). The additional molecule, CE-108, is almost identical to rimocidin but lacks 2 carbon atoms and 4 hydrogen atoms, giving it a molecular formula of $C_{37}H_{57}NO_{14}$. So far to date CE-108 has not been associated with cultures of *S. rimosus* SR1, but examination of the MS data revealed an m/z peak at 740.4229 present only in the SR1 sample (Figure 6.3 C). During the elucidation of the rimocidin biosynthetic pathway, extra catalytic domains were identified in module 12 with respect to the biosynthetic cluster from *S. diastaticus*. These domains were predicted to be active, and if so, would produce rimocidin* (Figure 6.4 A). Rimocidin* would have a lower molecular mass from that of rimocidin, due to the removal of an oxygen atom and the addition of two hydrogen atoms, therefore a molecular mass of 753.4299 would be expected. Analysis of the MS data identified a molecule with the predicted molecular formula of $C_{39}H_{63}NO_{13}$ at an m/z of 754.4387, this molecule was only present in *S. rimosus* SR1 (Figure 6.4 B). It is likely this is the alternative form of rimocidin, produced due to active KR, DH and ER domains within module 12. If this is the case then an alternative form of CE-108 (CE-108*) would be expected as well, due to the shared biosynthetic pathway, and this molecule would have a molecular mass of 725.3934. At an m/z of 726.4056 a MS peak was detected that gave a predicted molecular formula of $C_{37}H_{59}NO_{13}$, which would correspond to CE-108 lacking an oxygen atom but gaining two hydrogen atoms.

A

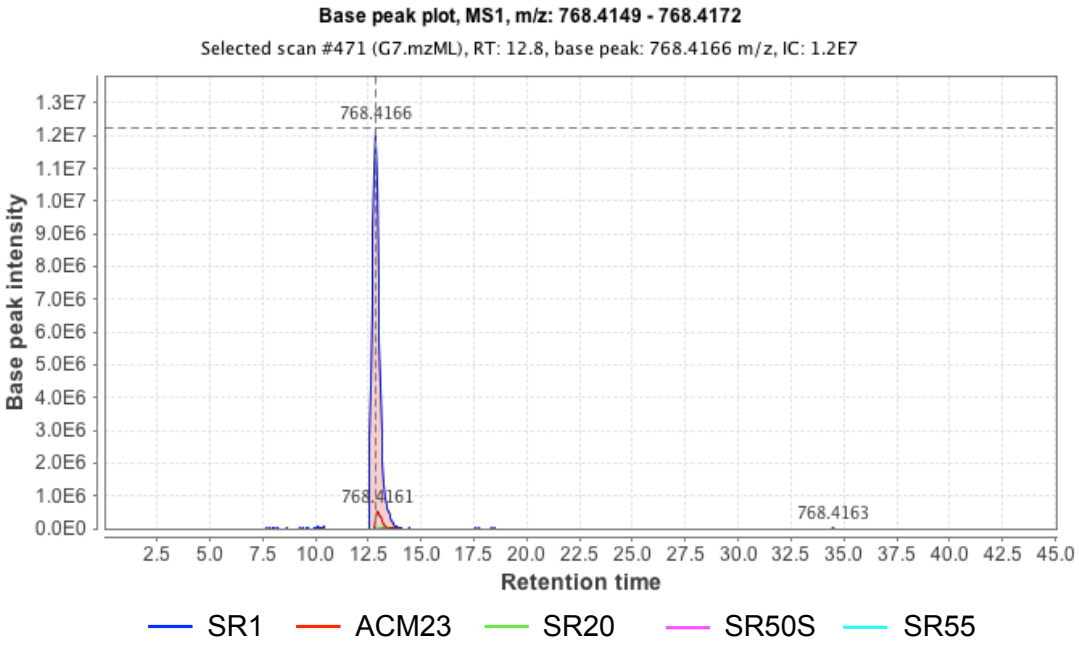


Rimocidin



CE-108

B



C

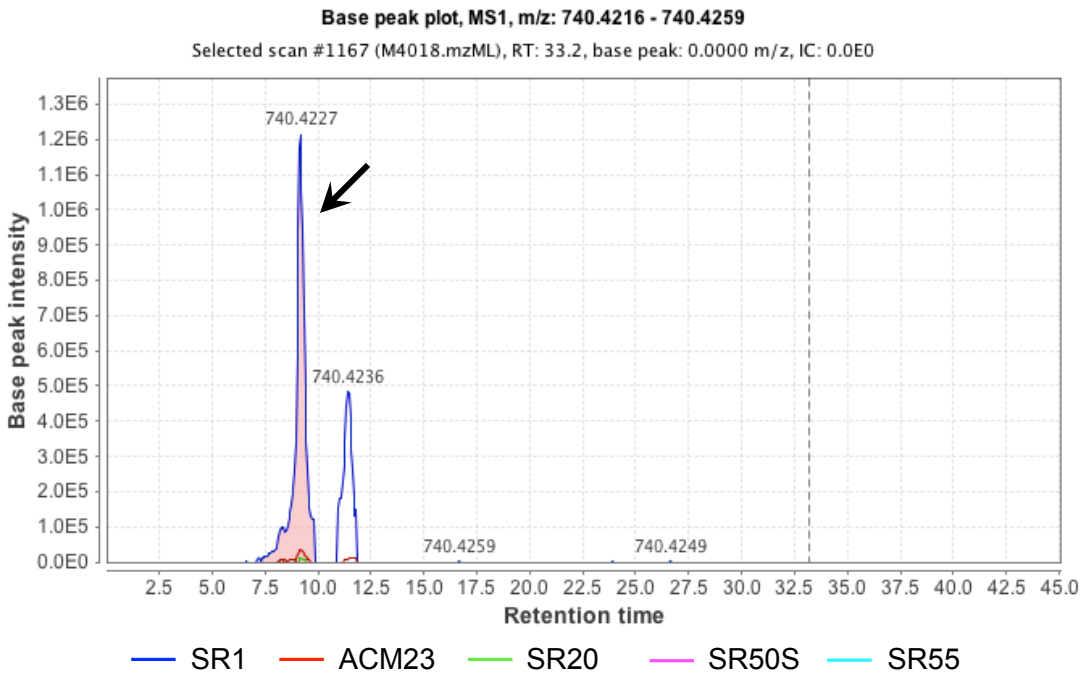
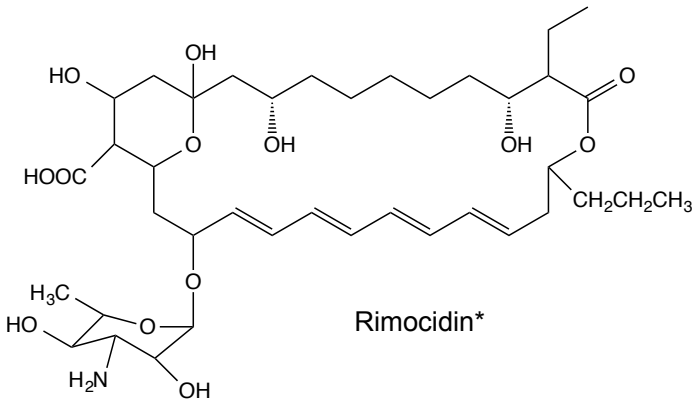
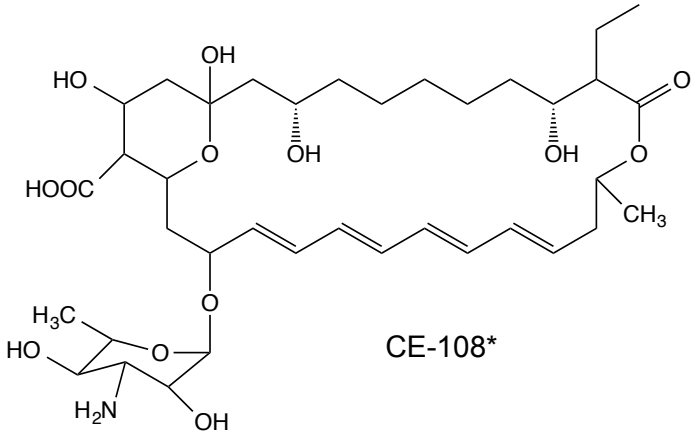


Figure 6.3 Molecular structures of rimocidin and CE-108, and corresponding chromatograms. The peak corresponding to the molecule of interest is highlighted with a black arrow. **A.** Molecular structure of both rimocidin and CE-108. R = CH_2-CH_2-CH_3 (rimocidin) and CH_3 (CE-108). **B.** Chromatogram of rimocidin. **C.** Chromatogram of CE-108

A

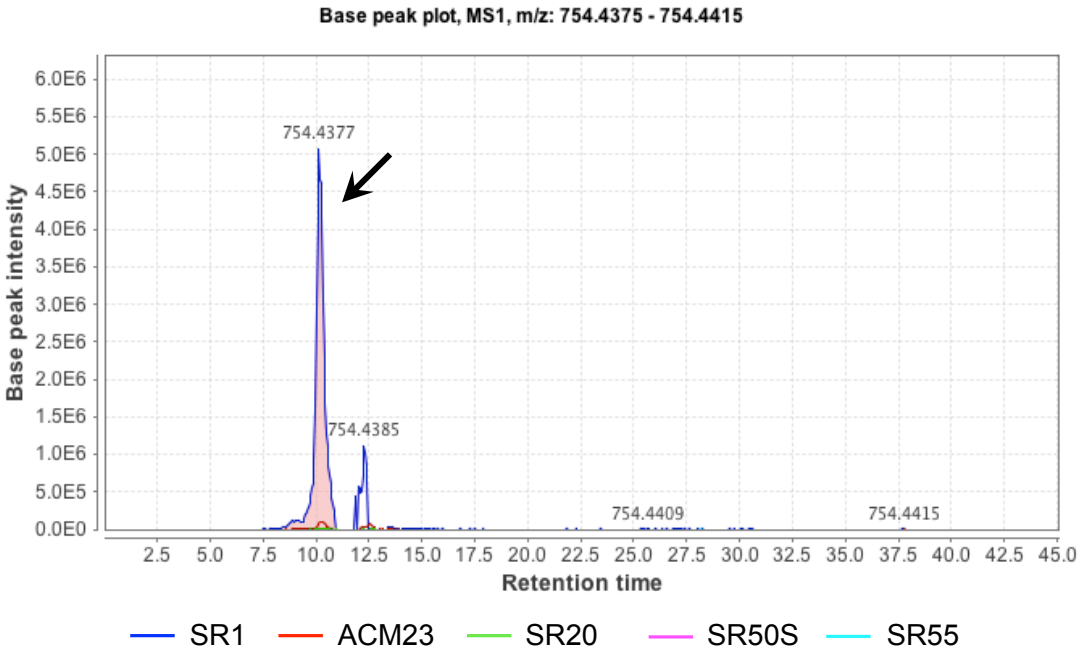


Rimocidin*



CE-108*

B



C

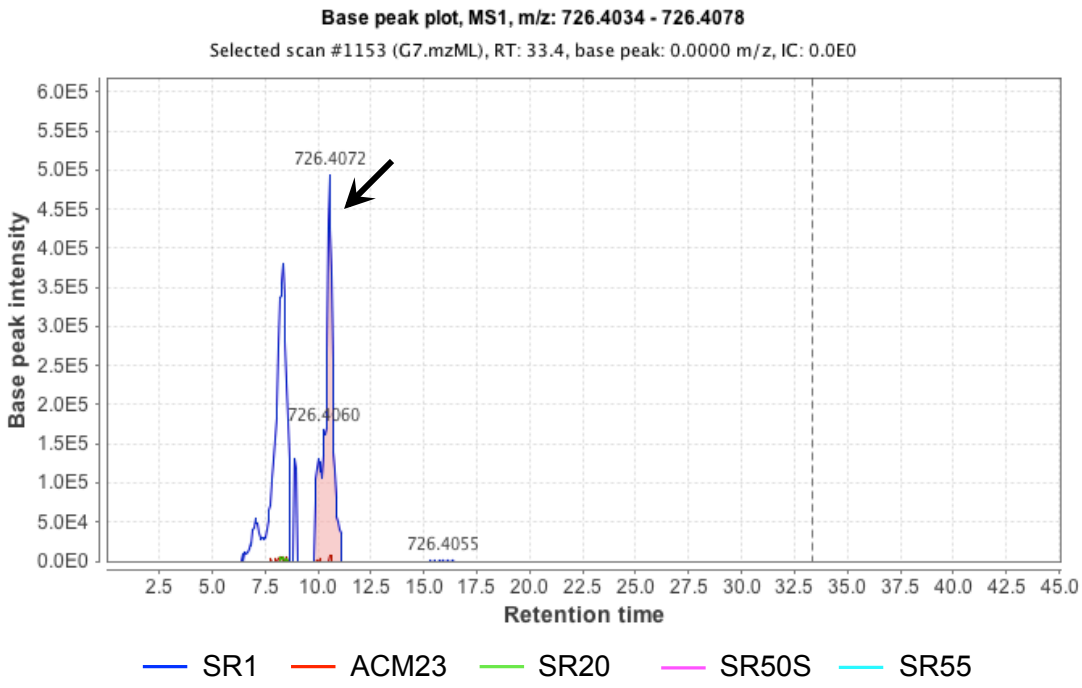
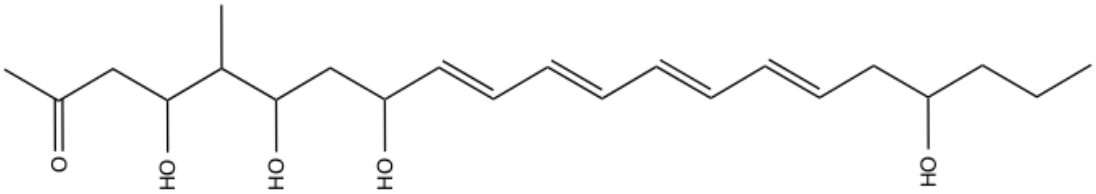


Figure 6.4 Molecular structures of rimocidin* and CE-108*, and corresponding chromatograms. The peak corresponding to the molecule of interest is highlighted with a black arrow. **A.** Molecular structure of both rimocidin* and CE-108*. **B.** Chromatogram of rimocidin*. **C.** Chromatogram of CE-108*

6.5 Identification of derivative 23 by HRMS

The rimocidin knockout mutant generated in Chapter 5 was predicted to produce a derivative of rimocidin with the molecular formula of $C_{22}H_{36}O_5$ and to contain a linear carbon backbone (Figure 6.5). Identification of this molecule in secondary metabolite extracts of ACM23 would confirm the predicted biosynthetic pathway of rimocidin and the function of gene *SRIM7671*. The calculated molecular mass of derivative 23 was 380.2563 and would be predicted to have an m/z of 381.2573 on the mass spectrum under positive ion conditions. Examination of the chromatograph data in MZmine 2.10 revealed a peak at 381.2034, which was only present in the secondary metabolite profile of ACM23. Although the observed mass and calculated mass were not identical both were within 5 ppm, and therefore within acceptable parameters (Webb *et al.,* 2004).

A



B

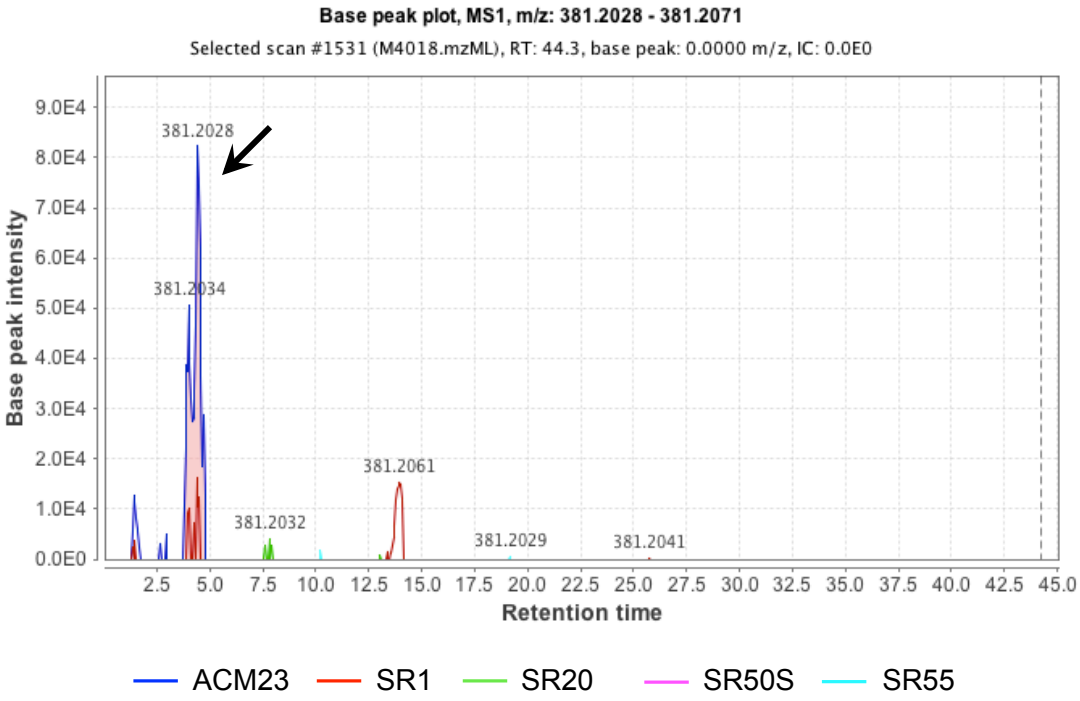


Figure 6.5 Molecular structure of derivative 23, and corresponding chromatogram. The peak corresponding to the molecule of interest is highlighted with a black arrow. **A.** Molecular structure of derivative 23. **B.** Chromatogram of derivative 23.

6.6 Discussion

The analysis of the genes associated with rimocidin allowed for confirmation of the previously predicted biosynthetic pathway. The only variation detected was that of 3 extra catalytic domains within module 12, whose presence would alter the final rimocidin molecule. Examining the amino acid sequence of the KR, DH and ER domains suggested these domains were active and so an alternative form of rimocidin was predicted to be within the extracts of *S. rimosus* SR1. Rimocidin* was readily detected in the SR1 extracts during MS analysis (Figure 6.4 B). There was a slight peak for ACM23, however a similar peak was detected on the chromatogram of rimocidin for this strain. However rimocidin is not produced by ACM23 as the rimocidin bioassays performed with extracts from this strain did not inhibit the growth of *Saccharomyces cerevisiae*. This suggests that these slight peaks seen on the chromatograms for both rimocidin and rimocidin* for ACM23 does not indicate the presence of either metabolite. Going by the base peak intensity of rimocidin, which was around 1.2E7 compared to 5.0E6 for rimocidin*, indicates that the rimocidin form is favoured during the biosynthesis process. The same was true for the CE-108 and the alternative form CE-108*, as base peak intensities for both were 1.2E6 and 5.0E5, respectively. This supports the theory that the extra module 12 domains are active but the efficiency of the catalytic activity, for the removal of the carbonyl group, is low.

Genomic analysis of the later production strains revealed the loss of the rimocidin cluster in *S. rimosus* SR20 and therefore SR50S (15883S) and SR55 (23383). It was decided to generate a deletion within the biosynthetic pathway rimocidin knockout in SR1 in order to establish any connection between rimocidin production and OTC productivity. The rimocidin gene knockout also provided a means of confirming the linkage of the predicted biosynthetic cluster with the rimocidin production phenotype. As mentioned previously, we were unable to detect rimocidin by MS data in ACM23 extracts, therefore supporting the bioassay data and confirming the loss of

rimocidin production in this strain. The placement of the gene disruption in *SRIM7671* would stop the synthesis of rimocidin after the acyl carrier protein (ACP) of module 8, therefore generating a carbon backbone containing 22 carbon atoms, 36 hydrogen atoms and 5 oxygen atoms. A molecule that corresponding to this was identified in the MS data for ACM23, but absent for the other strains, therefore *SRIM7671* can provisionally be associated with encoding the KR and ACP of module 7, plus all of module 8. However increased OTC productivity in ACM23 was not observed (Figure 6.2 A), as the base peak intensity was 2.5E6 compared to 5.0E6 for SR1. SR20 had the greatest base peak intensity for OTC, with SR55 having the smallest. The genetic stability of SR55 is highly variable, causing the strain to spontaneously lose OTC production, and it is likely this is the reason why SR55 produced the least OTC.

It should be noted that this is preliminary and in order to receive a true representation of OTC productivity and confirmation of the alternative forms of rimocidin and CE-108, these experiments should be performed in triplicate. In addition tandem MS should be performed for the elucidation of the molecular structures of molecules predicted to be rimocidin* and CE-108*.

Chapter 7

7 Purification and initial characterisation of the putative SARP, OtcR

7.1 Introduction

From the SNC and genomic deletion comparison between the *S. rimous* strains described in Chapter 4, one of the non-synonymous SNCs identified for strain SR55 (23383) was selected for further investigation. This SNC was found within the gene *SRIM7546*, renamed here *otcR*, which lies just downstream of *otrB*. Bioinformatical analysis indicated that this gene encoded a *Streptomyces* antibiotic regulatory protein (SARP). A transcriptional regulatory (Trans-Reg) domain (SMART accession number SM00862) was found within the amino-terminal (N-terminal) region of OtcR and a bacterial transcriptional activator domain (BTAD) within the carboxy-terminal (C-terminal) (Figure 7.1 A). The Trans-Reg domain is reminiscent of the winged helix-turn-helix (wHTH) domains of OmpR family proteins (Martínez-Hackert & Stock 1997; Hennecke *et al.,* 2005; Narayanan *et al.,* 2012) and are a common feature of SARPs (Arias *et al.,* 1999; Sheldon *et al.,* 2002; Novakova *et al.,* 2011), these domains are also thought to play a role in DNA binding. MEME (Bailey *et al.,* 2009) analysis of Trans-Reg domains from SARPs and OmpR family proteins found three motifs of interest. One of the motifs identified was shared only by N-terminal Trans-Reg domains, another motif was shared by C-terminal Trans-Reg domains and finally a motif was discovered that was common to all the Trans-Reg domains analysed (Figure 7.1 B). It is within this shared motif that the non-synonymous SNC of SR55 is found, phenylalanine substituting for leucine (Figure 7.1 C). The leucine at this position is very well conserved amongst Trans-Reg domains, with the exception of SnorA (also referred to as SnoA) and DnrI (Ylihonko *et al.,* 1996), in which an isoleucine or phenylalanine are substituted respectively. It is possible that substitution by phenylalanine

alters the DNA-binding properties of the Trans-Reg region and may be a reason for increased OTC productivity in SR55, assuming OtcR is a regulator of the *otc* biosynthetic cluster. Therefore it was decided to study this gene further to elucidate its function and the significance of the non-synonymous SNC within SR55's copy of *otcR*. The most direct method of establishing the role of *otcR* was to generate gene knockouts in both SR1 (G7) and SR55, with the addition of DNA-binding studies for both the wild type (SR1) and mutated (SR55) proteins.

REDirect (Datsenko & Wanner 2000; Gust *et al.,* 2003) was the method chosen for the generation of gene knockouts as a cosmid library was available for strain SR20 (Almutairi, 2007) that could be used. This method allowed for the generation of an *otcR* deleted cosmid that could be used to delete *otcR* in both SR1 and SR55 through intergenic conjugation (Mazodier *et al.,* 1989; Phornphisutthimas *et al.,* 2010) or electroporation (Pigac & Schrempf 1995). The REDirect method utilises the enhanced recombination rate of λ Red (γ, β, exo), to incorporate linear DNA, in this case a PCR product of an antibiotic resistant cassette, into the selected cosmid, deleting the gene of choice.

For the investigation of the DNA-binding properties of the OtcR protein, pure fractions of the protein were required. Initially band shift assays were to be undertaken to identify DNA-binding capabilities of the two forms of the OtcR protein by comparing the differences DNA mobility. The pET expression system was selected for use in the production of OtcR, with pET28a being the plasmid of choice. By using the pET system up to 50 % of the proteins present in the cell could be OtcR. The pET system uses the powerful T7 bacteriophage expression system (Studier and Moffatt, 1986) and pET28a encodes a fusion tag in the form of a poly-histidine tag (HIS-tag) that encodes 6 histidine residues, either at the N-terminus or C-terminus of the recombinant protein. This HIS-tag fusion allowed the use of nickel affinity chromatography for purification of the OtcR protein. *otcR* amplified from SR1 and SR55 were blunt-end cloned into SmaI-digested pUC19 (Yanisch-Perron et al, 1985) and sub-cloned into pET28a using NdeI and HindIII sites (as

described in Chapter 2.7.2). These plasmids were then used to transform *E. coli* BL21 (DE3) for production of OtcR and subsequent purification.

OmpR-Family transcriptional regulators, as well as many SARPs, are known to bind to repeat regions of DNA sequence within the promoters of their target genes (Wietzorrek & Bibb 1997; Arias *et al.,* 1999; Hennecke *et al.,* 2005; Narayanan *et al.,* 2012). Previous analysis of the promoter sequences of the genes *otcC, otcX, otcY1, otcY2-1* and *otcY2-2,* reveled tandem repeats (McDowall *et al.,* 1999), and suggests that a SARP might control OTC production. If these tandem repeats are associated with regulation via OtcR then these repeats would be present in all genes regulated by OtcR. Therefore, further examination of the upstream regions of the *otc* genes was undertaken.

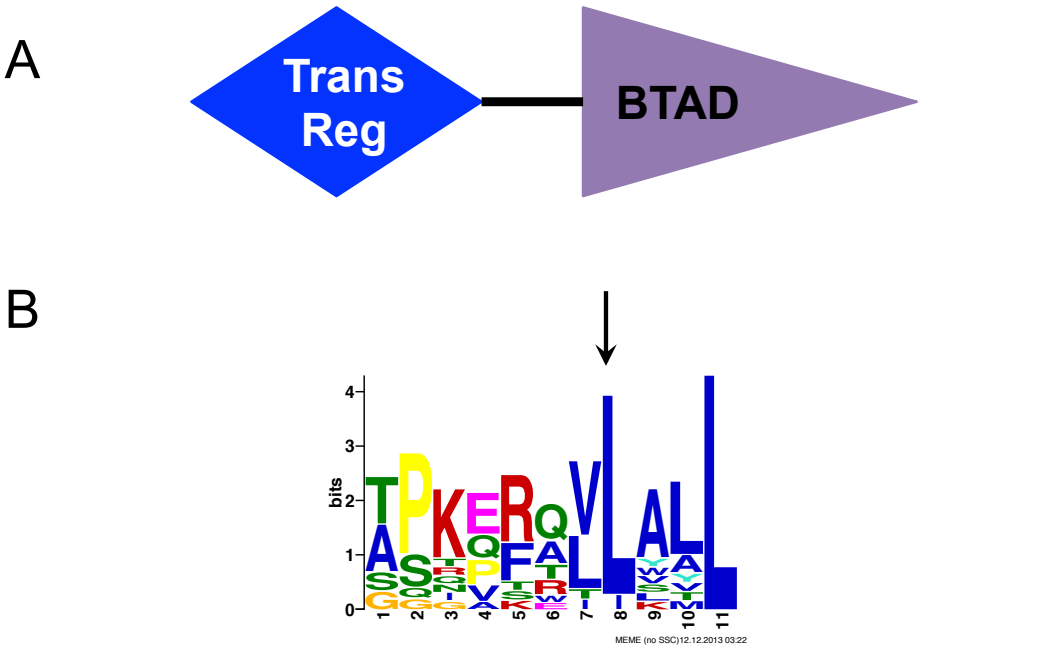
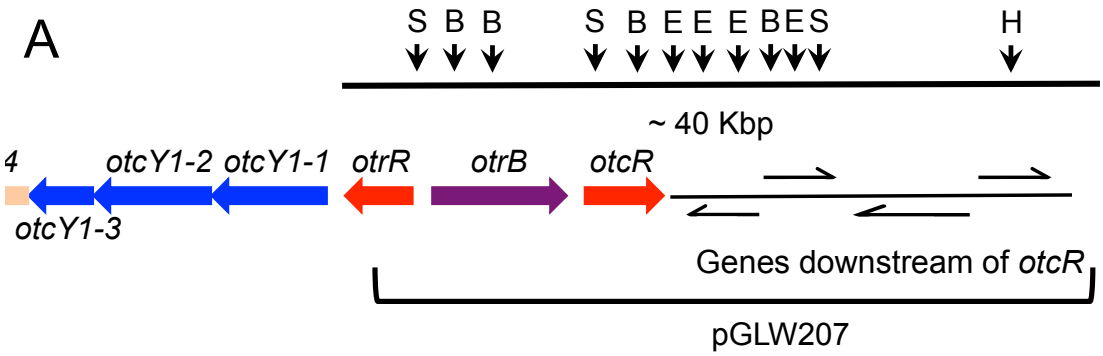
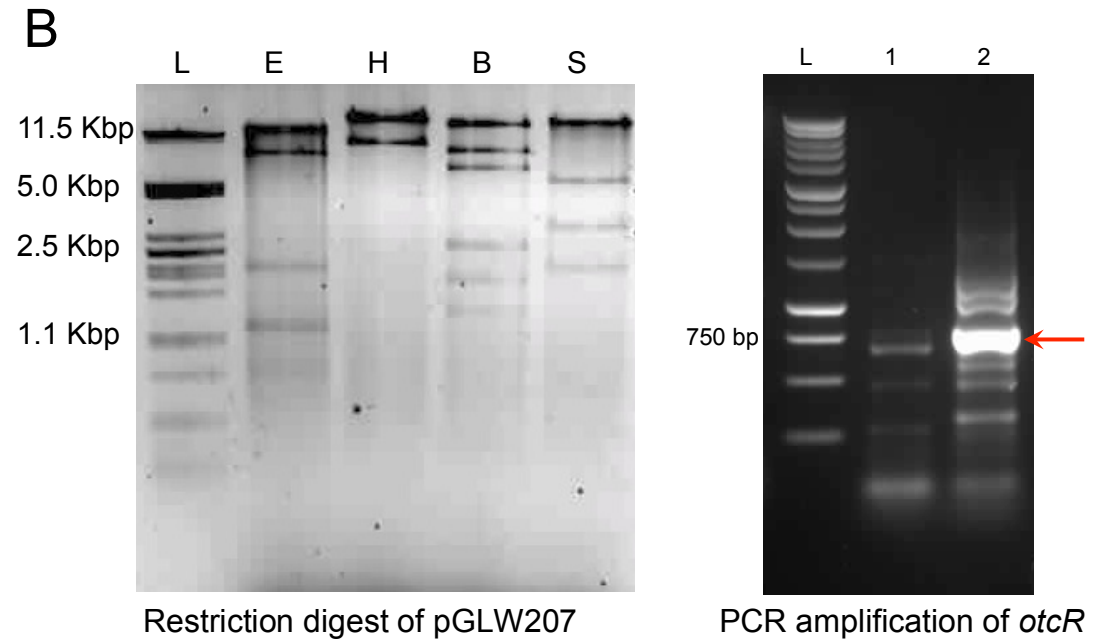


Figure 7.1 Bioinformatic analysis of OtcR. **A.** Domain architecture of OtcR. **B.** Trans-Reg domain identified using MEME. **C.** Amino acid alignment of N-terminal and C-terminal Trans-Reg domains. DnrI, *S peucetius*, ActII-ORF4, AfsR, *S. coelicolor* A3(2), PapR4, *S pristinaespiralis*, SrrY, *S. rochei*, SsfT1, *S. sp.* SF2575, SnorA, *S. nogalater*, NocR, *Nocardia uniformis*, EmbR, *Mycobacterium tuberculosis*, KdpE, *E. coli*, PhoP, *Bacillus subtilis*, CpxR, *Shigella flexneri*, OmpR, *Salmonella typhimurium*, ToxR, *Vibrio cholerae.*

7.2 Deletion of *otcR* from *S. rimosus* SR1 and SR55

To generate an *otcR* knockout strain using the REDirect method it was necessary to select a cosmid from the *S. rimosus* SR20 (M4018) library (Almutairi, 2007) in order to delete *otcR* from the cosmid with a selectable marker. The cosmid library available for strain SR20 has not been fully sequenced, however previous analysis by colony hybridisation (Almutairi, 2007) identified 5 cosmids containing regions of the OTC biosynthetic cluster. One of these cosmids, named pGLW207, contained the right arm of the biosynthetic cluster and within it *otcR* (Figure 7.2 A). This cosmid was selected for the REDirect protocol, which uses PCR-targeted mutagenesis, as not only was the *otcR* gene present but it was predicted that there was sufficient homologous regions flanking *otcR* to permit an allelic replacement of this gene with the selectable marker. Confirmation of the pGLW207 cosmid was achieved by restriction digest analysis, plus PCR amplification of the *otcR* gene (Figure 7.2 B). The apramycin resistance cassette (aac(3)IV) was chosen as the selectable marker for *otcR* disruption. The apramycin cassette was isolated from the plasmid pIJ773 by digesting with EcoRI and HindIII. The cassette was then PCR amplified using the forward primer otcR-KO-F and the reverse primer otcR-KO-R (Chapter 2.6, Table 2.6). At this point the apramycin cassette was used to transform *E.coli* BW25113 containing both the plasmid pIJ790 and the cosmid pGLW207, allowing λ RED-mediated gene replacement to take place. The resulting transformants were apramycin resistant, as well as kanamycin and carbenicillin resistant due to the original cosmid pGLW207. Confirmation of *otcR* disruption was established by PCR (Figure 7.2 C) and the cosmid was renamed pOtcR-KO. Once the suicide vector pOtcR-KO was generated introduction of this plasmid into the chromosome SR1 and SR55 was attempted, via conjugation and electroporation.





Restriction digest of pGLW207

PCR amplification of *otcR*

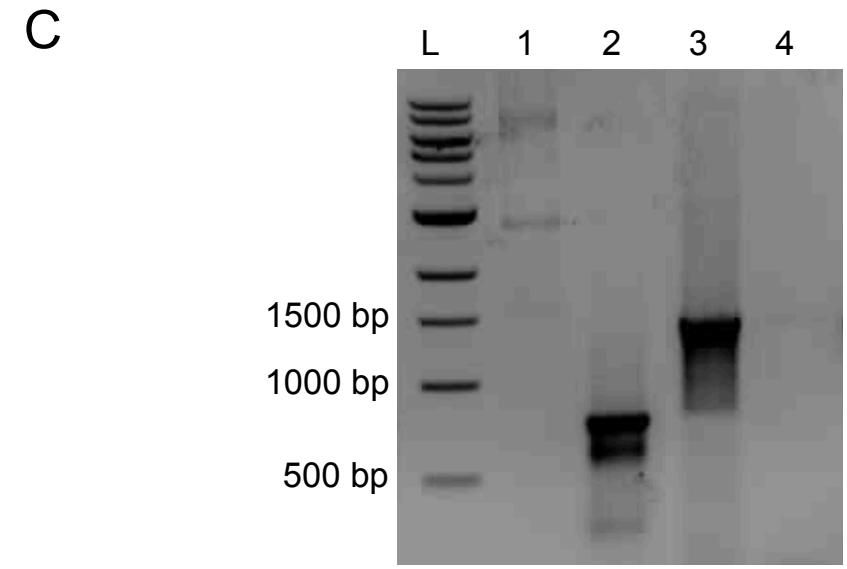


Figure 7.2 Isolation of cosmid pGLW207. **A.** Relative position of pGLW207 within the OTC gene cluster and *S. rimosus* SR1 genome, as well as restriction sites, as determined by previous analysis (Almutairi, 2007). As no sequence is available for the cosmid exact sites and sizes could not be determined. **B.** Confirmation of isolation of the cosmid pGLW207 by restriction digest and PCR. For the restriction digest confirmation gel (left) lane L = λPstI DNA ladder, lane E = EcoRI digest, lane H = HindIII digest, lane B = BamHI digest and lane S = SacI digest. For the PCR confirmation gel (right) lane L = 1 Kb Invitrogen DNA ladder, lane 1 = *S. rimosus* SR1 genomic DNA used as template and lane 2 = pGLW207 cosmid DNA used as template. **C.** PCR confirmation of *otcR* disruption and the generation of pOtcR-KO. Lane L = NEB 1 Kb DNA ladder, lane 1 = pOtcR-KO cosmid DNA with OtcR-FR primers, lane 2 = SR1 genomic DNA with OtcR-FR primers, 3 = pOtcR-KO cosmid DNA with apramycin (Apr-FR, Chapter 2.6) primers, 4 = SR1 genomic DNA with apramycin (Apr-FR, Chapter 2.6) primers.

Successful delivery of pOtcR-KO into the chromosome of *S. rimosus* and deletion of *otcR* should result in the recovery of apramycin resistant and kanamycin sensitive colonies following conjugation or electroporation of *S. rimosus* strains. *S. rimosus* colonies were obtained that were apramycin resistant and kanamycin sensitive. From these colonies genomic DNA was isolated in order to confirm the presence of the apramycin cassette and confirm disruption of *otcR* using PCR, however no positive results were obtained. As the presence of the apramycin cassette could not be confirmed and the *otcR* gene could be amplified, it was determined that these apramycin resistance colonies were false positives. Many attempts were made to obtain a true *otcR* knockout however none were successful.

7.3 Solubility optimisation and purification of OtcR via nickel affinity chromatography

As OtcR has never been over produced and purified before it was necessary to optimise this process. Initially pOtcR-SR1 (Chapter 2.7.2) was overexpressed in BL21 (DE3) at 30 °C and induced once the culture's optical density (OD) reached 0.4 using a final concentration of 0.1 M IPTG. In order to determine the best time after induction to harvest the cells for protein extraction, samples were taken at half hour time points from 2 hours until 4 hours. These samples were then separated on an SDS-PAGE gel and from examination of the gel, 3 hours after induction was deemed the best time for harvesting the cells (Figure 7.3). A high percentage of the protein present in the crude cell extracts was OtcR, therefore preparation of the samples for purification was undertaken. The cells were lysed by French pressure cell and the insoluble material removed by centrifugation. Upon examination of the resulting pellet and supernatant it became apparent that OtcR was insoluble, as most of the protein was present in the cell pellet rather than the cell lysate (Figure 7.3). It was therefore necessary to improve the solubility of the protein to increase efficiency of the purification process.

In order to improve the solubility of OtcR different detergents, reductants and salt concentrations were tested. Different detergents were tested first of all. Small-scale samples of OtcR were prepared and resuspended in resuspension buffer (100 mM Tris, 250 mM NaCl and 20 mM imidazole) supplemented with 1 % final concentrations of either SDS, sarkosyl, Tween-20 or Triton X-100 (Figure 7.4 A). From the results obtained it was determined that 1 % SDS had the most significant effect on OtcR solubility compared with the other detergents tested. Next different reductants were tested, this was achieved in a similar manner to the detergent examination but with the resuspension buffer supplemented with either 5 % 1-thioglycerol, 1 mM β-mecaptoethanol or 10 mM DTT (not shown) (Figure 7.4 B). The reductants were also tested in the presence of 1 % SDS to resolve if the reductants improved solubility better alone or in conjunction with the detergent (Figure 7.4 B). It was concluded that both 1-thioglycerol and β-mecaptoethanol were the best reductants for solubilising OtcR and in the presence of 1 % SDS improved solubility even further. It was decided to use the resuspension buffer supplemented with 1 % SDS and 1 mM β-mecaptoethanol for further protein purification experiments. This buffer was renamed soluble buffer.

At this point new protein preparations were made for purification using the soluble buffer. Initial attempts at purification via nickel affinity chromatography using the soluble buffer failed; no protein bound to the column. It was hypothesised that the presence of 1 % SDS within the soluble buffer was denaturing the proteins preventing binding, and therefore the crude sample was subjected to dialysis. The dialysis was performed in a stepwise manner, starting with dialysis buffer containing 0.5 % SDS reduced to 0.1 % SDS over an 8 hour period at room temperature, then reduced to 0 % SDS for an overnight incubation at 4 °C. After the dialysis purification by nickel chromatography was attempted again, however the same results were observed. It was then decided to attempt ion exchange chromatography. An anion exchanger was selected, as the theoretical isoelectric point (pI) of OtcR was pH 6.5, and so a pH of 8 was chosen, as the overall charge of the

OtcR protein would be negative at this pH. However once again no proteins could be eluted from the column.

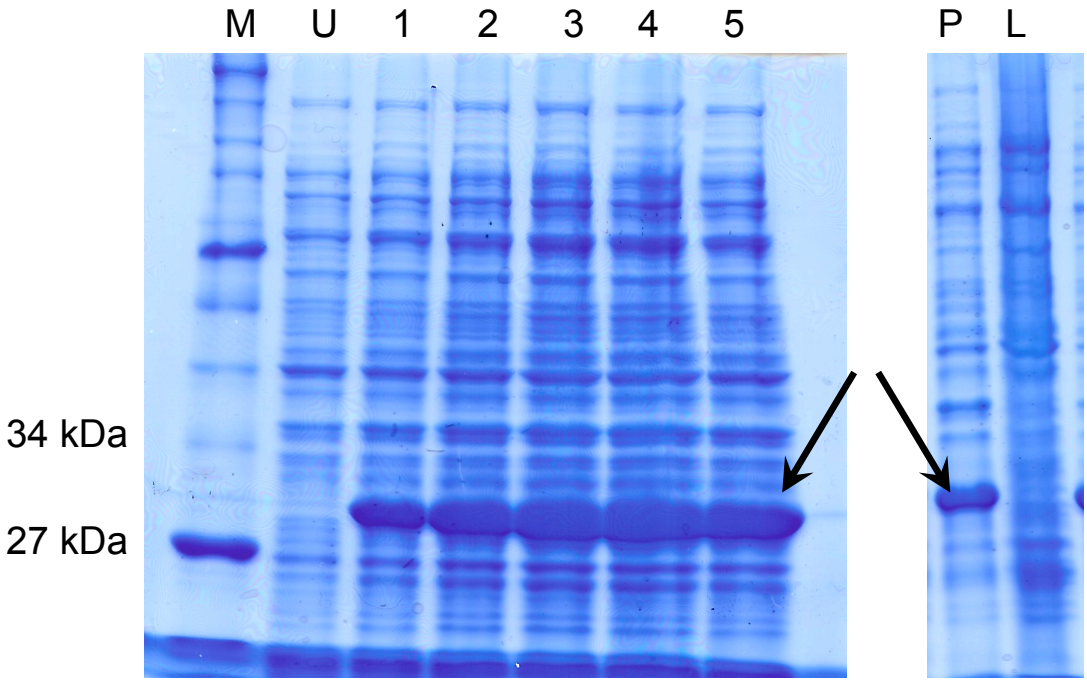


Figure 7.3 SDS-PAGE of OtcR overexpression at 30 °C. Lane M contains the molecular weight marker. Lane U contains crude cell extract of uninduced sample, lane 1 is crude cell extract 2 hours after induction with lane 2-5 samples taken at 30 minute intervals up to a total of 4 hours. Lane P is insoluble extract present in the pellet and lane L is soluble extract present in the lysate. The 29.7 kDa histagged OtcR is indicated by the black arrows.

As neither method of purification was successful with the soluble buffer it was decided to use the original resuspension buffer without supplementation. Preliminary purification using nickel affinity chromatography with the resuspension buffer showed a high degree of binding by non-specific proteins. As the total percentage of OtcR present in the crude sample was low, it was necessary to increase the specificity of binding. To improve specificity of protein binding, specificity tests were undertaken using different pH conditions and with ammonium thiocyanate. pH is known to affect the solubility, activity and stability of proteins (Skoulakis & Goodfellow 2003; Talley & Alexov 2010) and therefore establishing the best pH was essential not only for purification but also for subsequent protein studies. As mentioned previously, the theoretical pI of OtcR was pH 6.5 and so a pH range of 7.5 – 8.5 was chosen, as this range was around neutral and close to the pI. Ammonium thiocyanate is a chaotrope and therefore can destablise proteins and consequently aid solubility (Lawal 2006; Leibly *et al.,* 2012). It was for this reason that this compound was selected to improve solubility of OtcR and aid binding specificity. Samples were prepared for nickel affinity chromatography as before using the resuspension buffer at each pH to be tested and supplemented with 50 mM ammonium thiocyanate where appropriate. Fractions were eluted by an elution gradient of Buffer B (resuspension buffer with 1 M imidazole) from 0 – 100 % over 20 minutes at a rate of 1 mL per minute. The obtained fractions were then analysed by SDS-PAGE, as shown in Figure 7.5. From these results it was determined that pH 8 (Figure 7.5 B) was most effective at increasing binding specificity of OtcR and that ammonium thiocyanate had little or no effect and it was decided not to use ammonium thiocyanate in further tests. Although binding of non-specific proteins was lower at pH 8, there was still necessary to subject the sample to further purification.

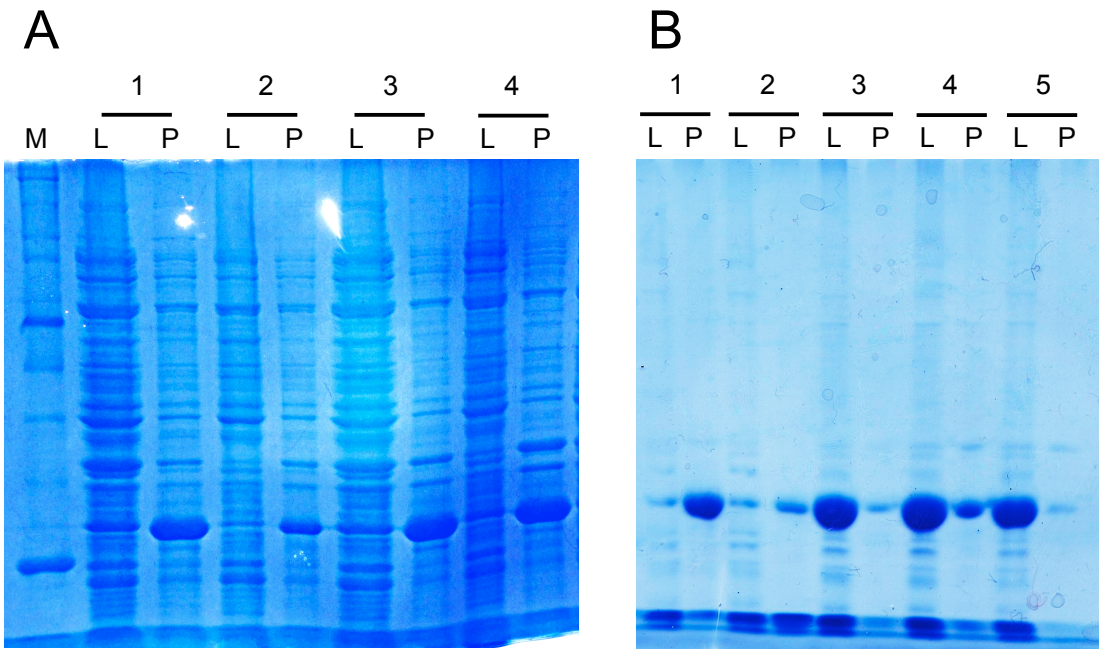


Figure 7.4 SDS-PAGE gel of OtcR solubility tests. Lane L is crude extract present in the lysate (supernatant), lane P is crude extract present in the pellet. Lane M is the molecular weight marker. **A.** SDS-PAGE gel of different detergents and the effect on OtcR solubility. 1. 1 % SDS 2. 1 % sarkosyl 3. 1 % tween 20 4. 1 % triton X-100 **B.** SDS-PAGE gel of different reductants and reductants with detergents and the effect on OtcR solubility. 1. 5 % 1-thioglycerol 2. 1 mM β-mecaptoethanol 3. 1 %1-thioglycerol and SDS 4. 1 % SDS and 10 mM DTT 5. β-mecaptoethanol and 1 % SDS.

Imidazole is used to elute any bound protein as it dissociates these proteins from the nickel column by competing for binding with the nickel ions. For this reason imidazole is used to reduce non-specific binding by out competing proteins with lower affinity for nickel, when added in high enough concentrations to the column wash buffer. However if imidazole concentrations are too high then binding of the protein of interest may be lower or abolished. It was decided to use 50 mM imidazole in the resuspension buffer, rather than the previous 20 mM, to determine if non-specific binding could be lowered, whilst still maintaining sufficient binding by OtcR. Once again nickel affinity chromatography was undertaken and eluted fractions were analysed by SDS-PAGE. Increasing the imidazole concentration in the resuspension buffer was successful in reducing non-specific binding as a single protein band was obtained (Figure 7.6 B) at around 30 kDa. The resuspension buffer supplemented with 50 mM imidazole was renamed Buffer A (100 mM Tris, 250 mM NaCl, 50 mM imidazole, pH 8) and used for all subsequent attempts to purify OtcR. From the elution profile seen in Figure 7.6 A it was determined that OtcR is eluted at around 12 % Buffer B.

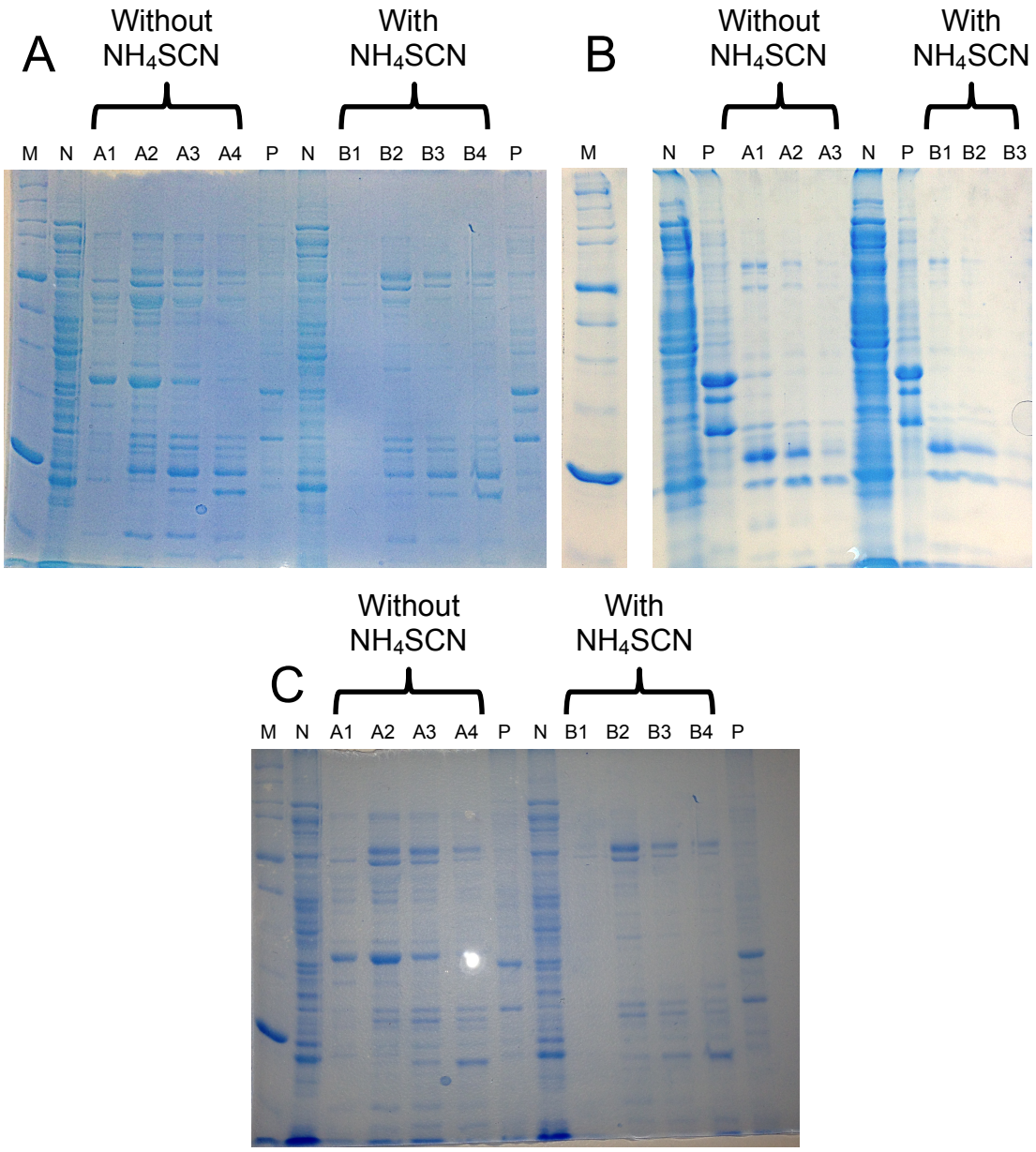
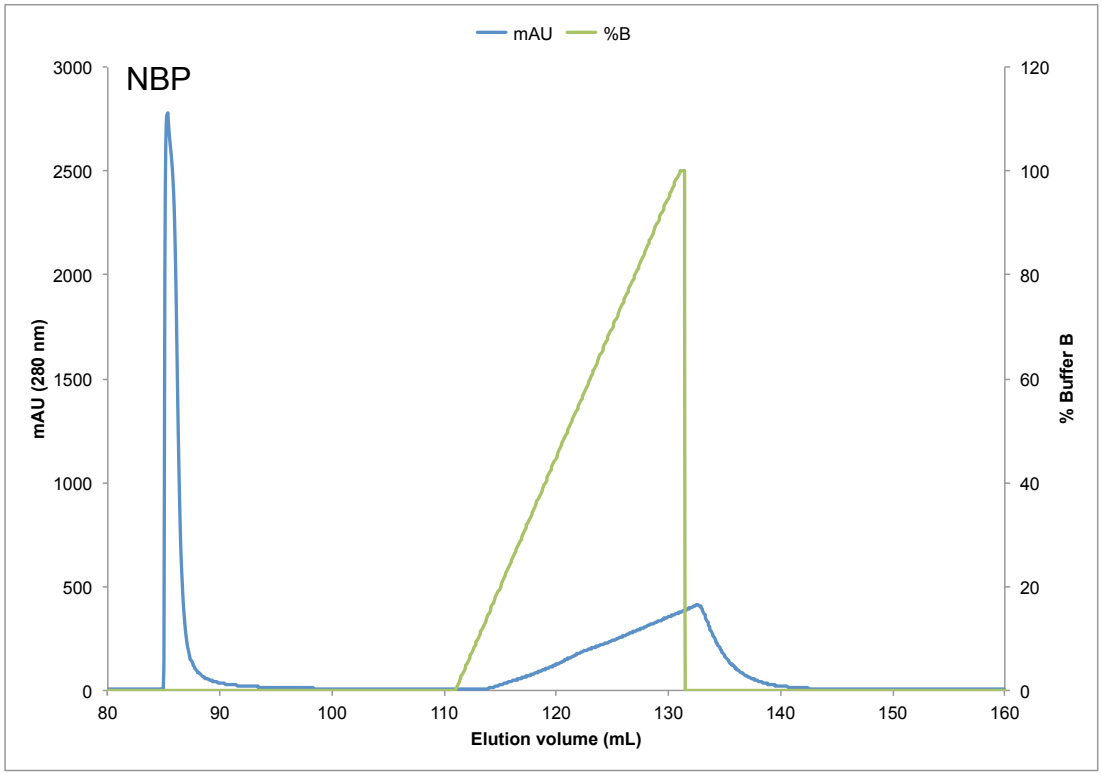


Figure 7.5 Binding specificity of OtcR using nickel affinity chromatography. Binding specificity of OtcR was tested using different pH conditions and in the presence or absence of 50 mM ammonium thiocyanate (NH_4SCN). Lane M is the molecular weight marker, lane N is protein that did not bind to the nickel column (flow through) and lane P is pellet. Lane A are eluted fractions without ammonium thiocyanate and lane B are eluted fractions with ammonium thiocyanate. **A.** pH 8.5 with and without ammonium thiocyanate. **B.** pH 8.0 with and without ammonium thiocyanate. **C.** pH 7.5 with and without ammonium thiocyanate.

A



B

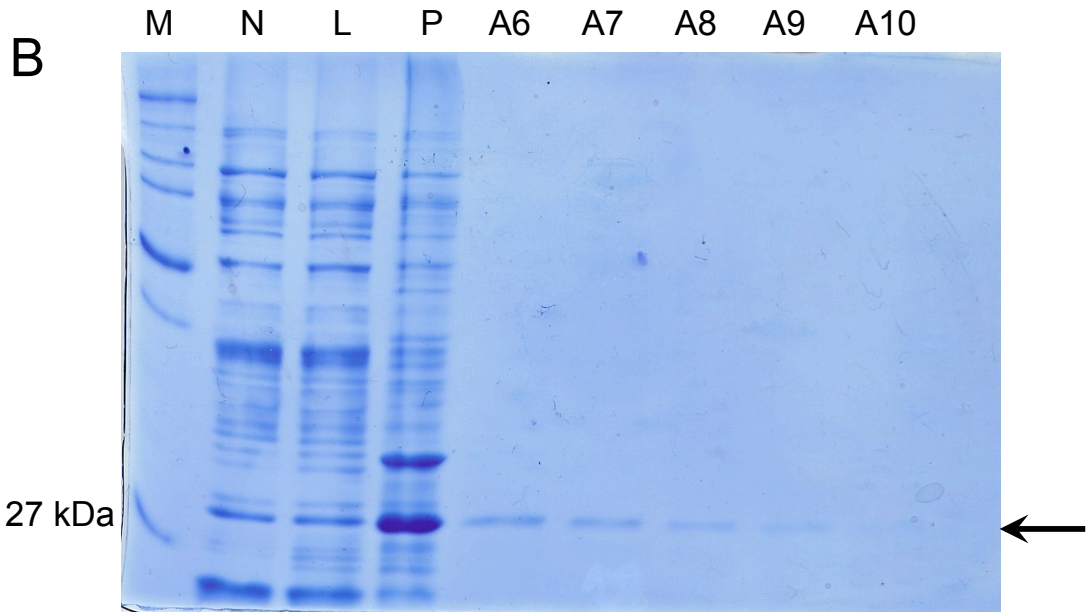


Figure 7.6 Nickel affinity chromatography of OtcR. **A.** Elution profile of OtcR from crude extract. **B.** SDS-PAGE gel of bound protein eluted by increasing percentage of Buffer B. A6 – A10 represents the fractions loaded onto the gel. OtcR is highlighted by the black arrow.

7.4 Investigation of the DNA-binding properties of OtcR

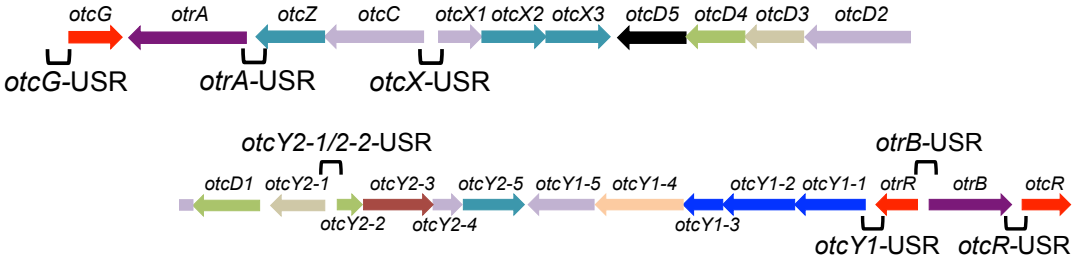
Once OtcR from both a SR1 and SR55 background was purified band shift assays could be performed. As mentioned previously tandem repeats within the promoter regions of some otc related genes were identified and it was hypothesised that these repeats may play a part in regulation of OTC production at the level of transcription (McDowall *et al.,* 1999). If *otc* biosynthetic genes are regulated by OtcR in this way, then those genes with direct tandem repeats within their promoters might be the target of OtcR. Examination of the upstream regions (USR) of the genes within the OTC cluster (Figure 7.7 A) did not reveal any new tandem repeats; this suggests that the genes associated with resistance and regulation are controlled by a different means. As noted by McDowall *et al* the tandem repeats within the promoters of *otcC, otcX, otcY1, otcY2-1* and *otcY2-2* share a high degree of similarity with repeats identified in the promoters of ACT and daunorubicin/doxorubicin (DNR/DXR) biosynthesis genes (Tang *et al.,* 1996; Wietzorrek & Bibb 1997). Since then direct repeats have been discovered within the USR of genes belonging to the fredericamycin (FDM) and valanimycin (VLM) biosynthetic clusters, whose regulation is controlled by the SARPs FdmR1 and VlmI, respectively (Chen, *et al.,* 2008b; Garg & Parry 2010). The repeats of ACT, DNR/DXR, FDM and VAL are heptameric, spaced either 4 or 15 nucleotides apart, occur either in pairs or as a trio and are located 8 nucleotides upstream of the -10 region (Tang *et al.,* 1996; Wietzorrek & Bibb 1997; Chen *et al.,* 2008b; Garg & Parry 2010). For this reason the OTC tandem repeats were examined as 7 nucleotides long with 4 nucleotides in between, rather than 9 nucleotides long and 2 nucleotides apart, as they were previously documented (McDowall *et al.,* 1999). When the repeats associated with OTC, ACT, DNR/DXR, FDM and VLM were examined together a consensus sequence of TCGA[G/A][C/G] could be readily identified (Figure 7.7 B) but unlike the other antibiotic clusters only one promoter (*otcY1*) associated with OTC had 3 direct repeats. Although one repeat lies on the -35 region and another between the -35 and -10

regions, previous predictions (Wietzorrek & Bibb 1997) and DNA modeling (Figure 7.7 C) suggest the -10 region lies on the opposite side of the DNA strand to the repeats. This would ensure both the SARP and RNA polymerase (RNAP) have access to the DNA and as isomerisation occurs at the -10 region the template strand would be available to the active site of RNAP implying transcription would not be adversely effected (Browning & Busby 2004; Ghosh *et al.,* 2010). It has been shown for DnrI, the SARP regulator of DNR/DXR, that changing the nucleotide spacing between the repeats is detrimental for DnrI binding (Sheldon *et al.,* 2002), so it is likely the SARP binds first and helps guide the RNAP to the promoter (Tanaka *et al.,* 2007). Together these observations would suggest a positive role for OtcR in the regulation of the OTC biosynthesis genes.

Once potential targets of OtcR were identified EMSA could be performed. Initial protein binding assays were performed with 500 nM of OtcR however none were successful and each DNA probe did not shift (Figure 7.8). The USR of *otrB* was used as a negative control and the USR of *otcR* was tested to ensure the prediction of no likely SARP binding regions was correct. As the Trans-Reg domain was within the N-terminal it is possible that the HIS-tag present prevented DNA binding due to misfolding of the OtcR protein. Cleavage of the HIS-tag domain would prevent misfolding and overcome such an issue, however a sufficient quantity of OtcR was never obtained to allow for HIS-tag cleavage and DNA binding assays to be performed. Alternatively a cofactor might be required for DNA binding activity and therefore without, OtcR could not bind its targets. For both these reasons a different method was employed.

In order to overcome the issue of the N-terminal HIS-tag and/or the requirement of a cofactor it was decided to use lysate from *S. rimosus* SR1, SR50S and SR55 when performing DNA binding assays. Due to the use of crude lysate specific time points were selected of 18 and 46 hours after inoculation, as these time points were predicted to occur before and at induction, respectively, of *otcC* and *otcX* expression (McDowall *et al.,* 1999).

A



B

otcC	5'-GCTGGACGG-N_4-TTTAATGCC-3'	*dpsEF*	5'-CGTCAAGCG-N_4-TCGAGCGCG-3'
otcX	5'-GCTCGAAGT-N_4-TCGAAGCGG-3'	*actI-ORF1*	5'-CTTCGAGCC-N_4-TCGAGCCAC-3'
otcY2-1	5'-GGTCGGTCC-N_4-TCGAACGGG-3'	*actIII*	5'-CCTTGACCG-N_4-TCGAGCCGG-3'
otcY2-2	5'-GCTCGAAGA-N_4-TCGAAGACC-3'	*fdmC*	5'-CTTCGAGCC-N_4-TCGAGGCGG-3'
*dnrG**	5'-CTTCCAGAG-N_4-TCGAGCGGC-3'	*vlmH*	5'-TGTCAGGTG-N_4-TCAGGGAGG-3'

otcY1 5'-CGTCGAATC-N_4-TCCAAAC-N_4-TTGAGGTCC-3'

C

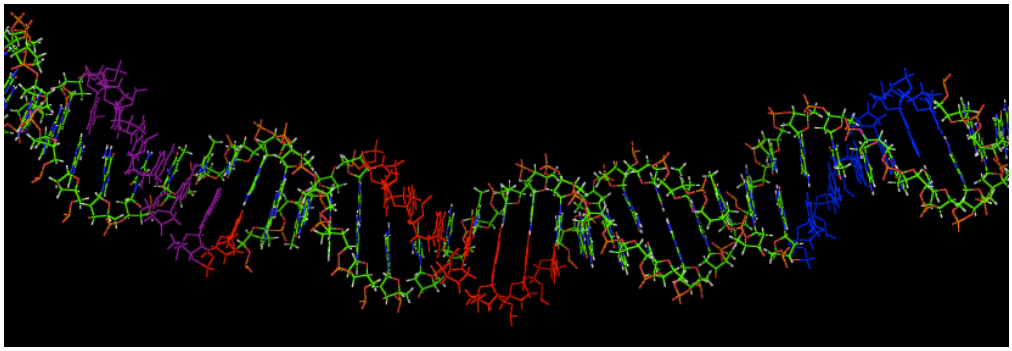


Figure 7.7 Analysis of tandem repeats in the USR of *otc* biosynthetic cluster genes. **A.** Relative position of the USR regions examined for tandem repeats and used for DNA probes in band shift assays. **B.** Comparison of putative and verified SARP binding sites present in the OTC, ACT, DNR/DXR, FDM and VLM biosynthetic gene clusters. *dnrG** represents *dnrG-dpsABCD* **C.** 3D DNA model of the promoter region of *otcX* generated using Model.it (Vlahovicek & Pongor 2000). Coloured segments of DNA correspond to: the -10 region = blue, -35 region = purple and the tandem repeats = red.

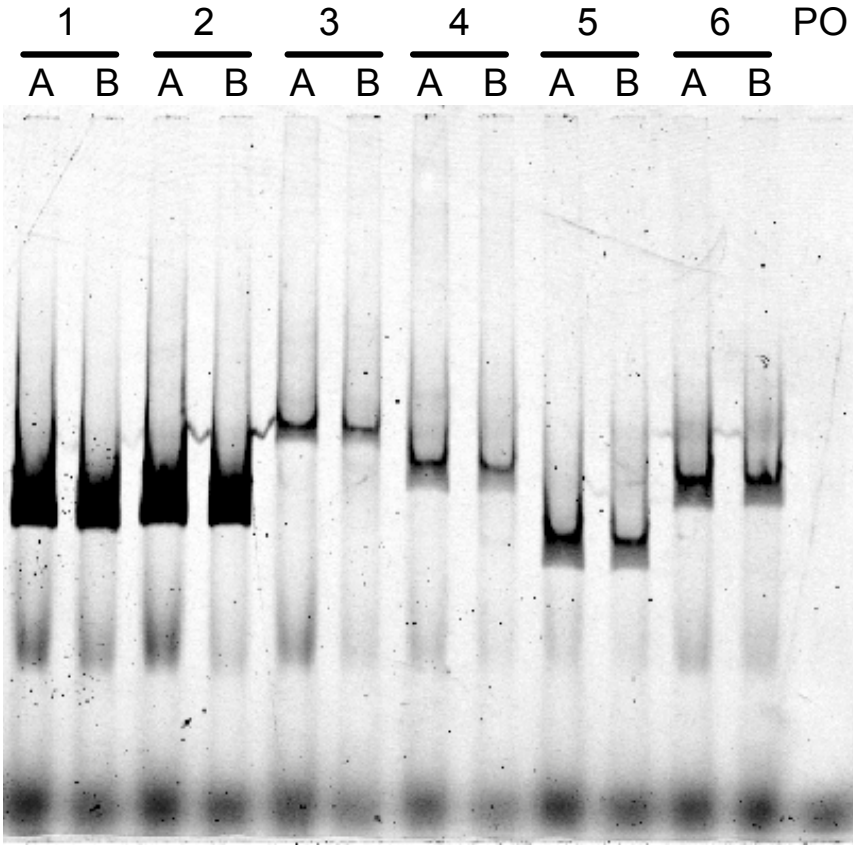


Figure 7.8 DNA binding assay with purified OtcR-SR1. Lane PO is protein only. Lane A for each sample is DNA only and lane B for each sample is DNA and protein. Total target DNA for each test was 3 ng, with the exception of *otcX* and *otcY2-2* of which 100 ng was used. For all samples containing protein 500 nM of OtcR-SR1 was used. Gene upstream regions are 1. *otcX* 2. *otcY2-2* 3. *otcY2-1* 4. *otcY1* 5. *otrB* 6. *otcR*. Gene upstream regions were PCR amplified using M13 (-21) Cy5 fluorophore-labeled primers from pUC19 inserts.

EMSA were attempted using the crude lysate from SR1, SR50S and SR55 from the 18 and 46 hour time points for the following USR; *otcR, otrB/R, otcY1, otcY2-1, otcY2-2, otcX/C, otrA* and *otcG*. For each EMSA *devA*-USR from *S. coelicolor* A3(2) was used as a positive control for 18 hour samples and as a negative control for 46 hour samples, as *devA* has been shown to be expressed up until 24 hours of growth (Hoskisson *et al.,* 2006; Clark & Hoskisson 2011). However at both time points for all samples *devA*-USR shifted, and further to this no observable difference was detected between SR1, SR50S and SR55 for *otc* cluster specific USR probes (Figure 7.9). This suggests that either unrelated protein binding to the *otc* cluster USR probes or that the mechanism of regulation of these genes lies outside the OTC biosynthetic cluster.

It was decided to also use crude lysate from SR1, SR50S (15883S) and SR55 to identify any proteins that bound to the USR of *otcR* by exploiting the strong bond formed between biotin and streptavidin, by labelling the USR of *otcR* with biotin and binding it to streptavidin beads (Holmberg *et al.,* 2005). Biotin-labeled otcR-USR-F primer was combined with otcR-USR-R to amplify the *otcR* USR resulting in a biotin-labeled *otcR*-USR probe. The biotin labeled *otcR*-USR probe was bound to 1.5 mg of Dynabeads® M-270 Streptavidin magnetic beads, enabling immobilisation of the probe to a Dynamag™-2 magnet (see Chapter 2.11 for full details). Two sets of Dynabeads® M-270 Streptavidin magnetic beads were prepared for each *S. rimosus* strain and for both time points, one with the *otcR*-USR probe and another without, resulting in twelve preparations of Dynabeads® M-270 Streptavidin magnetic beads. At this point crude *S. rimosus* lysate was washed over the immobilised probe, allowing for the identification of bound and unbound protein profiles for each strain and time point (Figure 7.10). From the protein profiles two SDS-PAGE gel bands were selected for identification by via the Mascot server using the results from tandem mass spectrometry, after trypsin digesting the proteins. These bands were selected due to significant differences between either time points or other strain's protein profiles.

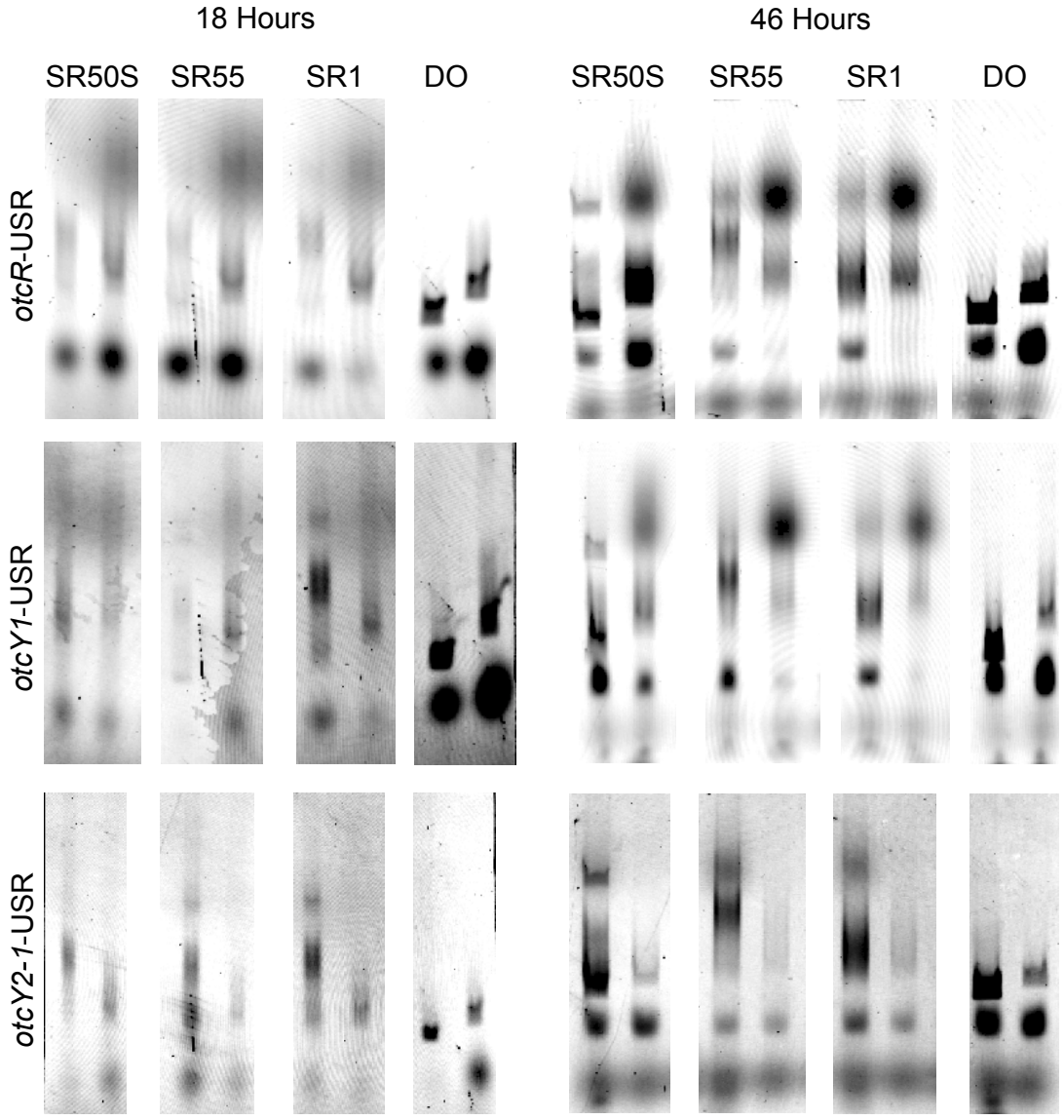


Figure 7.9 Selected EMSA from crude *S. rimosus* lysate. The first lane of each band shift is *devA*-USR from *S. coelicolor* A3(2), the second lane of each band shift is the USR noted to the left of the figure. Lane DO is DNA probe only. Crude lysate used was 3 µL from crude *Streptomyces* cell lysate (Chapter 2.9). Each sample contained around 4 ng of DNA target. Gene USR were PCR amplified using M13 (-21) cyanine 5 (Cy5) fluorophore-labeled primers from pUC19 inserts.

Unfortunately the protein profiles for SR1 and SR55 did not have any significant difference to SR50S at around 28 kDa (Figure 7.10), the molecular weight of OtcR. This suggests that OtcR was not bound to the *otcR*-USR probe under these conditions and therefore not pulled down. The band predicted to be OtcR overexpressed in *E. coli* was also subjected to Mascot analysis in order to confirm the protein was OtcR. Mass spectrometry and Mascot analysis was carried out by Prof. Andrew Pitt at the University of Aston.

Mascot analysis identified OtcR-OE as a SARP family transcriptional regulator (GI: 490078890), with a statistically significant score of 1029. Correct identification of OtcR validates the method and confirms successful overexpression of the *otcR* gene in *E. coli*. Of the bands selected for Mascot analysis SR1-18 was the most intriguing in relation to OTC biosynthesis. Sample SR1-18 was selected due to its presence in the protein profiles of both SR1 and SR55 at the 18 hour time point but absent from the 46 hour time point (Figure 7.10), this suggested that it was a potential regulator of *otcR*. It is possible this band was also present in the 18 hour protein profile of SR50S, however due to the poor quality of Coomassie blue staining the same region was barely visible. Mascot analysis identified a multifunctional aminopeptidase A (GI: 490085040), with a statistically significant Mascot score of 9665. This protein was subsequently identified as SRIM2039, a predicted cytosol aminopeptidase protein, and a homologue of SCO2179. SCO2179 knockout mutants of *S. coelicolor* A3(2) have increased sporulation and actinorhodin production (Song *et al.,* 2013), suggesting a negative role in these processes.

As OTC is first detected around 30 hours of growth (McDowall *et al.,* 1999; Yu, *et al.,* 2012a) and as SRIM2039 is only present at the 18 hour time point, this suggests SRIM2039 is a repressor of *otcR* and plays a similar role to that of SCO2179 in *S. coelicolor* A3(2). SR50S-46 was selected for Mascot analysis as it was only present in the protein profile of SR50S at 46 hours. SR50S-46 was in both the *otcR*-USR probe and beads only samples, which implies it was binding to the beads rather than *otcR*-USR. However as

there was SR50S-46 was the only band present at that size in all the samples, it was decided this was a unique feature of SR50S and possibly related to the absence of the OTC cluster in this strain. SR50S-46 was identified as a 30S ribosomal protein S4 (GI: 490082640) with a Mascot score of 1396, which is statistically significant. Therefore SR50S-46 is SRIM1195, a 30S ribosomal protein S4 or RspD. The mode of action of OTC is to prevent protein synthesis by binding reversibly to the 30S ribosomal subunit (Doyle *et al.,* 1991; Chopra & Roberts 2001; Petković *et al.,* 2006), and one mechanism of resistance is to block this translational arrest by protecting the ribosome, in *S. rimosus* this function is carried out by OtrA (Butler *et al.,* 1989; Doyle *et al.,* 1991). As SR50S is lacking the *otc* cluster any modification of the 30S ribosomal subunit, due to OTC or OtrA, would not occur and it may be due to this that SRIM1195 from SR50S is able to bind to the beads, suggesting SRIM1195 may be a target for OTC or it plays a role in resistance.

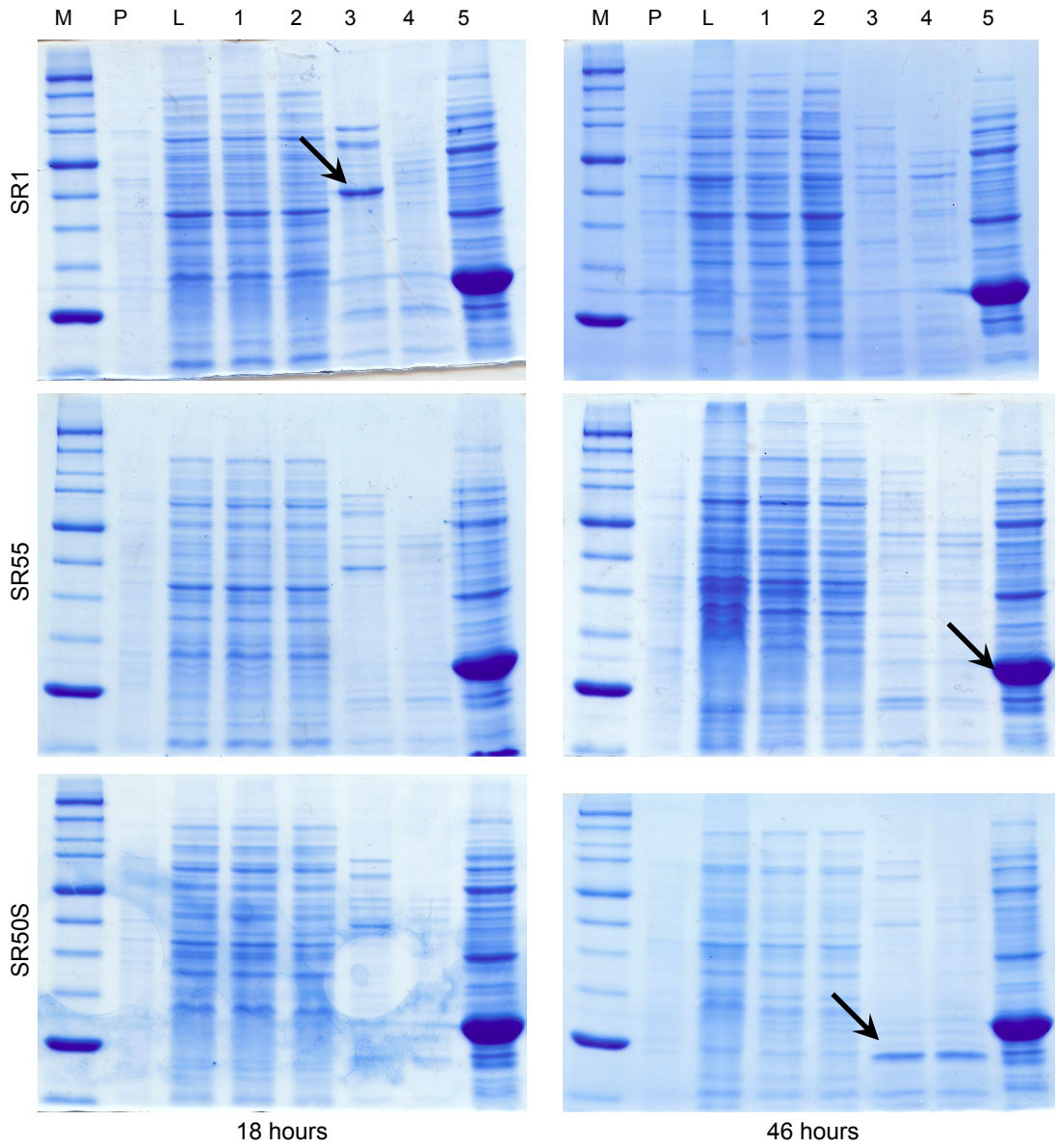


Figure 7.10 Whole cell extract protein profiles of SR1, SR55 and SR50S at 18 and 46 hour growth points. Lane M is the molecular weight marker, lane P is the pellet, lane L is lysate, lane 1 is non-bound protein washed over streptavidin beads with biotin labeled *otcR* USR, lane 2 is non-bound protein washed over streptavidin beads only, lane 3 is bound protein washed over streptavidin beads with biotin labeled *otcR* USR region, lane 4 is bound protein washed over streptavidin beads only and lane 5 is OtcR overexpressed in *E. coli*. Bands selected for Mascot analysis are indicated by black arrows and named from top to bottom: SR1-18, OtcR-OE and SR50S-46.

7.5 Discussion

Many antibiotic biosynthetic gene clusters have a SARP regulator associated (Wietzorrek & Bibb 1997; Arias *et al.,* 1999; Sheldon *et al.,* 2002; Liu *et al.,* 2013), however until now no SARP regulator had been determined for the *otc* cluster, therefore it was significant to identify the potential SARP, *otcR*. Two other regulators associated with the *otc* cluster are known, *otrR* and *otcG*, thought to be a regulator of *otrB* and a positive regulator of OTC production, respectively (Petković *et al.,* 2006; Lenik 2009). However it should be noted that *otcG* is not likely to be the sole regulator of OTC production as deletion of the gene did not abolish OTC production (Lenik 2009). Therefore, it was desirable to establish the potential role *otcR* played in OTC production.

Attempts to disrupt *otcR* by employing the REDirect method were unsuccessful most likely due to the antibiotic selective marker, apramycin. Both conjugation and electroporation were used to attempt to deliver the disruption cosmid and in both cases false positive colonies were obtained. These colonies grew on apramycin but the knockout cassette could not be detected. In addition, SR1 colonies, that lacked the disruption cosmid, grew on the negative control plates even when apramycin concentrations were increased to 600 μg/mL. For this reason it is recommended that use a different antibiotic resistance marker, possibly thiostrepton is used to disrupt the *otcR* gene in future. Overexpression of OtcR in *E.coli* was relatively simple with 3 hours growth after induction with IPTG at 30 °C being sufficient to produce large quantities of OtcR. However it became apparent during initial attempts to purify OtcR that the majority of the protein was insoluble. Although solubility of the protein was improved another problem had to be overcome, the soluble OtcR protein no longer bound to the nickel column. One potential reason for non-binding was the presence of 1 % SDS in the resuspension buffer which has the potential to cause protein denaturation (Michaux *et al.,* 2008); this would also explain why no proteins appeared to bind to the column. To address this issue the cell lysate was dialysed in

order to refold the proteins and remove SDS from the sample (Umetsu 2003). However the dialysed sample failed to bind the column suggesting the proteins within the sample had not refolded. It was decided at this point to attempt purification without improving the solubility of OtcR.

Once pure samples of OtcR were obtained protein-DNA binding EMSA were undertaken. However, OtcR did not bind to the USR of any of the *otc* cluster genes tested. As the HIS-tag was encoded at the N-terminal of the OtcR protein and this is the location of the Trans-reg and potential DNA binding domain, it is likely the HIS-tag interfered with this domain and brought about misfolding of the protein (Halliwell *et al.,* 2001) which could explain why OtcR did not bind to any of the USR. The HIS-tag could be cleaved from the protein in order to resolve this problem, however two alternative methods might also be employed. The HIS-tag could have been fused to the C-terminal as the domain of the protein as this region is predicted to be a BTAD and not believed to bind DNA (Alderwick *et al.,* 2006; Tanaka *et al.,* 2007). It should be noted the BTAD of DnrI is required for DNA binding, by the Trans-Reg domain, and DNR/DXR production in *S. peucetius* (Sheldon *et al.,* 2002). Another solution would have been to use an overexpression plasmid that did not encode a fusion tag, such as pET21a, this would allow for purification of OtcR in its native form and require no further processing, such as removal of a fusion tag. As OtcR was predicted to have DNA binding capabilities heparin agarose affinity chromatography was used as DNA binding proteins are known to bind to heparin (Nyborg *et al.,* 1988; Marton *et al.,* 1989). Heparin agarose chromatography works in a similar manner to cation exchange due to the high level of anionic sulphate it contains, and although OtcR could be isolated using this method the samples were not pure or of sufficiently high enough quantity to follow up with another form of chromatography. As such the only OtcR available for the band shift assays was N-terminally HIS-tagged. It is also possible OtcR does not bind to any of the USR of the *otc* genes, this however seems unlikely due to the tandem repeats present within the *otc* genes, suggesting regulation at a transcriptional level by a SARP. Given the high similarity OtcR has to SARP

regulators and its proximity to the *otc* cluster, it is the most likely candidate for this form of regulation.

As attempts to analyse the DNA-binding capabilities of OtcR were unsuccessful an alternative method was used in the hopes of avoiding the pitfalls of the previous method. Use of the biotin-streptavidin method limited the target DNA being tested to the USR of *otcR*, nevertheless this was useful for identifying any protein that binds this region. However if OtcR does not bind its own USR region then the experiment would fail to prove OtcR has DNA-binding capabilities. From the observed differences between the protein profiles of the biotin-streptavidin experiment for SR1, SR50S and SR55 at 18 and 46 hour time points (Figure 7.10) two bands were selected for further examination, SR1-18 and SR50S-46. Of the two proteins identified SRIM2039, from the SR1-18 band, was predicted to play a role in the regulation of *otcR*, as it is a homologue of SCO2179 a protein thought to be involved in sporulation and actinorhodin production in *S. coelicolor* A3(2) (Song *et al.,* 2013). Further to this no protein band was detected at the same position in the 46 hour samples, suggesting the role SRIM2039 plays is growth dependent. As both OTC and ACT are produced at around the same time during growth (McDowall *et al.,* 1999; Hesketh *et al.,* 2001; Bruheim *et al.,* 2002; Lu *et al.,* 2007; Yu *et al.,* 2012b; Yu *et al.,* 2012a), it is possible that *otcR* has the same expression pattern as *actII*-ORF4, and therefore *otcR* would be expressed after 20 hours (Huang 2001; Uguru *et al.,* 2005). However, this would need to be confirmed by gene expression analysis of *otcR*. Combined with the observation that SRIM2039 only binds the *otcR* USR region at 18 and not 46 hours then it seems likely SRIM2039 negatively regulates expression of *otcR* until at least 20 hours. It is also worth noting that in a similar biotin-streptavidin experiment carried out with *S. coelicolor* A3(2) after 2 days of growth, using the promoter region of *actII-ORF4* as the DNA target, SCO2179, the SRIM2039 homologue, was not detected (Park *et al.,* 2009). This further supports the hypothesis that SRIM2039 plays a repressive role in the expression of pathway specific regulators of antibiotic biosynthetic clusters in a growth dependent manner. There are homologues

of SRIM2039 in all the *Streptomyces* species analysed in Chapter 3, as well as other *Streptomyces* species (Song *et al.,* 2013). Given how ubiquitous SRIM2039 is, it is possible this protein is a common control mechanism of antibiotic production during exponential growth in *Streptomyces* species.

SR50S-46 was identified as the 30S ribosomal protein S4 or RpsD, encoded by *SRIM1195*. The 30S subunit requires the association of 20 ribosomal proteins to the 16S ribosomal RNA (rRNA) for its assembly (Nowotny & Nierhaus 1988; Bellur & Woodson 2009; Kim *et al.,* 2014). The S4 ribosomal protein is central to this process as it is one of the first recruited to the rRNA, therefore initiating the ribosome assembly process, it controls its own and other ribosomal proteins translation and induces conformational rearrangement of the 16S rRNA (Nowotny & Nierhaus 1988; Powers & Noller 1995; Bellur & Woodson 2009; Kim *et al.,* 2014). S4 folds a 5 way junction (5WJ) in the 16S rRNA 5' domain which gathers 30S ribosome assembly, and binding to the 16S in this way stablises the 5WJ and a pseudoknot between helix (h) 18 of 16S rRNA and its internal loop (Bellur & Woodson 2009; Mayerle & Woodson 2013; Kim *et al.,* 2014). The h18 loop must fold correctly in order for protein synthesis to occur as the h18 loop interacts with transfer RNA (tRNA) in the mature ribosome (Kim *et al.,* 2014). Correct folding of the h18 loop only occurs after S4 binds the 16S rRNA (Powers & Noller 1995) and recently it has been shown that the folded h18 stabilises the docked h3, without this 30S assembly is impaired (H. Kim *et al.,* 2014). The nature of the h3 and h18 loop could also determine the order in which other proteins are recruited, as it has been suggest S12 could be rejected from binding to the complex until the S16 protein has been bound due to the folded state of h18 and the docked state of h3 (Kim *et al.,* 2014). Taken together these studies demonstrate the importance of S4 in the formation of the 30S ribosome. It has recently been shown that the ribosome protection protein TetO from *Campylobacter jejuni* interacts with h18, resulting in a twisted conformation for the helix which would dislodge tetracycline (Li *et al.,* 2013b). The twisted conformation of h18 would prevent normal protein synthesis and therefore the normal configuration of the helix would need to

be restored immediately (Li *et al.,* 2013b). The exact mechanism employed by OtrA in OTC resistance is not as well established as those for TetO and TetM (Connell *et al.,* 2002; Connell *et al.,* 2003; Dönhöfer *et al.,* 2012; W. Li *et al.,* 2013b), however it is assumed that the function of OtrA is the same (Doyle *et al.,* 1991; Chopra & Roberts 2001). If it is assumed that OtrA also interacts with h18 and alters the helix's conformation then rapid restoration would be necessary for normal protein synthesis. As there appears to be a surplus of S4 ribosomal protein present in the 46 hour sample of SR50S (Figure 7.10), the OTC non-producer, but none in SR1 or SR55 it could be hypothesised that S4 restores h18 configuration. In this way S4 ribosomal protein would be constantly repairing the ribosome and therefore there would be no excess available to bind the beads. It should be noted however that OtrA only shares 33 % identity with TetO and 32 % with TetM from *Enterococcus faecalis*, whereas TetO from *C. jejuni* and TetM share 76 % identity. Further to this binding of tetracycline to the ribosome has been shown to decrease in the presence of TetM but not in the presence of TetO (Doyle *et al.,* 1991), so it is possible that the protection conferred by OtrA is functionally the same but mechanistically different.

No direct tandem repeats, as seen with the OTC, ACT, DNR/DXR, FDM and VLM biosynthesis genes (Figure 7.7 B), were detected during the examination of the USR of *otcR*, however four repeat regions with a consensus of TTCC[G/C][C/T][T/C][G/T]CC were discovered (Figure 7.11 A). Of the four repeats three are predicted to be located on the same side of the DNA helix (Figure 7.11 B) and are located 43 to 84 base pairs upstream of the predicted transcriptional start point. In DNR/DXR regulation DnrN positively regulates *dnrI* by binding 37 to 55 base pairs and 62 to 100 base pairs upstream of the *dnrI* transcription start point (Furuya & Hutchinson 1996) which corresponds to the repeat regions of the USR of *otcR*. OtcG contains a N-terminal receiver response domain and a C-terminal LuxR-type HTH domain reminiscent of the NarL family of proteins (Baikalov *et al.,* 1996), as do DnrN and RedZ, the positive regulator of *redD* who's product regulates RED biosynthesis genes (Otten *et al.,* 2000). Similarities between

OtcG and DnrN, as well as parallels between the repeats in the USR of *otcR* and the DnrN binding region suggest a positive role for OtcG in the regulation of *otcR*. As deletion of *otcG* did not eliminate OTC production (Lenik 2009) its role in the regulation of *otcR* is likely to be supportive. The receiver response domain of OtcG has at least 4 of the 5 conserved amino acids that are believed to be required for phosphorylation (Baikalov *et al.,* 1996; Zhang *et al.,* 2003; Wang *et al.,* 2009a), unlike RedZ and DnrN, and most notably the aspartate (D59) that is phosphorylated (Baikalov *et al.,* 1996; Zhang *et al.,* 2003; Sanders *et al.,* 1989). Further to this, mutation of the phosphorylated aspartate (D59) in DnrN did not eliminate DnrN activity and no phosphorylation of the wild type protein could be detected (Furuya & Hutchinson 1996) suggesting DnrN is not controlled by its phosphorylation state. It is possible OtcG's function is still regulated by its phosphorylation state due to the presence of at least 4 of the necessary amino acids but, as demonstrated with DnrN, these amino acid residues may merely be an artifact left over from its ancestral origin. However if phosphorylation of OtcG does occur then the mechanism of this phosphorylation must be considered, as no cognate sensor kinase is present. It has been noted previously that acetyl phosphate can function as a phosphate donor for response regulators in the absence of a sensor kinase and acetyl phosphate concentrations increase when carbon flow into glycolysis is greater than the flow into the tricarboxylic acid cycle (TCA) (McCleary *et al.,* 1993). OTC production occurs as *S. rimosus* cultures enter the stationary phase of growth and it is during this time that expression of glycolysis related genes increases and the production of TCA cycle enzymes decrease (Navarro Llorens *et al.,* 2010). Under these conditions acetyl phosphate concentrations would increase and therefore be able to donate phosphate to OtcG. If it is assumed that the repeats in the USR of *otcR* are DNA binding motifs for OtcG and that in order to bind DNA OtcG must be phosphorylated then a hypothesis of the regulation of *otcR* can be formed (Figure 7.12). OtcG enhances OTC production but is not absolutely required, this would imply OtcG is not fully responsible for the expression of *otcR*. As SRIM2039 was found bound to

the USR of *otcR* at 18 hours but not 46 hours then it can be proposed that it prevents expression of *otcR* while bound (Figure 7.12). During the transition from exponential growth phase to stationary phase SRIM2039 becomes unbound from the *otcR* promoter and then phosphorylated OtcG binds leading to *otcR* being highly expressed; as levels of phosphorylated OtcG diminish there would be a transition from high level of expression to low (Figure 7.12). If this model is correct then it would explain why OtcG was not detected during the biotin streptavidin experiment at 46 hours (Figure 7.10) and why expression of *otcC* and *otcX* declines after 46 hours (McDowall *et al.,* 1999). Analysis of the gene expression profiles of *otcR* and the *otc* biosynthesis genes, in a wild type strain and in *otcG* and *otcR* deletion mutants, would be required to truly understand the role that both *otcR* and *otcG* play in OTC production.

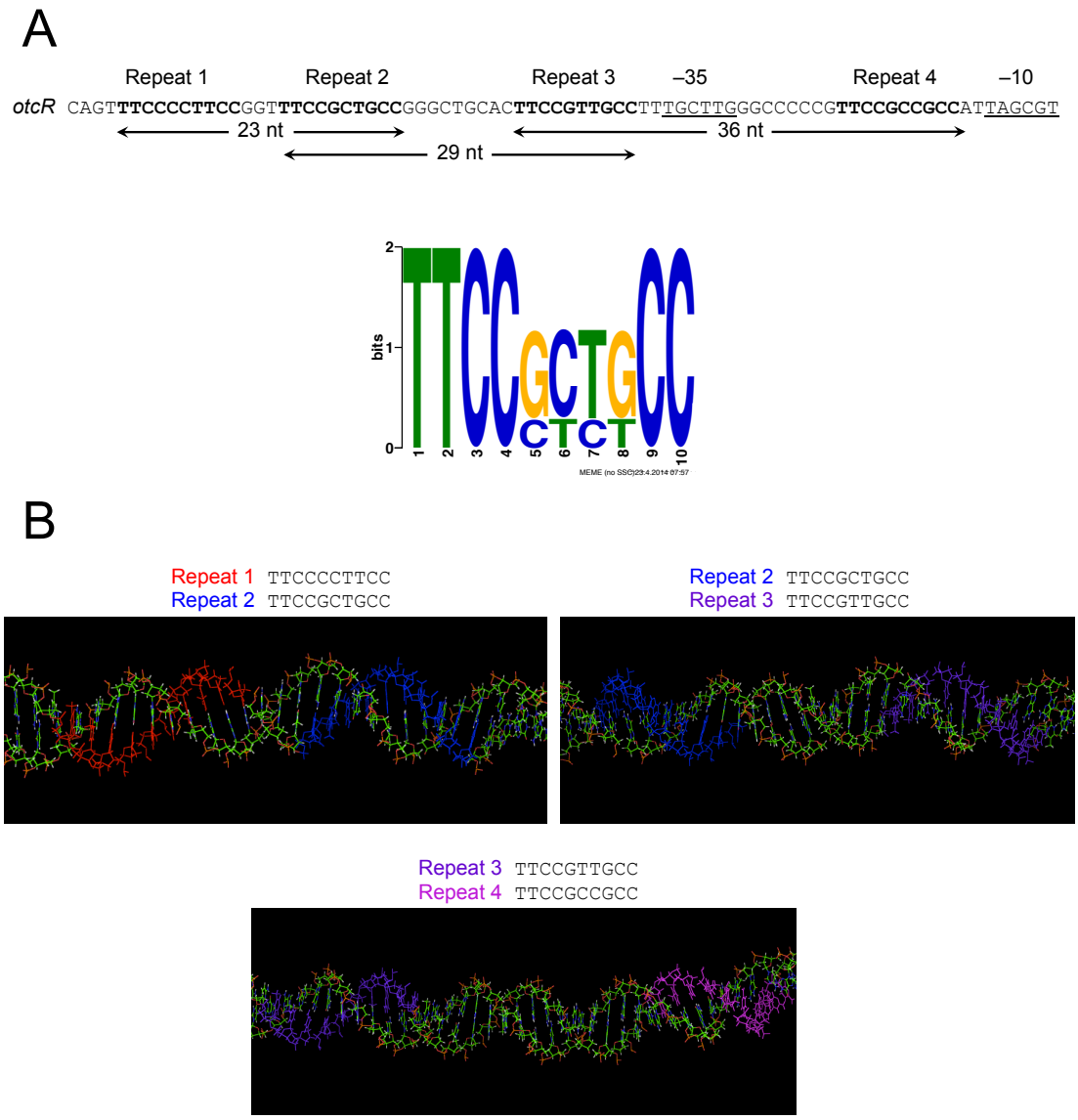


Figure 7.13 Overview of four potential DNA binding motifs in the USR of *otcR*. **A.** Relative position of the repeat regions identified in the USR of the *otcR* gene and the consensus sequence motif. Putative -10 and -35 regions were identified using general consensus information (Hawley & McClure 1983; Harley & Reynolds 1987) and comparison to those identified previously (McDowall *et al.,* 1999) **B.** 3D DNA model of the sequence in A. generated using Model.it. The text colour of each repeat corresponds to the same coloured segment of DNA, red = repeat 1, blue = repeat 2, purple = repeat 3 and magenta = repeat 4.

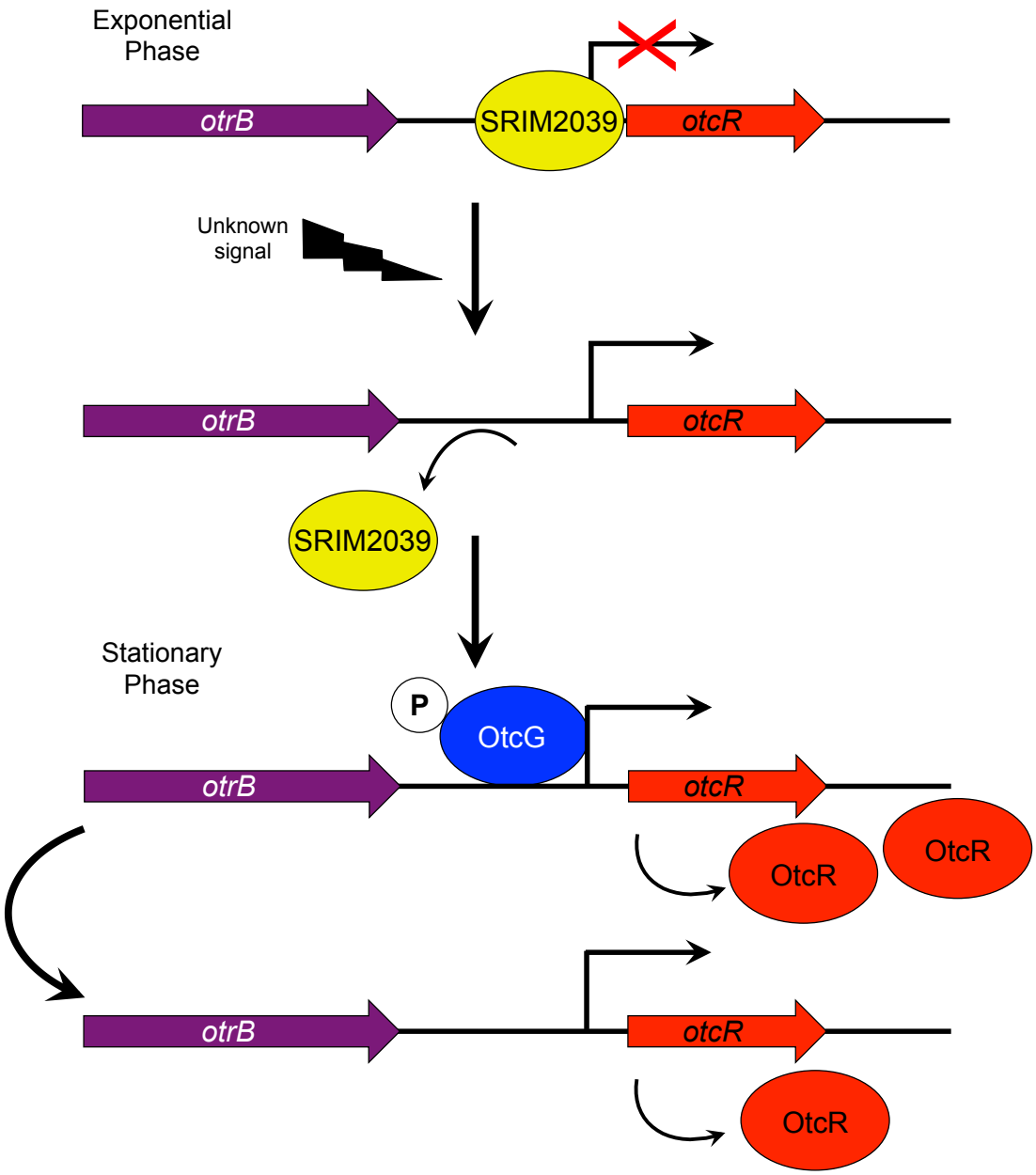


Figure 7.12 Putative model for the regulation of *otcR*. Phosphorylated OtcG is indicated by the letter P inside an open circle.

Chapter 8

8 General discussion and future work

8.1 General Discussion

When *S. rimosus* G7 (SR1) was discovered by Pfizer and oxytetracycline (OTC) was first isolated in 1950 (Finlay & Hobby 1950), this strain only produced 1 g/L of the antibiotic. However over the years that followed, the productivity of *S. rimosus* was increased to more than 55 g/L of OTC through mutagenesis and screening for high producers. Throughout those years new high producing strains were produced sequentially: M4018 (SR20) and 15883 (SR50), due to treatment with N-methyl-N'-nitro-N'-nitrosogaunidine (MNNG) and UV irradiation (Petković *et al.,* 2006). Although this treatment increased OTC productivity, it also produced genetically unstable strains, leading to the isolation of *S. rimosus* 15883S (SR50S), a strain with a spontaneous deletion of the OTC biosynthetic cluster (McDowall *et al.,* 1999). However the availability of all 4 strains has provided the opportunity to analyse the precise genomic differences that occurred in each and link these changes to OTC productivity. Before the work for this thesis was undertaken genome sequencing had been completed for *S. rimosus* M4018, 15883S and 23383, therefore during this work genome sequencing of *S. rimosus* G7 was performed. The genome sequence of *S. rimosus* G7 was then used as a reference for the identification of the genomic changes that took place during the forced evolution of the later production strains (Pethick *et al.,* 2013).

The identification of genomic changes between G7 and the production strains was achieved by whole genome comparisons. The objective of these comparisons was to identify those genomic differences that might be responsible for increased OTC production, thus providing a genomic change to be tested in *S. rimosus* G7. From the analysis undertaken during the

comparison chapter (Chapter 4), almost 600 genes were found to be deleted and over 512 CDS were shown to contain non-synonymous SNCs. Both genes lost and non-synonymous SNCs, along with sInDELs, were the primary focus of the comparisons when determining the effect on OTC productivity. Due to the time constraints and the sheer volume of data, limiting the breadth was necessary but still provides a useful starting point for this study. However, synonymous and intergenic SNCs are likely to play a part in OTC productivity as well. Synonymous mutations, although do not change the amino acid encoded, do alter codon usage, which can lower protein production (Agashe *et al.,* 2013) as the required transfer RNA (tRNA) needs to be expressed. An example of this in *Streptomyces* is *bldA* as it encodes a leucine tRNA that is associated with the rare codon UUA (Chater & Chandra 2008). The expression of genes containing the TTA codon are dependent on the expression of *bldA*, and for this reason no essential genes contain the rare TTA codon (Hengst *et al.,* 2010; Liu *et al.,* 2013). The presence of the TTA codon in genes provides an extra level of regulation due to the requirement of BldA, with *bldA* expression upregulated as the Streptomyces growth begins to slow and enter stationary phase (Liu *et al.,* 2013). For this reason *bldA* is associated with antibiotic production and development, as *bldA* mutants are defective in both (Chater & Chandra 2008). In *S. coelicolor* A3(2) antibiotic production is disrupted in *bldA* mutants due to the presence of the TTA codon in the pathway specific regulators *actII-ORF4* and *redZ* (Liu *et al.,* 2013). Further to this, substitution of the TTA codon in *actII-ORF4* with another leucine-associated codon, TTG, could restore actinorhodin (ACT) production in *bldA* mutants (Fernández-Moreno *et al.,* 1991). Additionally the expression of *adpA* (*bldH*) in *S. coelicolor* A3(2) is also dependent on *bldA*, once again demonstrated by substitution of the TTA codon with TTG leading to partial restoration of aerial hyphae formation (Takano *et al.,* 2003). Therefore, the TTA codon provides a useful mechanism for controlling the expression of genes and removing competition from other pathways that require BldA (Chater & Chandra 2008; Liu *et al.,* 2013). Given the role played by *bldA* in antibiotic production, both

otcG and *otcR* were examined in order to identify the presence of a TTA codon. Only *otcG* contained the rare codon, which potentially ensures *otcR* expression cannot occur unless there is sufficient OtcG available, which will only occur if *bldA* is being expressed. Additionally, the presence of SNCs that converted a codon either to or from TTA was investigated, as those genes associated with antibiotic regulation that require *bldA* expression would be relieved or placed under the control of BldA if the TTA codon was absent or present. Of the total 1013 SNCs present in all of the production strains, only SR50S had a non-synonymous SNC that converted a TTC codon to TTA, present in a CDS encoding a hypothetical protein.

As well as altering protein production synonymous mutations can also decrease the number of single point mutations required to cause a non-synonymous mutation to arise in a particular gene (Cambray & Mazel 2008). This can be thought of as a synonymous mutation converting a tyrosine codon from TAC to TAT, although the amino acid encoded is still tyrosine the codon sequence would only require a single mutation to convert it into the histidine codon CAT. Further to this, it has been demonstrated that synonymous mutations can influence the adaptive stability of a gene by either increasing or decreasing its potential for non-synonymous mutations to arise (Cambray & Mazel 2008). Intergenic SNCs can be considered in a similar way to synonymous SNCs, on the surface these SNCs do not appear to influence the gene or encoded protein. However depending on the location of the intergenic SNC gene expression may be altered. Binding to target regions of DNA by regulators can be abolished by introducing point mutations within these site-specific DNA regions. In some cases this requires mutating all sites within the target or as little as two (Hirano *et al.,* 2008; Tucker *et al.,* 2010). Depending on the target site sequence and affinity of the regulator, SNCs within an intergenic region could alter expression of a gene.

After identifying the loss of the rimocidin cluster in M4018 it was necessary to determine if loss of the rimocidin cluster lead to increased OTC production. Therefore, a *rim* knockout mutant was produced in order to test

this hypothesis. Although it was not shown that the loss of the rimocidin cluster in M4018 caused the primary increase in OTC production for that strain, it seems likely. As mentioned previously (Chapter 5) the rimocidin knockout mutant, ACM23, only disrupted the rimocidin biosynthetic pathway after module 8 and therefore 8 molecules of acetate would still be incorporated. In order to prevent any acetate being redirected into rimocidin production a knockout mutant of *rimA* should be produced as this gene encoding the loading domain of the rimocidin biosynthetic cluster. Alternatively a knockout mutant of the putative regulator of the *rim* cluster, *SRIM7669*, could be produced, with the additional benefit of elucidating the role SRIM7669 plays in *rim* regulation. Recently the genome sequence of another high OTC producer, *S. rimosus* R6-500, has become available (Baranasic *et al.,* 2014). Unlike M4018 this strain still contains the rimocidin cluster, however examination of the DNA sequence for *rimA*, within R6-500 shows a non-synonymous SNC within the carboxylic acid-CoA ligase domain. Further to this initial rimocidin bioassays, using agar plugs of R6-500 show no zones of inhibition when tested against *Saccharomyces cerevisiae*, demonstrating that R6-500 does not produce rimocidin as a final product (Figure 8.1). It is likely that the mutation in RimA prevents its activity and therefore rimocidin production cannot be initiated, thus leading to an increase in OTC production. If this is the case then it is interesting to see two different mechanisms for eliminating rimocidin production for increased OTC productivity.

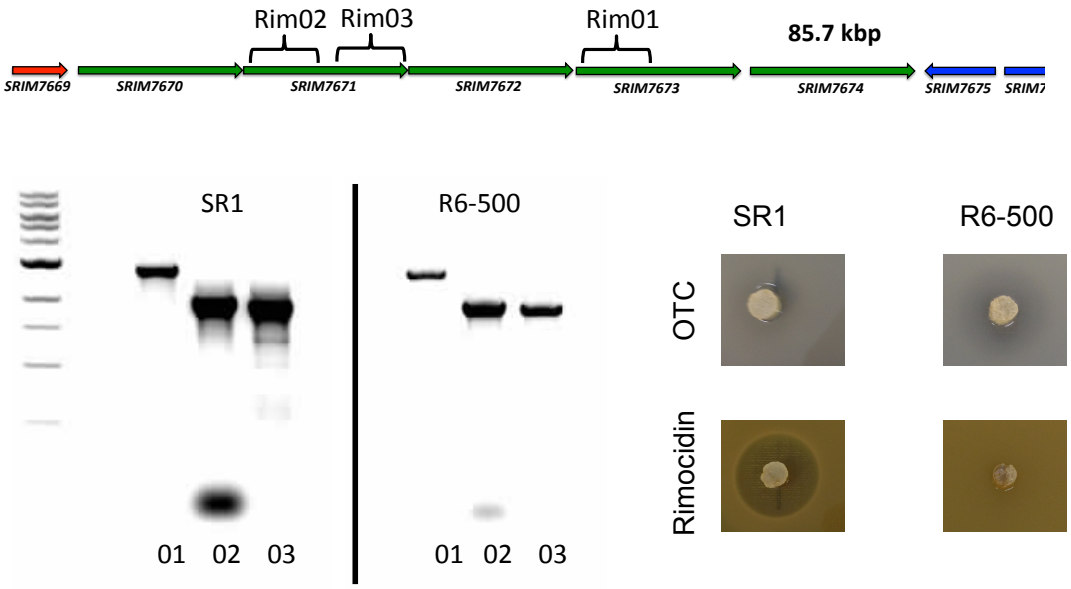


Figure 8.1 Confirmation of *rim* cluster presence and lack of rimocidin biosynthesis by R6-500. PCR primer pairs Rim01, Rim02 and Rim03 were used, represented by 01, 02 and 03. *E. coli* was used as the test organism for OTC assays, and *S. cerevisiae* for rimocidin assays.

Before the complete genome sequence of *S. rimosus* G7 was obtained, only two regulators were associated with the *otc* biosynthetic cluster, OtcG and OtrR. Of those two regulators only OtcG has been shown to have a positive role in OTC production, whereas OtrR is thought to be a repressor of *otrB* expression (Petković *et al.,* 2006; Lenik 2009). Identification of a gene that encodes a transcriptional regulator that lay downstream of *otrB* provided an opportunity to better understand *otc* regulation. Although attempts to knockout *otcR* were unsuccessful, as well as the attempts to purify OtcR in a form that could be used for DNA binding experiments, the preliminary data suggests it is a regulator of the *otc* biosynthetic genes. The repeat regions within the promoters of the *otc* genes, reminiscent of putative and verified SARP binding sites for other antibiotic gene clusters, and sequence similarity within the Trans-Reg domains of those SARPs and OtcR further supports this. Whether OtcR plays a positive or negative role in the expression of the *otc* genes has yet to be determined. The majority of SARPs play a positive role in the expression of their cognate biosynthetic cluster, as seen with TylT and TylS, PapR1, PapR2 and PapR4, and SrrY and SrrZ, for tylosin, pristinamycin and lankamycin production respectively (Suzuki *et al.,* 2010; Mast *et al.,* 2010; Khaliq *et al.,* 2014). In addition those SARP regulators that contain a N-terminal Trans-Reg domain, with a C-terminal BTAD domain have the conserved leucine within the LALL motif (Chapter 7, Figure 7.1 B). OtcR from strain 23383 has the conserved leucine substituted by phenylalanine and it is possible this alters DNA binding affinity. As the same amino acid substitution is present in the positive SARP regulator of daunorubicin/doxorubicin (DNR/DXR), DnrI (Chapter 7, Figure 7.1 C), it is unlikely this change prevents binding. Overall it seems likely that OtcR plays a positive role in OTC production, but wet data would need to be obtained to confirm this.

Overall the data obtained in Chapter 4 should provide a directed approach to strain improvement by either supporting an hypothesis about the importance of a gene in relation to increased antibiotic production, i.e the *zwf*

gene (Tang *et al.,* 2011), or by giving a steer towards genes that should be deleted or mutated for improved production. As seen with the data obtained in Chapter 3, applying the GET_HOMOLOGUES algorithm to bacterial genomes of different species allows for the rapid identification of homologues, as well as paralogues within an organism. GET_HOMOLOGUES could be applied to the genome of an organism, of which increased antibiotic production is desired, and the genome of *S. rimosus.* In this way, any potential targets for mutation or deletion found within *S. rimosus* can be easily identified in the organism of choice, by finding its homologue.

8.2 Future Work

The large number of genomic differences between G7 and the production strains makes it difficult to pinpoint those differences that are actually linked with increased OTC productivity. One of the most prominent SEED categories associated with the deletion profiles and mutations, of the production strains, was secondary metabolism. This would suggest that the production of other secondary metabolites direct resources away from OTC production. However, this would mean that those secondary metabolites are being produced at the same time as OTC. In addition, deleted genes and mutations that have occurred in the production strains may affect precursor supply in a way that is not immediately clear. This could be via primary metabolism, expression of other genes or response to external stimuli. For this reason examination of the metabolomic profiles of G7, M4018, 15883S and 23383 in correlation with the genomic profiles should be undertaken. Further to this, if undertaken in a quantitative manner, then the concentrations of particular primary and secondary metabolites can be determined. In doing so, any differences could potentially be linked to a genomic difference, if not aid identification of those genomic differences that are linked to increased OTC production. In addition tandem mass spectrometry could be utilised, to not only identify secondary metabolites

from the predicted clusters within *S. rimosus* but also to aid in elucidation of their structures.

The true relationship between rimocidin and OTC productivity needs to be examined further. Disruption of *rimA* or *SRIM7669* would be ideal for this purpose, as both would eliminate rimocidin production, if the predicted role of SRIM7669 is correct. In addition, as rimocidin production would never be initiated no carbon would be diverted from OTC production and therefore the correlation between rimocidin and OTC productivity would be more readily identified.

Disruption of *otcR* is essential for confirming the role it plays in OTC production. Complementation of *otcR* in G7 with a copy from 23383 would demonstrate the role the non-synonymous SNC plays in increased OTC production, if any at all. In addition, examination of the identified tandem repeats within the promoters of the *otc* genes should be analysed, once successful purification of OtcR in a form compatible with DNA-binding assays has been obtained. Most likely encoding a HIS-tag at the C-terminal of the protein will resolve this issue. Alternatively, a plasmid encoding a maltose binding protein could be utilised, as this can aid solubility as well as purification (Childers *et al.,* 2007). If OtcR binds to the upstream regions of the otc genes then the precise target regions could be determined. Site-directed mutagenesis can be applied to the tandem repeats for this purpose but also to examine the change in DNA binding affinity of OtcR from 23383. If the phenylalanine substitution increases binding affinity then a greater number of nucleotide sites within the tandem repeats would need to be altered to disrupt binding.

As OtcG has been shown to play a positive role in OTC production, it is necessary to understand the precise role it plays. The model proposed here suggests that OtcG acts as a positive regulator of *otcR* (Chapter 7, Figure 7.14), by binding to the repeats identified within the upstream region of otcR (Chapter 7, Figure 7.13 A). Further to this, it was suggested that OtcG would only bind while in a phosphorylated state, enhancing the expression of *otcR*. Once the levels of phosphorylated OtcG diminish then

this enhanced expression would decrease. If this is the case then *otcR* expression would be lower in an *otcG* mutant when compared to the wild type.

Chapter 9

9 References

Agashe, D., Martinez-Gomez, N.C., Drummond, D.A., Marx, C.J. (2013) Good Codons, Bad Transcript: Large Reductions in Gene Expression and Fitness Arising from Synonymous Mutations in a Key Enzyme. Molecular Biology and Evolution. 30(3), 549–560.

Ahmed, S., Craney, A., Pimentel-Elardo, S.M., Nodwell, J.R. (2012) A Synthetic, Species-Specific Activator of Secondary Metabolism and Sporulation in Streptomyces coelicolor. ChemBioChem. 14(1), 83–91.

Aizawa, T., Kim, SY., Takahashi, S., Koshita, M., Tani, M., Futamura, Y., Osada, H., Funa, N. (2014) Alkyldihydropyrones, new polyketides synthesized by a type III polyketide synthase from Streptomyces reveromyceticus. Journal of Antibiotics.

Ajithkumar, V., Prasad, R. (2010) Modulation of dnrN expression by intracellular levels of DnrO and daunorubicin in Streptomyces peucetius. FEMS Microbiology Letters. 306(2), 160–167.

Akey, D.L., Razelun, J.R., Tehranisa, J., Sherman, D.H., Gerwick, W.H., Smith, J.L. (2010) Crystal Structures of Dehydratase Domainsfrom the Curacin Polyketide Biosynthetic Pathway. Structure/Folding and Design. 18(1), 94–105.

Akopiants, K., Florova, G., Li, C., Reynolds, K.A. (2005) Multiple pathways for acetate assimilation in Streptomyces cinnamonensis. Journal of Industrial Microbiology & Biotechnology. 33(2), 141–150.

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

Alderwick, L.J., Molle, V., Kremer, L., Cozzone, A.J., Dafforn, T.R., Besra, G.S., Fütterer, K. (2006) Molecular structure of EmbR, a response element of Ser/Thr kinase signaling in Mycobacterium tuberculosis. Proceedings of the National Academy of Sciences of the United States of America. 103(8), 2558–2563.

Alekshun, M.N., Levy, S.B., Mealy, T.R., Seaton, B.A., Head, J.F. (2001) The crystal structure of MarR, a regulator of multiple antibiotic resistance, at 2.3 A resolution. Nature Structural Biology. 8(8), 710–714.

Allenby, N.E.E., Laing, E., Bucca, G., Kierzek, A.M., Smith, C.P. (2012) Diverse control of metabolism and other cellular processes in Streptomyces

coelicolor by the PhoP transcription factor: genome-wide identification of in vivo targets. Nucleic Acids Research. 40(19), 9543–9556.

Almutairi, A.S. (2007) Heterologous expression of the oxytetracycline gene cluster from S. rimosus M4018 in different streptomyces species. University of Strathclyde

Ames, B.D., Nguyen, C., Bruegger, J., Smith, P., Xu, W., Ma, S., Wong, E., Wong, S., Xie, X., Li, J.W.-H., Vederas, J.C., Tang, Y., Tsai, S.-C. (2012) Crystal structure and biochemical studies of the trans-acting polyketide enoyl reductase LovC from lovastatin biosynthesis. Proceedings of the National Academy of Sciences. 109(28), 11144–11149.

Anand, S., Mohanty, D. (2012) Modeling holo-ACP:DH and holo-ACP:KR complexes of modular polyketide synthases: a docking and molecular dynamics study. BMC Structural Biology. 12(1), 1–1.

Anand, S., Prasad, M.V.R., Yadav, G., Kumar, N., Shehara, J., Ansari, M.Z., Mohanty, D. (2010) SBSPKS: structure based sequence analysis of polyketide synthases. Nucleic Acids Research. 38(Web Server), W487–W496.

Anderson, A.S., Wellington, E.M. (2001) The taxonomy of Streptomyces and related genera. International Journal Of Systematic And Evolutionary Microbiology. 51(3), 797–814.

Angiuoli, S.V., Salzberg, S.L. (2011) Mugsy: fast multiple alignment of closely related whole genomes. Bioinformatics. 27(3), 334–342.

Anton, N., Santos-Aberturas, J., Mendes, M.V., Guerra, S.M., Martin, J.F., Aparicio, J.F. (2007) PimM, a PAS domain positive regulator of pimaricin biosynthesis in Streptomyces natalensis. Microbiology. 153(9), 3174–3183.

Antón, N., Mendes, M.V., Martín, J.F., Aparicio, J.F. (2004) Identification of PimR as a positive regulator of pimaricin biosynthesis in Streptomyces natalensis. Journal of Bacteriology. 186(9), 2567–2575.

Aparicio, J.F., Fouces, R., Mendes, M.V., Olivera, N., Martin, J.F. (2000) A complex multienzyme system encoded by five polyketide synthase genes is involved in the biosynthesis of the 26-membered polyene macrolide pimaricin in Streptomyces natalensis. Chemistry & Biology. 7(11), 895–905.

Arias, P., Fernández-Moreno, M.A., Malpartida, F. (1999) Characterization of the pathway-specific positive transcriptional regulator for actinorhodin biosynthesis in Streptomyces coelicolor A3(2) as a DNA-binding protein. Journal of Bacteriology. 181(22), 6958–6968.

Arthur, C.J., Szafranska, A., Evans, S.E., Findlow, S.C., Burston, S.G., Owen, P., Clark-Lewis, I., Simpson, T.J., Crosby, J., Crump, M.P. (2005) Self-Malonylation Is an Intrinsic Property of a Chemically Synthesized Type II

Polyketide Synthase Acyl Carrier Protein †,‡. Biochemistry. 44(46), 15414–15421.

Arthur, C.J., Williams, C., Pottage, K., Płoskoń, E., Findlow, S.C., Burston, S.G., Simpson, T.J., Crump, M.P., Crosby, J. (2009) Structure and Malonyl CoA-ACP Transacylase Binding of Streptomyces coelicolorFatty Acid Synthase Acyl Carrier Protein. ACS Chemical Biology. 4(8), 625–636.

Austin, M.B., Noel, J.P. (2002) The chalcone synthase superfamily of type III polyketide synthases. Natural Product Reports. 20(1), 79–110.

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., Zagnitko, O. (2008) The RAST Server: Rapid Annotations using Subsystems Technology. BMC Genomics. 9(1), 75.

Bai, L., Qi, X., Zhang, Y., Yao, C., Guo, L., Jiang, R., Zhang, R., Li, Y. (2013) A new GntR family regulator Ste1 in Streptomyces sp. 139. Applied Microbiology and Biotechnology. 97(19), 8673–8682.

Baikalov, I., Schröder, I., Kaczor-Grzeskowiak, M., Grzeskowiak, K., Gunsalus, R.P., Dickerson, R.E. (1996) Structure of the Escherichia coli response regulator NarL. Biochemistry. 35(34), 11053–11061.

Bailey, R.A. (1984) Quasi-complete latin squares: construction and randomization. Journal of the Royal Statistical Society. Series B (Methodological), 323–334.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., Noble, W.S. (2009) MEME SUITE: tools for motif discovery and searching. Nucleic Acids Research. 37(Web Server), W202–W208.

Ball, S., Bessell, C.J., Mortimer, A. (1957) The production of polyenic antibiotics by soil streptomycetes. Journal of General Microbiology. 17(1), 96–103.

Banskota, A.H., McAlpine, J.B., Sørensen, D., Aouidate, M., Piraee, M., Alarco, A.-M., Omura, S., Shiomi, K., Farnet, C.M., Zazopoulos, E. (2006) Isolation and identification of three new 5-alkenyl-3, 3 (2H)-furanones from two Streptomyces species using a genomic screening approach. The Journal of Antibiotics. 59(3), 168–176.

Baranasic, D., Gacesa, R., Starcevic, A., Zucko, J., Blazic, M., Horvat, M., Gjuracic, K., Fujs, S., Hranueli, D., Kosec, G., Cullum, J., Petkovic, H. (2013) Draft Genome Sequence of Streptomyces rapamycinicus Strain NRRL 5491, the Producer of the Immunosuppressant Rapamycin. Genome Announcements. 1(4), e00581-13.

Baranasic, D., Zucko, J., Nair, M., Pain, A., Long, P.F., Hranueli, D., Cullum, J., Starcevic, A. (2014) Genome Sequences of the Oxytetracycline Production Strain Streptomyces rimosus R6-500 and Two Mutants with Chromosomal Rearrangements. Genome Announcements. 2(4), e00517–14.

Barbieri, C.M., Wu, T., Stock, A.M. (2013) Comprehensive Analysis of OmpR Phosphorylation, Dimerization, and DNA Binding Supports a Canonical Model for Activation. Journal of Molecular Biology. 425(10), 1612–1626.

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

Bellur, D.L., Woodson, S.A. (2009) A minimized rRNA-binding site for ribosomal protein S4 and its implications for 30S assembly. Nucleic Acids Research. 37(6), 1886–1896.

Beltran-Alvarez, P., Cox, R.J., Crosby, J., Simpson, T.J. (2007) Dissecting the Component Reactions Catalyzed by the Actinorhodin Minimal Polyketide Synthase †. Biochemistry. 46(50), 14672–14681.

Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, 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., Hopwood, D.A. (2002) Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). Nature. 417(6885), 141–147.

Berdy, J. (2012) Thoughts and facts about antibiotics: Where we are now and where we are heading. The Journal of Antibiotics. 65(8), 385–395.

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

Binnie, C., Warren, M., Butler, M.J. (1989) Cloning and heterologous expression in Streptomyces lividans of Streptomyces rimosus genes involved in oxytetracycline biosynthesis. Journal of Bacteriology. 171(2), 887–895.

Bisang, C., Long, P.F., Cortés, J., Westcott, J., Crosby, J., Matharu, A.L., Cox, R.J., Simpson, T.J., Staunton, J., Leadlay, P.F. (1999) A chain initiation factor common to both modular and aromatic polyketide synthases. Nature. 401(6752), 502–505.

Bishop, A., Fielding, S., Dyson, P., Herron, P. (2004) Systematic Insertional

Mutagenesis of a Streptomycete Genome: A Link Between Osmoadaptation and Antibiotic Production. Genome Research. 14(5), 893–900.

Blin, K., Medema, M.H., Kazempour, D., Fischbach, M.A., Breitling, R., Takano, E., Weber, T. (2013) antiSMASH 2.0--a versatile platform for genome mining of secondary metabolite producers. Nucleic Acids Research. 41(W1), W204–W212.

Bonnett, S.A., Whicher, J.R., Papireddy, K., Florova, G., Smith, J.L., Reynolds, K.A. (2013) Structural and Stereochemical Analysisof a Modular Polyketide Synthase Ketoreductase Domain Required for the Generation of a cis-Alkene. Chemistry & Biology. 20(6), 772–783.

Brooks, B.E., Buchanan, S.K. (2008) Signaling mechanisms for activation of extracytoplasmic function (ECF) sigma factors. Biochimica et Biophysica Acta (BBA) - Biomembranes. 1778(9), 1930–1945.

Browning, D.F., Busby, S.J.W. (2004) The regulation of bacterial transcription initiation. Nature Reviews Microbiology. 2(1), 57–65.

Bruheim, P., Sletta, H., Bibb, M.J., White, J., Levine, D.W. (2002) High-yield actinorhodin production in fed-batch culture by a Streptomyces lividans strain overexpressing the pathway-specific activator gene actII-ORF4. Journal of Industrial Microbiology & Biotechnology. 28(2), 103–111.

Bursy, J., Kuhlmann, A.U., Pittelkow, M., Hartmann, H., Jebbar, M., Pierik, A.J., Bremer, E. (2008) Synthesis and Uptake of the Compatible Solutes Ectoine and 5-Hydroxyectoine by Streptomyces coelicolor A3(2) in Response to Salt and Heat Stresses. Applied and Environmental Microbiology. 74(23), 7286–7296.

Bushley, K.E., Raja, R., Jaiswal, P., Cumbie, J.S., Nonogaki, M., Boyd, A.E., Owensby, C.A., Knaus, B.J., Elser, J., Miller, D., Di, Y., McPhail, K.L., Spatafora, J.W. (2013) The Genome of Tolypocladium inflatum: Evolution, Organization, and Expression of the Cyclosporin Biosynthetic Gene Cluster J. Heitman, ed. PLoS Genetics. 9(6), e1003496.

Butler, M.J., Friend, E.J., Hunter, I.S., Kaczmarek, F.S., Sugden, D.A., Warren, M. (1989) Molecular cloning of resistance genes and architecture of a linked gene cluster involved in biosynthesis of oxytetracycline by Streptomyces rimosus. Molecular & General Genetics : MGG. 215(2), 231–238.

Cambray, G., Mazel, D. (2008) Synonymous Genes Explore Different Evolutionary Landscapes. PLoS Genetics. 4(11), –e1000256.

Cane, D.E. (2010) Programming of Erythromycin Biosynthesis by a Modular Polyketide Synthase. Journal of Biological Chemistry. 285(36), 27517–27523.

Cane, D.E., He, X., Kobayashi, S., Omura, S., Ikeda, H. (2006) Geosmin biosynthesis in Streptomyces avermitilis. Molecular cloning, expression, and mechanistic study of the germacradienol/geosmin synthase. The Journal of Antibiotics. 59(8), 471–479.

Cao, B., Yao, F., Zheng, X., Cui, D., Shao, Y., Zhu, C., Deng, Z., You, D. (2012) Genome Mining of the Biosynthetic Gene Cluster of the Polyene Macrolide Antibiotic Tetramycin and Characterization of a P450 Monooxygenase Involved in the Hydroxylation of the Tetramycin B Polyol Segment. ChemBioChem. 13(15), 2234–2242.

Carmody, M., Byrne, B., Murphy, B., Breen, C., Lynch, S., Flood, E., Finnan, S., Caffrey, P. (2004) Analysis and manipulation of amphotericin biosynthetic genes by means of modified phage KC515 transduction techniques. Gene. 343(1), 107–115.

Carreras, C.W., Khosla, C. (1998) Purification and in vitro reconstitution of the essential protein components of an aromatic polyketide synthase. Biochemistry. 37(8), 2084–2088.

Carver, T., Harris, S.R., Berriman, M., Parkhill, J., McQuillan, J.A. (2012) Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. Bioinformatics. 28(4), 464–469.

Castaldo, G., Zucko, J., Heidelberger, S., Vujaklija, D., Hranueli, D., Cullum, J., Wattana-Amorn, P., Crump, M.P., Crosby, J., Long, P.F. (2008) Proposed Arrangement of Proteins Forming a Bacterial Type II Polyketide Synthase. Chemistry & Biology. 15(11), 1156–1165.

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

Challis, G.L. (2008) Mining microbial genomes for new natural products and biosynthetic pathways. Microbiology. 154(6), 1555–1569.

Chan, D.I., Vogel, H.J. (2010) Current understanding of fatty acid biosynthesis and the acyl carrier protein. Biochemical Journal. 430(1), 1–19.

Chan, Y.A., Podevels, A.M., Kevany, B.M., Thomas, M.G. (2008) Biosynthesis of polyketide synthase extender units. Natural Product Reports. 26(1), 90.

Chater, K.F. (2001) Regulation of sporulation in Streptomyces coelicolor A3 (2): a checkpoint multiplex? Current Opinion in Microbiology. 4(6), 667–673.

Chater, K.F. (2006) Streptomyces inside-out: a new perspective on the bacteria that provide us with antibiotics. Philosophical Transactions of the

Royal Society B: Biological Sciences. 361(1469), 761–768.

Chater, K.F., Chandra, G. (2008) The use of the rare UUA codon to define 'Expression Space' for genes involved in secondary metabolism, development and environmental adaptation in Streptomyces. The Journal of Microbiology. 46(1), 1–11.

Chávez, A., Forero, A., Sánchez, M., Rodríguez-Sanoja, R., Mendoza-Hernández, G., Servín-Gonzalez, L., Sánchez, B., García-Huante, Y., Rocha, D., Langley, E., Ruiz, B., Sánchez, S. (2010) Interaction of SCO2127 with BldKB and its possible connection to carbon catabolite regulation of morphological differentiation in Streptomyces coelicolor. Applied Microbiology and Biotechnology. 89(3), 799–806.

Chen, S., Huang, X., Zhou, X., Bai, L., He, J., Jeong, K.J., Lee, S.Y., Deng, Z. (2003) Organizational and Mutational Analysis of a Complete FR-008/Candicidin Gene Cluster Encoding a Structurally Related Polyene Complex. Chemistry & Biology. 10(11), 1065–1076.

Chen, Y., Deng, W., Wu, J., Qian, J., Chu, J., Zhuang, Y., Zhang, S., Liu, W. (2008) Genetic Modulation of the Overexpression of Tailoring Genes eryK and eryG Leading to the Improvement of Erythromycin A Purity and Production in Saccharopolyspora erythraea Fermentation. Applied and Environmental Microbiology. 74(6), 1820–1828.

Chen, Y., Wendt-Pienkowski, E., Shen, B. (2008) Identification and Utility of FdmR1 as a Streptomyces Antibiotic Regulatory Protein Activator for Fredericamycin Production in Streptomyces griseus ATCC 49344 and Heterologous Hosts. Journal of Bacteriology. 190(16), 5587–5596.

Chevreux, B. (2004) Using the miraEST Assembler for Reliable and Automated mRNA Transcript Assembly and SNP Detection in Sequenced ESTs. Genome Research. 14(6), 1147–1159.

Childers, B.M., Weber, G.G., Prouty, M.G., Castaneda, M.M., Peng, F., Klose, K.E. (2007) Identification of residues critical for the function of the Vibrio cholerae virulence regulator ToxT by scanning alanine mutagenesis. Journal of Molecular Biology. 367(5), 1413–1430.

Cho, Y.H., Kim, E.J., Chung, H.J., Choi, J.H., Chater, K.F., Ahn, B.E., Shin, J.H., Roe, J.H. (2003) The pqrAB Operon Is Responsible for Paraquat Resistance in Streptomyces coelicolor. Journal of Bacteriology. 185(23), 6756–6763.

Chopra, I., Roberts, M. (2001) Tetracycline Antibiotics: Mode of Action, Applications, Molecular Biology, and Epidemiology of Bacterial Resistance. Microbiology and Molecular Biology Reviews. 65(2), 232–260.

Claessen, D., de Jong, W., Dijkhuizen, L., Wösten, H.A.B. (2006) Regulation

of Streptomyces development: reach for the sky! Trends in Microbiology. 14(7), 313–319.

Clark, L.C., Hoskisson, P.A. (2011) Duplication and Evolution of devA-Like Genes in Streptomyces Has Resulted in Distinct Developmental Roles S. Bereswill, ed. PLoS ONE. 6(10), e25049.

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.E., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M.A., Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J.E., Taylor, K., Whitehead, S., Barrell, B.G. (1998) Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature. 393(6685), 537–544.

Connell, S.R., Tracz, D.M., Nierhaus, K.H., Taylor, D.E. (2003) Ribosomal Protection Proteins and Their Mechanism of Tetracycline Resistance. Antimicrobial Agents and Chemotherapy. 47(12), 3675–3681.

Connell, S.R., Trieber, C.A., Stelzl, U., Einfeldt, E., Taylor, D.E., Nierhaus, K.H. (2002) The tetracycline resistance protein Tet(o) perturbs the conformation of the ribosomal decoding centre. Molecular Microbiology. 45(6), 1463–1472.

Connolly, D.M., Winkler, M.E. (1991) Structure of Escherichia coli K-12 miaA and characterization of the mutator phenotype caused by miaA insertion mutations. Journal of Bacteriology. 173(5), 1711–1721.

Contreras-Moreira, B., Vinuesa, P. (2013) GET_HOMOLOGUES, a Versatile Software Package for Scalable and Robust Microbial Pangenome Analysis. Applied and Environmental Microbiology. 79(24), 7696–7701.

Corre, C., Challis, G.L. (2009) New natural product biosynthetic chemistry discovered by genome mining. Natural Product Reports. 26(8), 977.

Coze, F., Gilard, F., Tcherkez, G., Virolle, M.-J., Guyonvarch, A. (2013) Carbon-Flux Distribution within Streptomyces coelicolor Metabolism: A Comparison between the Actinorhodin-Producing Strain M145 and Its Non-Producing Derivative M1146 M. Desvaux, ed. PLoS ONE. 8(12), e84151.

Das, A., Khosla, C. (2009) Biosynthesis of Aromatic Polyketides in Bacteria. Accounts of Chemical Research. 42(5), 631–639.

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

De Schrijver, A., De Mot, R. (1999) A subfamily of MalT-related ATP-dependent regulators in the LuxR family. Microbiology. 145 (6), 1287–1288.

Demain, A.L., Inamine, E. (1970) Biochemistry and regulation of streptomycin and mannosidostreptomycinase (alpha-D-mannosidase) formation. Bacteriological Reviews. 34(1), 1–19.

Doroghazi, J.R., Buckley, D.H. (2010) Widespread homologous recombination within andbetween Streptomyces species. The ISME Journal. 4(9), 1136–1143.

Doroghazi, J.R., Metcalf, W.W. (2013) Comparative genomics of actinomycetes with a focus on natural product biosynthetic genes. BMC Genomics. 14(1), 1–1.

Doyle, D., McDowall, K.J., Butler, M.J., Hunter, I.S. (1991) Characterization of an oxytetracycline-resistance gene, otrA, of Streptomyces rimosus. Molecular Microbiology. 5(12), 2923–2933.

Dönhöfer, A., Franckenberg, S., Wickles, S., Berninghausen, O., Beckmann, R., Wilson, D.N. (2012) Structural basis for TetM-mediated tetracycline resistance. Proceedings of the National Academy of Sciences. 109(42), 16900–16905.

Du, L., Lou, L. (2010) PKS and NRPS release mechanisms. Natural Product Reports. 27(2), 255.

Du, Y.L., Li, S.Z., Zhou, Z., Chen, S.F., Fan, W.M., Li, Y.Q. (2011) The pleitropic regulator AdpAch is required for natamycin biosynthesis and morphological differentiation in Streptomyces chattanoogensis. Microbiology. 157(5), 1300–1311.

Ducros, V.M.A., Lewis, R.J., Verma, C.S., Dodson, E.J., Leonard, G., Turkenburg, J.P., Murshudov, G.N., Wilkinson, A.J., Brannigan, J.A. (2001) Crystal structure of GerE, the ultimate transcriptional regulator of spore formation in Bacillus subtilis. Journal of Molecular Biology. 306(4), 759–771.

Elias, M., Wellner, A., Goldin-Azulay, K., Chabriere, E., Vorholt, J.A., Erb, T.J., Tawfik, D.S. (2012) The molecular basis of phosphate discrimination in arsenate-rich environments. Nature. 490(7422), 134–137.

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

Elliot, M.A., Bibb, M.J., Buttner, M.J., Leskiw, B.K. (2001) BldD is a direct regulator of key developmental genes in Streptomyces coelicolor A3(2). Molecular Microbiology. 40(1), 257–269.

Elliot, M.A., Buttner, M.J., Nodwell, J.R. (2007) Multicellular development in Streptomyces. Myxobacteria: Multicellularity and Differentiation, 419–438.

Elliot, M.A., Locke, T.R., Galibois, C.M., Leskiw, B.K. (2003) BldD from Streptomyces coelicoloris a non-essential global regulator that binds its own promoter as a dimer. FEMS Microbiology Letters. 225(1), 35–40.

Fernández-Moreno, M.A., Caballero, J., Hopwood, D.A., Malpartida, F. (1991) The act cluster contains regulatory and antibiotic export genes, direct targets for translational control by the bldA tRNA gene of Streptomyces. Cell. 66(4), 769–780.

Fernández-Moreno, M.A., Martinez, E., Caballero, J.L., Ichinose, K., Hopwood, D.A., Malpartida, F. (1994) DNA sequence and functions of the actVI region of the actinorhodin biosynthetic gene cluster of Streptomyces coelicolor A3(2). The Journal of Biological Chemistry. 269(40), 24854–24863.

Finlay, A.C., Hobby, G.L. (1950) Terramycin, a new antibiotic. Science. 111(2874), 85.

Fischer, M., Alderson, J., van Keulen, G., White, J., Sawers, R.G. (2010) The obligate aerobe Streptomyces coelicolor A3(2) synthesizes three active respiratory nitrate reductases. Microbiology. 156(10), 3166–3179.

Fischer, M., Falke, D., Sawers, R.G. (2013) A respiratory nitrate reductase active exclusively in resting spores of the obligate aerobe Streptomyces coelicolor A3(2). Molecular Microbiology. 89(6), 1259–1273.

Fischer, S., Brunk, B.P., Chen, F., Gao, X., Harb, O.S., Iodice, J.B., Shanmugam, D., Roos, D.S., Stoeckert, C.J., Jr. (2002) Using OrthoMCL to Assign Proteins to OrthoMCL-DB Groups or to Cluster Proteomes Into New Ortholog Groups. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Friend, E.J., Hopwood, D.A. (1971) The linkage map of Streptomyces rimosus. Journal of General Microbiology. 68(2), 187–197.

Fu, H.L., Meng, Y., Ordonez, E., Villadangos, A.F., Bhattacharjee, H., Gil, J.A., Mateos, L.M., Rosen, B.P. (2009) Properties of Arsenite Efflux Permeases (Acr3) from Alkaliphilus metalliredigens and Corynebacterium glutamicum. Journal of Biological Chemistry. 284(30), 19887–19895.

Fujii, T., Gramajo, H.C., Takano, E., Bibb, M.J. (1996) redD and actII-ORF4, pathway-specific regulatory genes for antibiotic production in Streptomyces coelicolor A3 (2), are transcribed in vitro by an RNA polymerase holoenzyme containing sigma hrdD. Journal of Bacteriology. 178(11), 3402–3405.

Furuya, K., Hutchinson, C.R. (1996) The DnrN protein of Streptomyces peucetius, a pseudo-response regulator, is a DNA-binding protein involved in the regulation of daunorubicin biosynthesis. Journal of Bacteriology. 178(21), 6310–6318.

Galm, U., Wang, L., Wendt-Pienkowski, E., Yang, R., Liu, W., Tao, M., Coughlin, J.M., Shen, B. (2008) In Vivo Manipulation of the Bleomycin Biosynthetic Gene Cluster in Streptomyces verticillus ATCC15003 Revealing New Insights into Its Biosynthetic Pathway. Journal of Biological Chemistry. 283(42), 28236–28245.

Galperin, M.Y. (2010) Diversity of structure and function of response regulator output domains. Current Opinion in Microbiology. 13(2), 150–159.

Garcia, I., Vior, N.M., Braña, A.F., González-Sabin, J., Rohr, J., Moris, F., Méndez, C., Salas, J.A. (2012) Elucidating the Biosynthetic Pathway for the Polyketide-Nonribosomal Peptide Collismycin A: Mechanism for Formation of the 2,2. Chemistry & Biology. 19(3), 399–413.

Garg, R.P., Parry, R.J. (2010) Regulation of valanimycin biosynthesis in Streptomyces viridifaciens: characterization of VlmI as a Streptomyces antibiotic regulatory protein (SARP). Microbiology. 156(2), 472–483.

Gay, D., You, Y.-O., Keatinge-Clay, A., Cane, D.E. (2013) Structure and Stereospecificity of the Dehydratase Domain from the Terminal Module of the Rifamycin Polyketide Synthase. Biochemistry, 131125100749007.

Ghosh, T., Bose, D., Zhang, X. (2010) Mechanisms for activating bacterial RNA polymerase. FEMS Microbiology Reviews, 34, 611–627

Gilberthorpe, N.J., Poole, R.K. (2008) Nitric oxide homeostasis in Salmonella typhimurium: roles of respiratory nitrate reductase and flavohemoglobin. The Journal of biological chemistry. 283(17), 11146–11154.

Gomez-Escribano, J.P., Martin, J.F., Hesketh, A., Bibb, M.J., Liras, P. (2008) Streptomyces clavuligerus relA-null mutants overproduce clavulanic acid and cephamycin C: negative regulation of secondary metabolism by (p)ppGpp. Microbiology. 154(3), 744–755.

Grant, J.R., Arantes, A.S., Stothard, P. (2012) Comparing thousands of circular genomes using the CGView Comparison Tool. BMC Genomics. 202(13), 2–8.

Gropp, M., Strausz, Y., Gross, M., Glaser, G. (2001) Regulation of Escherichia coli RelA Requires Oligomerization of the C-Terminal Domain. Journal of Bacteriology. 183(2), 570–579.

Gross, F., Luniak, N., Perlova, O., Gaitatzis, N., Jenke-Kodama, H., Gerth, K., Gottschalk, D., Dittmann, E., Müller, R. (2006) Bacterial type III polyketide synthases: phylogenetic analysis and potential for the production of novel secondary metabolites by heterologous expression in pseudomonads. Archives of Microbiology. 185(1), 28–38.

Guerra, S.M., Rodríguez-García, A., Santos-Aberturas, J., Vicente, C.M., Payero, T.D., Martín, J.F., Aparicio, J.F. (2012) LAL Regulators SCO0877

and SCO7173 as Pleiotropic Modulators of Phosphate Starvation Response and Actinorhodin Biosynthesis in Streptomyces coelicolor M. Otto, ed. PLoS ONE. 7(2), e31475.

Guilfoile, P.G., Hutchinson, C.R. (1992) The Streptomyces glaucescens TcmR protein represses transcription of the divergently oriented tcmR and tcmA genes by binding to an intergenic operator region. Journal of Bacteriology. 174(11), 3659–3666.

Guo, J., Zhang, X., Chen, Z., Wen, Y., Li, J. (2014) Two Adjacent and Similar TetR Family Transcriptional Regulator Genes, SAV577 and SAV576, Co-Regulate Avermectin Production in Streptomyces avermitilis. PLoS ONE. 9(6), e99224.

Guo, J., Zhang, X., Luo, S., He, F., Chen, Z., Wen, Y., Li, J. (2013) A Novel TetR Family Transcriptional Regulator, SAV576, Negatively Controls Avermectin Biosynthesis in Streptomyces avermitilis. PLoS ONE. 8(8), e71330.

Guo, X., Liu, T., Valenzano, C.R., Deng, Z., Cane, D.E. (2010) Mechanism and Stereospecificity of a Fully Saturating Polyketide Synthase Module: Nanchangmycin Synthase Module 2 and Its Dehydratase Domain. Journal of the American Chemical Society. 132(42), 14694–14696.

Gust, B., Kieser, T., Keith, C. (2003) PCR targeting system in Streptomyces coelicolor A3(2). John Innes Centre, Norwich Research Park, Colney, Norwich, 1–39.

Guzman, S. (2005) Pleiotropic effect of the SCO2127 gene on the glucose uptake, glucose kinase activity and carbon catabolite repression in Streptomyces peucetius var. caesius. Microbiology. 151(5), 1717–1723.

Guzmán, S., Ramos, I., Moreno, E., Ruiz, B., Rodríguez-Sanoja, R., Escalante, L., Langley, E., Sánchez, S. (2005) Sugar uptake and sensitivity to carbon catabolite regulation in Streptomyces peucetius var. caesius. Applied Microbiology and Biotechnology. 69(2), 200–206.

Hallam, S.J., Mincer, T.J., Schleper, C., Preston, C.M., Roberts, K., Richardson, P.M., DeLong, E.F. (2006) Pathways of Carbon Assimilation and Ammonia Oxidation Suggested by Environmental Genomic Analyses of Marine Crenarchaeota. PLoS Biology. 4(4), e95.

Halliwell, C.M., Morgan, G., Ou, C.-P., Cass, A.E.G. (2001) Introduction of a (Poly)histidine Tag in -Lactate Dehydrogenase Produces a Mixture of Active and Inactive Molecules. Analytical Biochemistry. 295(2), 257–261.

Hanel, F., Krugel, H., Fiedler, G. (1989) Arsenical Resistance of Growth and Phosphate Control of Antibiotic Biosynthesis in Streptomyces. Journal of General Microbiology. 135(3), 583–591.

Harley, C.B., Reynolds, R.P. (1987) Analysis of E. coli promoter sequences. Nucleic Acids Research. 15(5), 2343–2361.

Hawley, D.K., McClure, W.R. (1983) Compilation and analysis of Escherichia coli promoter DNA sequences. Nucleic Acids Research. 11(8), 2237–2255.

Haydock, S.F., Aparicio, J.F., Molnár, I., Schwecke, T., Khaw, L.E., König, A., Marsden, A.F., Galloway, I.S., Staunton, J., Leadlay, P.F. (1995) Divergent sequence motifs correlated with the substrate specificity of (methyl) malonyl-CoA: acyl carrier protein transacylase domains in modular polyketide synthases. FEBS Letters. 374(2), 246–248.

Hefti, M.H., Françoijs, K.-J., De Vries, S.C., Dixon, R., Vervoort, J. (2004) The PAS fold. European Journal of Biochemistry. 271(6), 1198–1208.

Heia, S., Borgos, S.E.F., Sletta, H., Escudero, L., Seco, E.M., Malpartida, F., Ellingsen, T.E., Zotchev, S.B. (2011) Initiation of Polyene Macrolide Biosynthesis: Interplay between Polyketide Synthase Domains and Modules as Revealed via Domain Swapping, Mutagenesis, and Heterologous Complementation. Applied and Environmental Microbiology. 77(19), 6982–6990.

Hengst, den, C.D.C., Tran, N.T.N., Bibb, M.J.M., Chandra, G.G., Leskiw, B.K.B., Buttner, M.J.M. (2010) Genes essential for morphological development and antibiotic production in Streptomyces coelicolor are targets of BldD during vegetative growth. Molecular Microbiology. 78(2), 361–379.

Hennecke, F., Muller, A., Meister, R., Strelow, A., Behrens, S. (2005) A ToxR-based two-hybrid system for the detection of periplasmic and cytoplasmic protein-protein interactions in Escherichia coli: minimal requirements for specific DNA binding and transcriptional activation. Protein Engineering Design and Selection. 18(10), 477–486.

Hertweck, C., Luzhetskyy, A., Rebets, Y., Bechthold, A. (2007) Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. Natural Product Reports. 24(1), 162.

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

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

Hillerich, B., Westpheling, J. (2006) A New GntR Family Transcriptional Regulator in Streptomyces coelicolor Is Required for Morphogenesis and Antibiotic Production and Controls Transcription of an ABC Transporter in

Response to Carbon Source. Journal of Bacteriology. 188(21), 7477–7487.

Hirano, S., Tanaka, K., Ohnishi, Y., Horinouchi, S. (2008) Conditionally positive effect of the TetR-family transcriptional regulator AtrA on streptomycin production by Streptomyces griseus. Microbiology (Reading, England). 154(Pt 3), 905–914.

Hitchman, T.S., Crosby, J., Byrom, K.J., Cox, R.J., Simpson, T.J. (1998) Catalytic self-acylation of type II polyketide synthase acyl carrier proteins. Chemistry & Biology. 5(1), 35–47.

Hodgson, D.A. (1982) Glucose repression of carbon uptake and metabolism in Streptomyces coelicolor A3(2) and its perturbation in mutants resistant to 2-deoxyglucose. Journal of General Microbiology. 128, 2417–2430.

Holmberg, A., Blomstergren, A., Nord, O., Lukacs, M., Lundeberg, J., Uhl n, M. (2005) The biotin-streptavidin interaction can be reversibly broken using water at elevated temperatures. Electrophoresis. 26(3), 501–510.

Hong, B., Phornphisutthimas, S., Tilley, E., Baumberg, S., McDowall, K.J. (2007) Streptomycin production by Streptomyces griseus can be modulated by a mechanism not associated with change in the adpA component of the A-factor cascade. Biotechnology Letters. 29(1), 57–64.

Hong, H.-J., Paget, M.S.B., Buttner, M.J. (2002) A signal transduction system in Streptomyces coelicolor that activates the expression of a putative cell wall glycan operon in response to vancomycin and other cell wall-specific antibiotics. Molecular Microbiology. 44(5), 1199–1211.

Horinouchi, S., Ohnishi, Y., Kang, D.K. (2001) The A-factor regulatory cascade and cAMP in the regulation of physiological and morphological development in Streptomyces griseus. Journal of Industrial Microbiology & Biotechnology. 27(3), 177–182.

Hoskisson, P.A., Rigali, S. (2009) Variation in form and function: the helix-turn-helix regulators of the GntR superfamily. Advances in Applied Microbiology. 69, 1–22.

Hoskisson, P.A., Rigali, S., Fowler, K., Findlay, K.C., Buttner, M.J. (2006) DevA, a GntR-Like Transcriptional Regulator Required for Development in Streptomyces coelicolor. Journal of Bacteriology. 188(14), 5014–5023.

Höjgård, S. (2012) Antibiotic resistance – why is the problem so difficult to solve? Infection Ecology & Epidemiology. 2(0), 397.

Hranueli, D., Pigac, J., Vesligaj, M. (1979) Characterization and persistence of actinophage RP2 isolated from Streptomyces rimosus ATCC 10970. Journal of General Microbiology. 114(2), 295–303.

Huang, J. (2001) Global analysis of growth phase responsive gene

expression and regulation of antibiotic biosynthetic pathways in Streptomyces coelicolor using DNA microarrays. Genes & Development. 15(23), 3183–3192.

Huber, J.A., Morrison, H.G., Huse, S.M., Neal, P.R., Sogin, M.L., Mark Welch, D.B. (2009) Effect of PCR amplicon size on assessments of clone library microbial diversity and community structure. Environmental Microbiology. 11(5), 1292–1302.

Hunt, A.C., Servin-Gonzalez, L., Kelemen, G.H., Buttner, M.J. (2005) The bldC Developmental Locus of Streptomyces coelicolor Encodes a Member of a Family of Small DNA-Binding Proteins Related to the DNA-Binding Domains of the MerR Family. Journal of Bacteriology. 187(2), 716–728.

Hutchings, M.I. (2004) Sensing and responding to diverse extracellular signals? Analysis of the sensor kinases and response regulators of Streptomyces coelicolor A3(2). Microbiology. 150(9), 2795–2806.

Hwang, K.-S., Kim, H.U., Charusanti, P., Palsson, B.Ø., Lee, S.Y. (2014) Biotechnology Advances. Biotechnology Advances. 32(2), 255–268.

Ishizuka, H., Horinouchi, S., Kieser, H.M., Hopwood, D.A., Beppu, T. (1992) A putative two-component regulatory system involved in secondary metabolism in Streptomyces spp. Journal of Bacteriology. 174(23), 7585–7594.

Jenke-Kodama, H., Dittmann, E. (2009) Evolution of metabolic diversity: Insights from microbial polyketide synthases. Phytochemistry. 70(15-16), 1858–1866.

Jeukens, J., Boyle, B., Kukavica-Ibrulj, I., Ouellet, M.M., Aaron, S.D., Charette, S.J., Fothergill, J.L., Tucker, N.P., Winstanley, C., Levesque, R.C. (2014) Comparative Genomics of Isolates of a Pseudomonas aeruginosa Epidemic Strain Associated with Chronic Lung Infections of Cystic Fibrosis Patients. PLoS ONE. 9(2), e87611.

Jez, J.M., Ferrer, J.L., Bowman, M.E., Austin, M.B., Schroder, J., Dixon, R.A., Noel, J.P. (2001) Structure and mechanism of chalcone synthase-like polyketide synthases. Journal of Industrial Microbiology & Biotechnology. 27(6), 393–398.

Jia, L., Tian, Y., Tan, H. (2007) SanT, a bidomain protein, is essential for nikkomycin biosynthesis of Streptomyces ansochromogenes. Biochemical and Biophysical Research Communications. 362(4), 1031–1036.

Jiang, H., Hutchinson, C.R. (2006) Feedback regulation of doxorubicin biosynthesis in Streptomyces peucetius. Research in Microbiology. 157(7), 666–674.

Jin, W., Kim, H.K., Kim, J.Y., Kang, S.G., Lee, S.H., Lee, K.J. (2004)

Cephamycin C production is regulated by relA and rsh genes in Streptomyces clavuligerus ATCC27064. Journal of Biotechnology. 114(1-2), 81–87.

Jones, G.H. (1994) Activation of ATP:GTP 3'-pyrophosphotransferase (guanosine pentaphosphate synthetase) in Streptomyces antibioticus. Journal of Bacteriology. 176(5), 1482–1487.

Jones, G.H. (1985) Regulation of Phenoxazinone Synthase Expression in Streptomyces-Antibioticus. Journal of Bacteriology. 163(3), 1215–1221.

Jones, G.H., Bibb, M.J. (1996) Guanosine pentaphosphate synthetase from Streptomyces antibioticus is also a polynucleotide phosphorylase. Journal of Bacteriology. 178(14), 4281–4288.

Kawamoto, S., Watanabe, M., Saito, N., Hesketh, A., Vachalova, K., Matsubara, K., Ochi, K. (2001) Molecular and Functional Analyses of the Gene (eshA) Encoding the 52-Kilodalton Protein of Streptomyces coelicolor A3(2) Required for Antibiotic Production. Journal of Bacteriology. 183(20), 6009–6016.

Keating, T.A., Walsh, C.T. (1999) Initiation, elongation, and termination strategies in polyketide and polypeptide antibiotic biosynthesis. Current Opinion in Chemical Biology. 3(5), 598–606.

Keatinge-Clay, A. (2008) Crystal Structure of the Erythromycin Polyketide Synthase Dehydratase. Journal of Molecular Biology. 384(4), 941–953.

Keatinge-Clay, A., Stroud, R. (2006) The Structure of a Ketoreductase Determines the Organization of the β-Carbon Processing Enzymes of Modular Polyketide Synthases. Structure. 14(4), 737–748.

Keatinge-Clay, A.T. (2007) A Tylosin Ketoreductase Reveals How Chirality Is Determined in Polyketides. Chemistry & Biology. 14(8), 898–908.

Kieser, T., Bibb, M J., Buttner, M J., Chater, K F., Hopwood, D A., (2000). Practical Streptomyces Genetics. Norwich, UK: The John Innes Foundation

Kern, A., Tilley, E., Hunter, I.S., Legiša, M., Glieder, A. (2007) Engineering primary metabolic pathways of industrial micro-organisms. Journal of Biotechnology. 129(1), 6–29.

Kerry, J., Hiney, M., Coyne, R., NicGabhainn, S., Gilroy, D., Cazabon, D., Smith, P. (1995) Fish feed as a source of oxytetracycline-resistant bacteria in the sediments under fish farms. Aquaculture. 131(1), 101–113.

Khaliq, S., Ghauri, M.A., Akhtar, K. (2014) Characterization of mutations in regulatory genes of Tyl cluster leading to overexpression of tylosin in mutant γ-1 of Streptomyces fradiae NRRL-2702. Applied Microbiology and Biotechnology. 98(2), 785–793.

Khosla, C. (2009) Structures and Mechanisms of Polyketide Synthases. The Journal of Organic Chemistry. 74(17), 6416–6420.

Kim, E.S., Hong, H.J., Choi, C.Y., Cohen, S.N. (2001) Modulation of Actinorhodin Biosynthesis in Streptomyces lividans by Glucose Repression of afsR2 Gene Transcription. Journal of Bacteriology. 183(7), 2198–2203.

Kim, H., Abeysirigunawarden, S.C., Chen, K., Mayerle, M., Ragunathan, K., Luthey-Schulten, Z., Ha, T., Woodson, S.A. (2014) Protein-guided RNA dynamics during early ribosome assembly. Nature. 506(7488), 334–338.

Kirby, R., Gan, T.-K., Hunter, I., Herron, P., Tilley, E. (2008) The genome of Streptomyces rimosus subsp. rimosus shows a novel structure compared to other Streptomyces using DNA/DNA microarray analysis. Antonie van Leeuwenhoek. 94(2), 173–186.

Kitani, S., Ikeda, H., Sakamoto, T., Noguchi, S., Nihira, T. (2009) Characterization of a regulatory gene, aveR, for the biosynthesis of avermectin in Streptomyces avermitilis. Applied Microbiology and Biotechnology. 82(6), 1089–1096.

Klassen, J.L., Adams, S.M., Bramhacharya, S., Giles, S.S., Goodwin, L.A., Woyke, T., Currie, C.R. (2011) Draft Genome Sequence of Streptomyces sp. Strain Wigar10, Isolated from a Surface-Sterilized Garlic Bulb. Journal of Bacteriology. 193(24), 6999–7000.

Knaggs, A.R., Knaggs, A.R., Knaggs, A.R. (1999) The biosynthesis of shikimate metabolites. Natural Product Reports. 16(4), 525–560.

Knirschová, R., Novakova, R., Feckova, L., Timko, J., Turna, J., Bistáková, J., Kormanec, J. (2007) Multiple regulatory genes in the salinomycin biosynthetic gene cluster of Streptomyces albus CCM 4719. Folia Microbiologica. 52(4), 359–365.

Koboldt, D.C., Chen, K., Wylie, T., Larson, D.E., McLellan, M.D., Mardis, E.R., Weinstock, G.M., Wilson, R.K., Ding, L. (2009) VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics. 25(17), 2283–2285.

Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., McLellan, M.D., Lin, L., Miller, C.A., Mardis, E.R., Ding, L., Wilson, R.K. (2012) VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Research. 22(3), 568–576.

Kong, D., Lee, M.-J., Lin, S., Kim, E.-S. (2013) Biosynthesis and pathway engineering of antifungal polyene macrolides in actinomycetes. Journal of Industrial Microbiology & Biotechnology. 40(6), 529–543.

Kuščer, E., Coates, N., Challis, I., Gregory, M., Wilkinson, B., Sheridan, R., Petković, H. (2007) Roles of rapH and rapG in Positive Regulation of

Rapamycin Biosynthesis in Streptomyces hygroscopicus. Journal of Bacteriology. 189(13), 4756–4763.

Kwak, J., McCue, L.A., Trczianka, K., Kendrick, K.E. (2001) Identification and Characterization of a Developmentally Regulated Protein, EshA, Required for Sporogenic Hyphal Branches in Streptomyces griseus. Journal of Bacteriology. 183(10), 3004–3015.

Kwakman, J.H., Postma, P.W. (1994) Glucose kinase has a regulatory role in carbon catabolite repression in Streptomyces coelicolor. Journal of Bacteriology. 176(9), 2694–2698.

Kwan, D.H., Leadlay, P.F. (2010) Mutagenesis of a Modular Polyketide Synthase Enoylreductase Domain Reveals Insights into Catalysis and Stereospecificity. ACS Chemical Biology. 5(9), 829–838.

Kwan, D.H., Sun, Y., Schulz, F., Hong, H., Popovic, B., Sim-Stark, J.C.C., Haydock, S.F., Leadlay, P.F. (2008) Prediction and Manipulation of the Stereochemistry of Enoylreduction in Modular Polyketide Synthases. Chemistry & Biology. 15(11), 1231–1240.

Labeda, D.P., Goodfellow, M., Brown, R., Ward, A.C., Lanoot, B., Vanncanneyt, M., SWINGS, J., Kim, S.B., Liu, Z., Chun, J., Tamura, T., Oguchi, A., Kikuchi, T., Kikuchi, H., Nishii, T., Tsuji, K., Yamaguchi, Y., Tase, A., Takahashi, M., Sakane, T., Suzuki, K.I., Hatano, K. (2011) Phylogenetic study of the species within the family Streptomycetaceae. Antonie van Leeuwenhoek. 101(1), 73–104.

Lal, R., Khanna, R., Kaur, H., Khanna, M., Dhingra, N., Lal, S., Gartemann, K.H., Eichenlaub, R., Ghosh, P.K. (1996) Engineering antibiotic producers to overcome the limitations of classical strain improvement programs. Critical Reviews in Microbiology. 22(4), 201–255.

Lal, R., Kumari, R., Kaur, H., Khanna, R., Dhingra, N., Tuteja, D. (2000) Regulation and manipulation of the gene clusters encoding type-I PKSs. Trends in Biotechnology. 18(6), 264–274.

Lanciano, P., Vergnes, A., Grimaldi, S., Guigliarelli, B., Magalon, A. (2007) Biogenesis of a respiratory complex is orchestrated by a single accessory protein. The Journal of Biological Chemistry. 282(24), 17468–17474.

Lawal, O.S. (2006) Kosmotropes and chaotropes as they affect functionality of a protein isolate. Food Chemistry. 95(1), 101–107.

Lefébure, T., Stanhope, M.J. (2007) Evolution of the core and pan-genome of Streptococcus: positive selection, recombination, and genome composition. Genome Biology. 8(5), R71.

Leibly, D.J., Nguyen, T.N., Kao, L.T., Hewitt, S.N., Barrett, L.K., Van Voorhis, W.C. (2012) Stabilizing Additives Added during Cell Lysis Aid in the

Solubilization of Recombinant Proteins. PLoS ONE. 7(12), e52482.

Lenik, U., Gormand, A., Magdevska, V., Fujs, S., Raspor, P., Hunter, I., Petkovic, H. (2009) Regulatory Elements in Tetracycline-Encoding Gene Clusters: the otcG Gene Positively Regulates the Production of Oxytetracycline in Streptomyces rimosus. Food Technology. Biotechnology. 47 (3), 323–330.

Letunic, I., Doerks, T., Bork, P. (2011) SMART 7: recent updates to the protein domain annotation resource. Nucleic Acids Research. 40(D1), D302–D305.

Li, H., Durbin, R. (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 26(5), 589–595.

Li, H., Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 25(14), 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics. 25(16), 2078–2079.

Li, H.-W., Zhi, X.-Y., Yao, J.-C., Zhou, Y., Tang, S.-K., Klenk, H.-P., Zhao, J., Li, W.-J. (2013) Comparative Genomic Analysis of the Genus Nocardiopsis Provides New Insights into Its Genetic Mechanisms of Environmental Adaptability C. Brochier-Armanet, ed. PLoS ONE. 8(4), e61528.

Li, L., Stoeckert, C.J., Roos, D.S. (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Research. 13(9), 2178–2189.

Li, W., Atkinson, G.C., Thakor, N.S., Allas, Ü., Lu, C.-C., Chan, K.-Y., Tenson, T., Schulten, K., Wilson, K.S., Hauryliuk, V., Frank, J. (2013) Mechanism of tetracycline resistance by ribosomal protection protein Tet(O). Nature Communications. 4, 1477.

Liew, C.W., Nilsson, M., Chen, M.W., Sun, H., Cornvik, T., Liang, Z.X., Lescar, J. (2012) Crystal Structure of the Acyltransferase Domain of the Iterative Polyketide Synthase in Enediyne Biosynthesis. Journal of Biological Chemistry. 287(27), 23203–23215.

Liu, G., Chater, K.F., Chandra, G., Niu, G., Tan, H. (2013) Molecular Regulation of Antibiotic Biosynthesis in Streptomyces. Microbiology and Molecular Biology Reviews. 77(1), 112–143.

Loman, N.J., Constantinidou, C., Chan, J.Z.M., Halachev, M., Sergeant, M., Penn, C.W., Robinson, E.R., Pallen, M.J. (2012) High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. Nature Reviews Microbiology. 10(9), 599–606.

Longo, F., Rampioni, G., Bondì, R., Imperi, F., Fimia, G.M., Visca, P., Zennaro, E., Leoni, L. (2013) A New Transcriptional Repressor of the Pseudomonas aeruginosa Quorum Sensing Receptor Gene lasR T. Coenye, ed. PLoS ONE. 8(7), e69554.

Lu, Y., Wang, W., Shu, D., Zhang, W., Chen, L., Qin, Z., Yang, S., Jiang, W. (2007) Characterization of a novel two-component regulatory system involved in the regulation of both actinorhodin and a type I polyketide in Streptomyces coelicolor. Applied Microbiology and Biotechnology. 77(3), 625–635.

MacGregor-Pryde, S.E. (1995) A molecular investigation of the otrB locus of Streptomyces rimosus. University of Glasgow

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

Manteca, A., Alvarez, R., Salazar, N., Yague, P., Sanchez, J. (2008) Mycelium Differentiation and Antibiotic Production in Submerged Cultures of Streptomyces coelicolor. Applied and Environmental Microbiology. 74(12), 3877–3886.

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

Martínez, L.F., Bishop, A., Parkes, L., Del Sol, R., Salerno, P., Sevcikova, B., Mazurakova, V., Kormanec, J., Dyson, P. (2009) Osmoregulation in Streptomyces coelicolor: modulation of SigB activity by OsaC. Molecular Microbiology. 71(5), 1250–1262.

Martínez-Hackert, E., Stock, A.M. (1997) The DNA-binding domain of OmpR: crystal structures of a winged helix transcription factor. Structure/Folding and Design. 5(1), 109–124.

Marton, L.S., Gulcher, J.R., Stefansson, K. (1989) Binding of hexabrachions to heparin and DNA. The Journal of Biological Chemistry. 264(22), 13145–13149.

Martín, J.-F., Liras, P. (2010) Engineering of regulatory cascades and networks controlling antibiotic biosynthesis in Streptomyces. Current Opinion in Microbiology. 13(3), 263–273.

Mast, Y., Weber, T., Gölz, M., Ort-Winklbauer, R., Gondran, A., Wohlleben, W., Schinko, E. (2010) Characterization of the 'pristinamycin supercluster' of Streptomyces pristinaespiralis. Microbial Biotechnology. 4(2), 192–206.

Mayerle, M., Woodson, S.A. (2013) Specific contacts between protein S4 and ribosomal RNA are required at multiple stages of ribosome assembly.

RNA. 19(4), 574–585.

Mazodier, P., Petter, R., Thompson, C. (1989) Intergeneric conjugation between Escherichia coli and Streptomyces species. Journal of Bacteriology. 171(6), 3583–3585.

McCleary, W.R., Stock, J.B., Ninfa, A.J. (1993) Is acetyl phosphate a global signal in Escherichia coli? Journal of Bacteriology. 175(10), 2793–2798.

McCormick, J.R., Flardh, K. (2011) Signals and regulators that govern Streptomyces development. FEMS Microbiology Reviews. 36(1), 206–231.

McCue, L.A. (2000) Functional Classification of cNMP-binding Proteins and Nucleotide Cyclases with Implications for Novel Regulatory Pathways in Mycobacterium tuberculosis. Genome Research. 10(2), 204–219.

McDowall, K.J., Thamchaipenet, A., Hunter, I.S. (1999) Phosphate control of oxytetracycline production by Streptomyces rimosus is at the level of transcription from promoters overlapped by tandem repeats similar to those of the DNA-binding sites of the OmpR family. Journal of Bacteriology. 181(10), 3025–3032.

McManus, P.S., Stockwell, V.O., Sundin, G.W., Jones, A.L. (2002) Antibiotic use in plant agriculture. Annual Review of Phytopathology. 40(1), 443–465.

Medema, M.H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M.A., Weber, T., Takano, E., Breitling, R. (2011) antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Research. 39(suppl), W339–W346.

Mendes, M.V., Recio, E., Antón, N., Guerra, S.M., Santos-Aberturas, J., Martín, J.F., Aparicio, J.F. (2007) Cholesterol Oxidases Act as Signaling Proteins for the Biosynthesis of the Polyene Macrolide Pimaricin. Chemistry & Biology. 14(3), 279–290.

Mendes, M.V., Tunca, S., Antón, N., Recio, E., Sola-Landa, A., Aparicio, J.F., Martín, J.F. (2007) The two-component phoR-phoP system of Streptomyces natalensis: Inactivation or deletion of phoP reduces the negative phosphate regulation of pimaricin biosynthesis. Metabolic Engineering. 9(2), 217–227.

Meng, Y.-L., Liu, Z., Rosen, B.P. (2004) As(III) and Sb(III) uptake by GlpF and efflux by ArsB in Escherichia coli. The Journal of Biological Chemistry. 279(18), 18334–18341.

Merrick, M.J. (1976) A morphological and genetic mapping study of bald colony mutants of Streptomyces coelicolor. Journal of General Microbiology. 96(2), 299–315.

Messer, W., Bellekes, U., Lother, H. (1985) Effect of dam methylation on the

activity of the E. coli replication origin, oriC. The EMBO Journal. 4(5), 1327–1332.

Méric, G., Yahara, K., Mageiros, L., Pascoe, B., Maiden, M.C.J., Jolley, K.A., Sheppard, S.K. (2014) A Reference Pan-Genome Approach to Comparative Bacterial Genomics: Identification of Novel Epidemiological Markers in Pathogenic Campylobacter. PLoS ONE. 9(3), e92798.

Michaux, C., Pomroy, N.C., Privé, G.G. (2008) Refolding SDS-Denatured Proteins by the Addition of Amphipathic Cosolvents. Journal of Molecular Biology. 375(5), 1477–1488.

Molzahn, S.W., Woods, R.A. (1972) Polyene resistance and the isolation of sterol mutants in Saccharomyces cerevisiae. Journal of General Microbiology. 72(2), 339–348.

Moore, B.S., Hertweck, C. (2002) Biosynthesis and attachment of novel bacterial polyketide synthase starter units. Natural Product Reports. 19(1), 70–99.

Narayanan, A., Paul, L.N., Tomar, S., Patil, D.N., Kumar, P., Yernool, D.A. (2012) Structure-Function Studies of DNA Binding Domain of Response Regulator KdpE Reveals Equal Affinity Interactions at DNA Half-Sites. PLoS ONE. 7(1), e30102.

Navarro Llorens, J.M., Tormo, A., Martínez-García, E. (2010) Stationary phase in gram-negative bacteria. FEMS Microbiology Reviews. 34(4), 476–495.

Niu, G., Li, L., Wei, J., Tan, H. (2013) Cloning, Heterologous Expression,and Characterization of the Gene Cluster Required for Gougerotin Biosynthesis. Chemistry & Biology. 20(1), 34–44.

Nothaft, H., Rigali, S., Boomsma, B., Swiatek, M., McDowall, K.J., van Wezel, G.P., Titgemeyer, F. (2010) The permease gene nagE2is the key to N-acetylglucosamine sensing and utilization in Streptomyces coelicolorand is subject to multi-level control. Molecular Microbiology. 75(5), 1133–1144.

Novakova, R., Rehakova, A., Kutas, P., Feckova, L., Kormanec, J. (2011) The role of two SARP family transcriptional regulators in regulation of the auricin gene cluster in Streptomyces aureofaciens CCM 3239. Microbiology. 157(6), 1629–1639.

Nowotny, V., Nierhaus, K.H. (1988) Assembly of the 30S subunit from Escherichia coli ribosomes occurs via two assembly domains which are initiated by S4 and S7. Biochemistry. 27(18), 7051–7055.

Nyborg, J.K., Dynan, W.S., Chen, I.S., Wachsman, W. (1988) Binding of host-cell factors to DNA sequences in the long terminal repeat of human T-cell leukemia virus type I: implications for viral gene expression. Proceedings

of the National Academy of Sciences of the United States of America. 85(5), 1457–1461.

Ohnishi, Y., Yamazaki, H., Kato, J.-Y., Tomono, A., Horinouchi, S. (2005) AdpA, a central transcriptional regulator in the A-factor regulatory cascade that leads to morphological development and secondary metabolism in Streptomyces griseus. Bioscience, Biotechnology, and Biochemistry. 69(3), 431–439.

Ohnuki, T., Imanaka, T., Aiba, S. (1985) Self-cloning in Streptomyces griseus of an str gene cluster for streptomycin biosynthesis and streptomycin resistance. Journal of Bacteriology. 164(1), 85–94.

Okamoto, S., Taguchi, T., Ochi, K., Ichinose, K. (2009) Biosynthesis of Actinorhodin and Related Antibiotics: Discovery of Alternative Routes for Quinone Formation Encoded in the act Gene Cluster. Chemistry & Biology. 16(2), 226–236.

Olano, C., Cano-Prieto, C., Losada, A.A., Bull, A.T., Goodfellow, M., Fiedler, H.P., Mendez, C., Salas, J.A. (2014) Draft Genome Sequence of Marine Actinomycete Streptomyces sp. Strain NTK 937, Producer of the Benzoxazole Antibiotic Caboxamycin. Genome Announcements. 2(4), e00534–14.

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. Nature Biotechnology. 25(4), 447–453.

Oren, A., Garrity, G.M. (2013) Then and now: a systematic review of the systematics of prokaryotes in the last 80 years. Antonie van Leeuwenhoek. 106(1), 43–56.

Otten, S.L., Olano, C., Hutchinson, C.R. (2000) The dnrO gene encodes a DNA-binding protein that regulates daunorubicin production in Streptomyces peucetius by controlling expression of the dnrN pseudo response regulator gene. Microbiology (Reading, England). 146 (6), 1457–1468.

Overbeek, R. (2005) The Subsystems Approach to Genome Annotation and its Use in the Project to Annotate 1000 Genomes. Nucleic Acids Research. 33(17), 5691–5702.

Paget, M.S., Leibovitz, E., Buttner, M.J. (1999) A putative two-component signal transduction system regulates sigmaE, a sigma factor required for normal cell wall integrity in Streptomyces coelicolor A3(2). Molecular Microbiology. 33(1), 97–107.

Pandey, R.C., Rinehart, K.L. (1977) Polyene antibiotics. VIII. The structure of rimocidin. The Journal of Antibiotics. 30(2), 146–157.

Pandza, K., Pfalzer, G., Cullum, J., Hranueli, D. (1997) Physical mapping shows that the unstable oxytetracycline gene cluster of Streptomyces rimosus lies close to one end of the linear chromosome. Microbiology. 143 (5), 1493–1501.

Park, S.-S., Yang, Y.-H., Song, E., Kim, E.-J., Kim, W.S., Sohng, J.K., Lee, H.C., Liou, K.K., Kim, B.-G. (2009) Mass spectrometric screening of transcriptional regulators involved in antibiotic biosynthesis in Streptomyces coelicolor A3(2). Journal of Industrial Microbiology & Biotechnology. 36(8), 1073–1083.

Park, S.R., Yoo, Y.J., Ban, Y.H., Yoon, Y.J. (2010) Biosynthesis of rapamycin and its regulation: past achievements and recent progress. Journal of Antibiotics. 63(8), 434–441.

Paulsen, I.T., Brown, M.H., Skurray, R.A. (1996) Proton-dependent multidrug efflux systems. Microbiological Reviews. 60(4), 575–608.

Peric-Concha, N., Borovicka, B., Long, P.F., Hranueli, D., Waterman, P.G., Hunter, I.S. (2005) Ablation of the otcC gene encoding a post-polyketide hydroxylase from the oxytetracyline biosynthetic pathway in Streptomyces rimosus results in novel polyketides with altered chain length. The Journal of biological chemistry. 280(45), 37455–37460.

Pethick, F.E., MacFadyen, A.C., Tang, Z., Sangal, V., Liu, T.T., Chu, J., Kosec, G., Petković, H., Guo, M., Kirby, R., Hoskisson, P.A., Herron, P.R., Hunter, I.S. (2013) Draft Genome Sequence of the Oxytetracycline-Producing Bacterium Streptomyces rimosus ATCC 10970. Genome Announcements. 1(2), e00063–13.

Petković, H., Cullum, J., Hranueli, D., Hunter, I.S., Peric-Concha, N., Pigac, J., Thamchaipenet, A., Vujaklija, D., Long, P.F. (2006) Genetics of Streptomyces rimosus, the Oxytetracycline Producer. Microbiology and Molecular Biology Reviews. 70(3), 704–728.

Petković, H., Thamchaipenet, A., Zhou, L.H., Hranueli, D., Raspor, P., Waterman, P.G., Hunter, I.S. (1999) Disruption of an aromatase/cyclase from the oxytetracycline gene cluster of Streptomyces rimosus results in production of novel polyketides with shorter chain lengths. The Journal of biological chemistry. 274(46), 32829–32834.

Phornphisutthimas, S., Sudtachat, N., Bunyoo, C., Chotewutmontri, P., Panijpan, B., Thamchaipenet, A. (2010) Development of an intergeneric conjugal transfer system for rimocidin-producing Streptomyces rimosus. Letters in Applied Microbiology. 50(5), 530–536.

Pickens, L.B., Kim, W., Wang, P., Zhou, H., Watanabe, K., Gomi, S., Tang, Y. (2009) Biochemical Analysis of the Biosynthetic Pathway of an Anticancer Tetracycline SF2575. Journal of the American Chemical Society. 131(48),

17677–17689.

Pickens, L.B., Tang, Y. (2009) Decoding and engineering tetracycline biosynthesis. Metabolic Engineering. 11(2), 69–75.

Pickens, L.B., Tang, Y. (2010) Oxytetracycline Biosynthesis. Journal of Biological Chemistry. 285(36), 27509–27515.

Pigac, J., Schrempf, H. (1995) A Simple and Rapid Method of Transformation of Streptomyces rimosus R6 and Other Streptomycetes by Electroporation. Applied and Environmental Microbiology. 61, 352–356.

Pluskal, T., Castillo, S., Villar-Briones, A., Orešič, M. (2010) MZmine 2: Modular framework for processing,visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics. 11(1), 395.

Polz, M.F., Cavanaugh, C.M. (1998) Bias in template-to-product ratios in multitemplate PCR. Applied and Environmental Microbiology. 64(10), 3724–3730.

Ponting, C.P., Aravind, L. (1997) PAS: a multifunctional domain family comes to light. Current Biology. 7(11), R674–677.

Powers, J.H. (2004) Antimicrobial drug development--the past, the present, and the future. Clinical Microbiology and Infection. 10 Suppl 4, 23–31.

Powers, T., Noller, H.F. (1995) A temperature-dependent conformational rearrangement in the ribosomal protein S4.16 S rRNA complex. The Journal of biological chemistry. 270(3), 1238–1242.

Prasanna, A.N., Mehra, S. (2013) Comparative Phylogenomics of Pathogenic and Non-Pathogenic Mycobacterium. PLoS ONE. 8(8), –e71248.

Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., Heger, A., Holm, L., Sonnhammer, E.L.L., Eddy, S.R., Bateman, A., Finn, R.D. (2011) The Pfam protein families database. Nucleic Acids Research. 40(D1), D290–D301.

Ramos, I., Guzmán, S., Escalante, L., Imriskova, I., Rodríguez-Sanoja, R., Sánchez, S., Langley, E. (2004) Glucose kinase alone cannot be responsible for carbon source regulation in Streptomyces peucetius var. caesius. Research in Microbiology. 155(4), 267–274.

Ramos, J.L., Martínez-Bueno, M., Molina-Henares, A.J., Terán, W., Watanabe, K., Zhang, X., Gallegos, M.T., Brennan, R., Tobes, R. (2005) The TetR family of transcriptional repressors. Microbiology and Molecular Biology Reviews. 69(2), 326–356.

Rausch, H., Vesligaj, M., Pocta, D., Biuković, G., Pigac, J., Cullum, J., Schmieger, H., Hranueli, D. (1993) The temperate phages RP2 and RP3 of

Streptomyces rimosus. Journal of General Microbiology. 139(10), 2517–2524.

Rawlings, B.J. (2001a) Type I polyketide biosynthesis in bacteria (Part A — erythromycin biosynthesis) (1994 to 2000). Natural Product Reports. 18(2), 190–227.

Rawlings, B.J. (2001b) Type I polyketide biosynthesis in bacteria (Part B) (1995 to mid-2000). Natural Product Reports. 18(3), 231–281.

Ren, J., Cui, Y., Zhang, F., Cui, H., Ni, X., Chen, F., Li, L., Xia, H. (2013) Enhancement of nystatin production by redirecting precursor fluxes after disruption of the tetramycin gene from Streptomyces ahygroscopicus. Microbiological Research, 1–7.

Reumerman, R.A., Tucker, N.P., Herron, P.R., Hoskisson, P.A., Sangal, V. (2013) Tool for rapid annotation of microbial SNPs (TRAMS): a simple program for rapid annotation of genomic variation in prokaryotes. Antonie van Leeuwenhoek. 104(3), 431–434.

Richardson, D.J., Berks, B.C., Russell, D.A., Spiro, S., Taylor, C.J. (2001) Functional, biochemical and genetic diversity of prokaryotic nitrate reductases. Cellular and Molecular Life Sciences. 58(2), 165–178.

Rigali, S., Nothaft, H., Noens, E.E.E., Schlicht, M., Colson, S., Muller, M., Joris, B., Koerten, H.K., Hopwood, D.A., Titgemeyer, F., 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. Molecular Microbiology. 61(5), 1237–1251.

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

Robinson, J.A. (1991) Polyketide Synthase Complexes: Their Structure and Function in Antibiotic Biosynthesis. Philosophical Transactions of the Royal Society B: Biological Sciences. 332(1263), 107–114.

Rodríguez, H., Rico, S., Díaz, M., Santamaría, R.I. (2013) Two-component systems in Streptomyces: key regulators of antibiotic complex pathways. Microbial Cell Factories. 12, 127.

Rodríguez-García, A., Santamarta, I., Pérez-Redondo, R., Martín, J.F., Liras, P. (2006) Characterization of a two-gene operon epeRA involved in multidrug resistance in Streptomyces clavuligerus. Research in Microbiology. 157(6), 559–568.

Rokem, J.S., Lantz, A.E., Nielsen, J. (2007) Systems biology of antibiotic

production by microorganisms. Natural Product Reports. 24(6), 1262.

Rong, X., Doroghazi, J.R., Cheng, K., Zhang, L., Buckley, D.H., Huang, Y. (2013) Classification of Streptomyces phylogroup pratensis (Doroghazi and Buckley, 2010) based on genetic and phenotypic evidence, and proposal of Streptomyces pratensis sp. nov. Systematic and Applied Microbiology. 36(6), 401–407.

Rong, X., Huang, Y. (2012) Systematic and Applied Microbiology. Systematic and Applied Microbiology. 35(1), 7–18.

Ruiz, B., Chávez, A., Forero, A., García-Huante, Y., Romero, A., Sánchez, M., Rocha, D., Sánchez, B., Rodríguez-Sanoja, R., Sánchez, S., Langley, E. (2010) Production of microbial secondary metabolites: Regulation by the carbon source. Critical Reviews in Microbiology. 36(2), 146–167.

Ryu, Y.G., Butler, M.J., Chater, K.F., Lee, K.J. (2006) Engineering of Primary Carbohydrate Metabolism for Increased Production of Actinorhodin in Streptomyces coelicolor. Applied and Environmental Microbiology. 72(11), 7132–7139.

Sadeghi, A., Soltani, B.M., Nekouei, M.K., Jouzani, G.S., Mirzaei, H.H., Sadeghizadeh, M. (2014) Microbiological Research. Microbiological Research. 169(9-10), 699–708.

Saito, A., Fujii, T., Yoneyama, T., Miyashita, K. (1998) glkA is involved in glucose repression of chitinase production in Streptomyces lividans. Journal of Bacteriology. 180(11), 2911–2914.

Saito, N. (2002) Genetic and Biochemical Characterization of EshA, a Protein That Forms Large Multimers and Affects Developmental Processes in Streptomyces griseus. Journal of Biological Chemistry. 278(8), 5902–5911.

Saito, N., Xu, J., Hosaka, T., Okamoto, S., Aoki, H., Bibb, M.J., Ochi, K. (2006) EshA Accentuates ppGpp Accumulation and Is Conditionally Required for Antibiotic Production in Streptomyces coelicolor A3(2). Journal of Bacteriology. 188(13), 4952–4961.

Sajish, M., Kalayil, S., Verma, S.K., Nandicoori, V.K., Prakash, B. (2009) The Significance of EXDD and RXKD Motif Conservation in Rel Proteins. Journal of Biological Chemistry. 284(14), 9115–9123.

Sanders, D.A., Gillece-Castro, B.L., Stock, A.M., Burlingame, A.L., Koshland, D.E. (1989) Identification of the site of phosphorylation of the chemotaxis response regulator protein, CheY. The Journal of biological chemistry. 264(36), 21770–21778.

Santos, C.L., Correia-Neves, M., Moradas-Ferreira, P., Mendes, M.V. (2012) A Walk into the LuxR Regulators of Actinobacteria: Phylogenomic

Distribution and Functional Diversity J. H. Badger, ed. PLoS ONE. 7(10), e46758.

Santos-Aberturas, J., Payero, T.D., Vicente, C.M., Guerra, S.M., Cañibano, C., Martín, J.F., Aparicio, J.F. (2011) Functional conservation of PAS–LuxR transcriptional regulators in polyene macrolide biosynthesis. Metabolic Engineering. 13(6), 756–767.

Santos-Aberturas, J., Vicente, C.M., Guerra, S.M., Payero, T.D., Martin, J.F., Aparicio, J.F. (2011) Molecular Control of Polyene Macrolide Biosynthesis: Direct Binding of the Regulator PimM to Eight Promoters of Pimaricin Genes and Identification of Binding Boxes. Journal of Biological Chemistry. 286(11), 9150–9161.

Santos-Aberturas, J., Vicente, C.M., Payero, T.D., Martín-Sánchez, L., Cañibano, C., Martín, J.F., Aparicio, J.F. (2012) Hierarchical Control on Polyene Macrolide Biosynthesis: PimR Modulates Pimaricin Production via the PAS-LuxR Transcriptional Activator PimM. PLoS ONE. 7(6), e38536.

Sánchez, S., Chávez, A., Forero, A., García-Huante, Y., Romero, A., Sánchez, M., Rocha, D., Sánchez, B., Avalos, M., Guzmán-Trampe, S., Rodríguez-Sanoja, R., Langley, E., Ruiz, B. (2010) Carbon source regulation of antibiotic production. Journal of Antibiotics. 63(8), 442–459.

Sánchez-Riego, A.M., López-Maury, L., Florencio, F.J. (2014) Genomic Responses to Arsenic in the Cyanobacterium Synechocystis sp. PCC 6803. PLoS ONE. 9(5), e96826.

Schneider, G. (2005) Enzymes in the biosynthesis of aromatic polyketide antibiotics. Current Opinion in Structural Biology. 15(6), 629–636.

Seco, E.M., Pérez-Zúñiga, F.J., Rolón, M.S., Malpartida, F. (2004) Starter Unit Choice Determines the Production of Two Tetraene Macrolides, Rimocidin and CE-108, in Streptomyces diastaticus var. 108. Chemistry & Biology. 11(3), 357–366.

Seipke, R.F., Crossman, L., Drou, N., Heavens, D., Bibb, M.J., Caccamo, M., Hutchings, M.I. (2011) Draft Genome Sequence of Streptomyces Strain S4, a Symbiont of the Leaf-Cutting Ant Acromyrmex octospinosus. Journal of Bacteriology. 193(16), 4270–4271.

Sekurova, O.N., Brautaset, T., Sletta, H., Borgos, S.E.F., Jakobsen, O.M., Ellingsen, T.E., Strom, A.R., Valla, S., Zotchev, S.B. (2004) In Vivo Analysis of the Regulatory Genes in the Nystatin Biosynthetic Gene Cluster of Streptomyces noursei ATCC 11455 Reveals Their Differential Control Over Antibiotic Biosynthesis. Journal of Bacteriology. 186(5), 1345–1354.

Sheldon, P.J., Busarow, S.B., Hutchinson, C.R. (2002) Mapping the DNA-binding domain and target sequences of the Streptomyces peucetius

daunorubicin biosynthesis regulatory protein, DnrI. Molecular Microbiology. 44(2), 449–460.

Shen, X.-L., Dong, H.-J., Hou, X.-P., Guan, W.-J., Li, Y.-Q. (2007) FtsY Affects Sporulation and Antibiotic Production by whiH in Streptomyces coelicolor. Current Microbiology. 56(1), 61–65.

Shen, Y., Yoon, P., Yu, T.W., Floss, H.G., Hopwood, D., Moore, B.S. (1999) Ectopic expression of the minimal whiE polyketide synthase generates a library of aromatic polyketides of diverse sizes and shapes. Proceedings of the National Academy of Sciences of the United States of America. 96(7), 3622–3627.

Shimada, T., Katayama, Y., Kawakita, S., Ogasawara, H., Nakano, M., Yamamoto, K., Ishihama, A. (2012) A novel regulator RcdA of the csgDgene encoding the master regulator of biofilm formation in Escherichia coli. Microbiology Open. 1(4), 381–394.

Shin, S.C., Ahn, D.H., Kim, S.J., Lee, H., Oh, T.-J., Lee, J.E., Park, H. (2013) Advantages of Single-Molecule Real-Time Sequencing in High-GC Content Genomes. PLoS ONE. 8(7), e68824.

Shu, D., Chen, L., Wang, W., Yu, Z., Ren, C., Zhang, W., Yang, S., Lu, Y., Jiang, W. (2008) afsQ1-Q2-sigQ is a pleiotropic but conditionally required signal transduction system for both secondary metabolism and morphological development in Streptomyces coelicolor. Applied Microbiology and Biotechnology. 81(6), 1149–1160.

Siedenburg, G., Jendrossek, D. (2011) Squalene-Hopene Cyclases. Applied and Environmental Microbiology. 77(12), 3905–3915.

Siguier, P., Gourbeyre, E., Chandler, M. (2014) Bacterial insertion sequences: their genomic impact and diversity. FEMS Microbiology Reviews, 38(5), 865-91.

Skoulakis, S., Goodfellow, J.M. (2003) The pH-dependent stability of wild-type and mutant transthyretin oligomers. Biophysical Journal. 84(5), 2795–2804.

Smanski, M.J., Casper, J., Peterson, R.M., Yu, Z., Rajski, S.R., Shen, B. (2012) Expression of the Platencin Biosynthetic Gene Cluster in Heterologous Hosts Yielding New Platencin Congeners. Journal of Natural Products. 75(12), 2158–2167.

Smanski, M.J., Peterson, R.M., Rajski, S.R., Shen, B. (2009) Engineered Streptomyces platensis Strains That Overproduce Antibiotics Platensimycin and Platencin. Antimicrobial Agents and Chemotherapy. 53(4), 1299–1304.

Smith, M.C., Burns, R.N., Wilson, S.E., Gregory, M.A. (1999) The complete genome sequence of the Streptomyces temperate phage straight phiC31:

evolutionary relationships to other viruses. Nucleic Acids Research. 27(10), 2145–2155.

So, A.D., Shah, T.A. (2014) New business models for antibiotic innovation. Upsala Journal of Medical Sciences. 119(2), 176–180.

Sola-Landa, A., Moura, R.S., Martin, J.F. (2003) The two-component PhoR-PhoP system controls both primary metabolism and secondary metabolite biosynthesis in Streptomyces lividans. Proceedings of the National Academy of Sciences of the United States of America. 100(10), 6133–6138.

Song, E., Rajesh, T., Lee, B.-R., Kim, E.-J., Jeon, J.-M., Park, S.-H., Park, H.-Y., Choi, K.-Y., Kim, Y.-G., Yang, Y.-H., Kim, B.-G. (2013) Deletion of an architectural unit, leucyl aminopeptidase (SCO2179), in Streptomyces coelicolor increases actinorhodin production and sporulation. Applied Microbiology and Biotechnology. 97(15), 6823–6833.

Srivatsan, A., Wang, J.D. (2008) Control of bacterial transcription, translation and replication by (p)ppGpp. Current Opinion in Microbiology. 11(2), 100–105.

Stack, D., Neville, C., Doyle, S. (2007) Nonribosomal peptide synthesis in Aspergillus fumigatus and other fungi. Microbiology. 153(5), 1297–1306.

Staley, J.T. (2006) The bacterial species dilemma and the genomic-phylogenetic species concept. Philosophical Transactions of the Royal Society B: Biological Sciences. 361(1475), 1899–1909.

Starcevic, A., Wolf, K., Diminic, J., Zucko, J., Ruzic, I.T., Long, P.F., Hranueli, D., Cullum, J. (2011) Recombinatorial biosynthesis of polyketides. Journal of Industrial Microbiology & Biotechnology. 39(3), 503–511.

Stevens, D.C., Henry, M.R., Murphy, K.A., Boddy, C.N. (2010) Heterologous Expression of the Oxytetracycline Biosynthetic Pathway in Myxococcus xanthus. Applied and Environmental Microbiology. 76(8), 2681–2683.

Stirrett, K., Denoya, C., Westpheling, J. (2008) Branched-chain amino acid catabolism provides precursors for the Type II polyketide antibiotic, actinorhodin, via pathways that are nutrient dependent. Journal of Industrial Microbiology & Biotechnology. 36(1), 129–137.

Strieker, M., TanoviÄ, A., Marahiel, M.A. (2010) Nonribosomal peptide synthetases: structures and dynamics. Current Opinion in Structural Biology. 20(2), 234–240.

Sugawara, K., Dohmae, N., Kasai, K., Saido-Sakanaka, H. (2003) Isolation and Identication of Novel ADP-Ribosylated Proteins from Streptomyces coelicolor A3(2). Bioscience, Biotechnology, and Biochemistry. 66(10), 2292–2296.

Sun, J., Hesketh, A., Bibb, M. (2001) Functional Analysis of relA and rshA, Two relA/spoT Homologues of Streptomyces coelicolor A3(2). Journal of Bacteriology. 183(11), 3488–3498.

Suzuki, M.T., Giovannoni, S.J. (1996) Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Applied and Environmental Microbiology. 62(2), 625–630.

Suzuki, T., Mochizuki, S., Yamamoto, A., Arakawa, K., Kinashi, H. (2010) Regulation of lankamycin biosynthesis in Streptomyces rochei by two SARP genes, srrY and srrZ. Bioscience, Biotechnology, and Biochemistry. 74(4), 819–827.

Swiatek, M.A., Gubbens, J., Bucca, G., Song, E., Yang, Y.H., Laing, E., Kim, B.G., Smith, C.P., van Wezel, G.P. (2013) The ROK Family Regulator Rok7B7 Pleiotropically Affects Xylose Utilization, Carbon Catabolite Repression, and Antibiotic Production in Streptomyces coelicolor. Journal of Bacteriology. 195(6), 1236–1248.

Takano, E. (2006) γ-Butyrolactones: Streptomyces signalling molecules regulating antibiotic production and differentiation. Current Opinion in Microbiology. 9(3), 287–294.

Takano, E., Tao, M., Long, F., Bibb, M.J., Wang, L., Li, W., Buttner, M.J., Bibb, M.J., Deng, Z.X., Chater, K.F. (2003) A rare leucine codon in adpA is implicated in the morphological defect of bldA mutants of Streptomyces coelicolor. Molecular Microbiology. 50(2), 475–486.

Talley, K., Alexov, E. (2010) On the pH-optimum of activity and stability of proteins. Proteins: Structure, Function, and Bioinformatics, 78(12), 2699-706.

Tanaka, A., Takano, Y., Ohnishi, Y., Horinouchi, S. (2007) AfsR Recruits RNA Polymerase to the afsS Promoter: A Model for Transcriptional Activation by SARPs. Journal of Molecular Biology. 369(2), 322–333.

Tang, L., Grimm, A., Zhang, Y.X., Hutchinson, C.R. (1996) Purification and characterization of the DNA-binding protein DnrI, a transcriptional factor of daunorubicin biosynthesis in Streptomyces peucetius. Molecular Microbiology. 22(5), 801–813.

Tang, Z., Xiao, C., Zhuang, Y., Chu, J., Zhang, S., Herron, P.R., Hunter, I.S., Guo, M. (2011) Improved oxytetracycline production in Streptomyces rimosus M4018 by metabolic engineering of the G6PDH gene in the pentose phosphate pathway. Enzyme and Microbial Technology. 49(1), 17–24.

Tettelin, H., Masignani, V., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.L., Angiuoli, S.V., Crabtree, J., Jones, A.L., Durkin, A.S., Deboy, R.T., Davidsen, T.M., Mora, M., Scarselli, M., Margarit y Ros, I., Peterson, J.D., Hauser, C.R., Sundaram, J.P., Nelson, W.C., Madupu, R., Brinkac, L.M.,

Dodson, R.J., Rosovitz, M.J., Sullivan, S.A., Daugherty, S.C., Haft, D.H., Selengut, J., Gwinn, M.L., Zhou, L., Zafar, N., Khouri, H., Radune, D., Dimitrov, G., Watkins, K., O'Connor, K.J.B., Smith, S., Utterback, T.R., White, O., Rubens, C.E., Grandi, G., Madoff, L.C., Kasper, D.L., Telford, J.L., Wessels, M.R., Rappuoli, R., Fraser, C.M. (2005) Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome". Proceedings of the National Academy of Sciences of the United States of America. 102(39), 13950–13955.

Tseng, H.C., Chen, C.W. (1991) A cloned ompR-like gene of Streptomyces lividans 66 suppresses defective melC1, a putative copper-transfer gene. Molecular Microbiology. 5(5), 1187–1196.

Tucker, N.P., Ghosh, T., Bush, M., Zhang, X., Dixon, R. (2010) Essential roles of three enhancer sites in 54-dependent transcription by the nitric oxide sensing regulatory protein NorR. Nucleic Acids Research. 38(4), 1182–1194.

Uguru, G.C., Stephens, K.E., Stead, J.A., Towle, J.E., Baumberg, S., McDowall, K.J. (2005) Transcriptional activation of the pathway-specific regulator of the actinorhodin biosynthetic genes in Streptomyces coelicolor. Molecular Microbiology. 58(1), 131–150.

Umetsu, M. (2003) How Additives Influence the Refolding of Immunoglobulin-folded Proteins in a Stepwise Dialysis System. Spectroscopic Evidence For Highly Efficient Refolding Of A Single-Chain Fv Fragment. Journal of Biological Chemistry. 278(11), 8979–8987.

Valenzano, C.R., You, Y.-O., Garg, A., Keatinge-Clay, A., Khosla, C., Cane, D.E. (2010) Stereospecificity of the Dehydratase Domain of the Erythromycin Polyketide Synthase. Journal of the American Chemical Society. 132(42), 14697–14699.

van Keulen, G., Alderson, J., White, J., Sawers, R.G. (2005) Nitrate respiration in the actinomycete Streptomyces coelicolor. Biochemical Society Transactions. 33(1), 210–212.

van Keulen, G., Jonkers, H.M., Claessen, D., Dijkhuizen, L., Wosten, H.A.B. (2003) Differentiation and Anaerobiosis in Standing Liquid Cultures of Streptomyces coelicolor. Journal of Bacteriology. 185(4), 1455–1458.

van Wezel, G.P., K ouml nig, M., Mahr, K., Nothaft, H., Thomae, A.W., Bibb, M., Titgemeyer, F. (2007) A New Piece of an Old Jigsaw: Glucose Kinase Is Activated Posttranslationally in a Glucose Transport-Dependent Manner in Streptomyces coelicolor A3(2). Journal of Molecular Microbiology and Biotechnology. 12(1-2), 67–74.

van Wezel, G.P., Mahr, K., König, M., Traag, B.A., Pimentel-Schmitt, E.F., Willimek, A., Titgemeyer, F. (2004) GlcP constitutes the major glucose uptake system of Streptomyces coelicolor A3(2). Molecular Microbiology.

55(2), 624–636.

van Wezel, G.P., McDowall, K.J. (2011) The regulation of the secondary metabolism of Streptomyces: new links and experimental advances. Natural Product Reports. 28(7), 1311.

Vasanthakumar, A., Kattusamy, K., Prasad, R. (2013) Regulation of daunorubicin biosynthesis in Streptomyces peucetius -feed forward and feedback transcriptional control. Journal of Basic Microbiology. 53(8), 636–644.

Vesligaj, M., Filipović, M., Pigac, J., Hranueli, D. (1981) Isolation of Streptomyces rimosus Mutants with Reduced Actinophage Susceptibility. Applied and Environmental Microbiology. 41(4), 986–991.

Vior, N.M., Olano, C., Garcia, I., Mendez, C., Salas, J.A. (2014) Collismycin A biosynthesis in Streptomyces sp. CS40 is regulated by iron levels through two pathway-specific regulators. Microbiology. 160(3), 467–478.

Vlahovicek, K., Pongor, S. (2000) Model.it: building three dimensional DNA models from sequence data. Bioinformatics. 16(11), 1044–1045.

Volpon, L., Lancelin, J.-M. (2002) Solution NMR structure of five representative glycosylated polyene macrolide antibiotics with a sterol-dependent antifungal activity. European Journal of Biochemistry. 269(18), 4533–4541.

Walker, J.E., Saraste, M., Runswick, M.J., Gay, N.J. (1982) Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. The EMBO Journal. 1(8), 945–951.

Wang, J., Zhao, G P. (2009) Biochemical and Biophysical Research Communications. Biochemical and Biophysical Research Communications. 386(1), 77–81.

Wang, L., Tian, X., Wang, J., Yang, H., Fan, K., Xu, G., Yang, K., Tan, H. (2009) Autoregulation of antibiotic biosynthesis by binding of the end product to an atypical response regulator. Proceedings of the National Academy of Sciences. 106(21), 8617–8622.

Wang, P., Gao, X., Chooi, Y.H., Deng, Z., Tang, Y. (2011) Genetic characterization of enzymes involved in the priming steps of oxytetracycline biosynthesis in Streptomyces rimosus. Microbiology. 157(8), 2401–2409.

Wang, P., Kim, W., Pickens, L.B., Gao, X., Tang, Y. (2012) Heterologous Expression and Manipulation of Three Tetracycline Biosynthetic Pathways. Angewandte Chemie International Edition. 51(44), 11136–11140.

Wang, P., Zhang, W., Zhan, J., Tang, Y. (2009) Identification of OxyE as an

Ancillary Oxygenase during Tetracycline Biosynthesis. ChemBioChem. 10(9), 1544–1550.

Wang, R., Mast, Y., Wang, J., Zhang, W., Zhao, G., Wohlleben, W., Lu, Y., Jiang, W. (2012) Identification of two-component system AfsQ1/Q2 regulon and its cross-regulation with GlnR in Streptomyces coelicolor. Molecular Microbiology. 87(1), 30–48.

Wang, S., Qiao, X., Liu, X., Zhang, X., Wang, C., Zhao, X., Chen, Z., Wen, Y., Song, Y. (2010) Complete genomic sequence analysis of the temperate bacteriophage phiSASD1 of Streptomyces avermitilis. Virology. 403(1), 78–84.

Wang, W., Shu, D., Chen, L., Jiang, W., Lu, Y. (2009) Cross-talk between an orphan response regulator and a noncognate histidine kinase in Streptomyces coelicolor. FEMS Microbiology Letters. 294(2), 150–156.

Watanabe, K., Praseuth, A.P., Wang, C.C. (2007) A comprehensive and engaging overview of the type III family of polyketide synthases. Current Opinion in Chemical Biology. 11(3), 279–286.

Watve, M., Tickoo, R., Jog, M., Bhole, B. (2001) How many antibiotics are produced by the genus Streptomyces? Archives of Microbiology. 176(5), 386–390.

Webb, K., Bristow, T., Sargent, M., Stein, B. (2004) Methodology for accurate mass measurement of small molecules. Best practice guide. LGC Limited, Tddington, 1–8.

Wei, J., Tian, Y., Niu, G., Tan, H. (2014) GouR, a TetR Family Transcriptional Regulator, Coordinates the Biosynthesis and Export of Gougerotin in Streptomyces graminearus. Applied and Environmental Microbiology. 80(2), 714–722.

Wei, W., Wang, W., Cao, Z., Yu, H., Wang, X., Zhao, J., Tan, H., Xu, H., Jiang, W., LI, Y. (2007) Comparative Analysis of Two-component Signal Transduction System in Two Streptomycete Genomes. Acta Biochimica et Biophysica Sinica. 39(5), 317–325.

Wei, Z.-H., Bai, L., Deng, Z., Zhong, J.-J. (2012) Impact of nitrogen concentration on validamycin A production and related gene transcription in fermentation of Streptomyces hygroscopicus 5008. Bioprocess and Biosystems Engineering. 35(7), 1201–1208.

White, J., Bibb, M. (1997) bldA dependence of undecylprodigiosin production in Streptomyces coelicolor A3(2) involves a pathway-specific regulatory cascade. Journal of Bacteriology. 179(3), 627–633.

Wietzorrek, A., Bibb, M. (1997) A novel family of proteins that regulates antibiotic production in streptomycetes appears to contain an OmpR-like

DNA-binding fold. Molecular Microbiology. 25(6), 1181–1184.

Wildermuth, H. (1970) Development and organization of the aerial mycelium in Streptomyces coelicolor. Journal of General Microbiology. 60(1), 43–50.

Willenbrock, H., Hallin, P.F., Wassenaar, T.M., Ussery, D.W. (2007) Characterization of probiotic Escherichia coli isolates with a novel pan-genome microarray. Genome Biology. 8(12), R267.

Williams, G.J. (2013) Engineering polyketide synthases and nonribosomal peptide synthetases. Current Opinion in Structural Biology. 23(4), 603–612.

Wilmes, P., Bowen, B.P., Thomas, B.C., Mueller, R.S., Denef, V.J., VerBerkmoes, N.C., Hettich, R.L., Northen, T.R., Banfield, J.F. (2010) Metabolome-proteome differentiation coupled to microbial divergence. mBio. 1(5), pii, e00246-10.

Wilson, D.J., Patton, S., Florova, G., Hale, V., Reynolds, K.A. (1998) The shikimic acid pathway and polyketide biosynthesis. Journal of Industrial Microbiology & Biotechnology. 20(5), 299–303.

Wilson, D.J., Xue, Y., Reynolds, K.A., Sherman, D.H. (2001) Characterization and Analysis of the PikD Regulatory Factor in the Pikromycin Biosynthetic Pathway of Streptomyces venezuelae. Journal of Bacteriology. 183(11), 3468–3475.

Xu, D., Seghezzi, N., Esnault, C., Virolle, M.J. (2010) Repression of Antibiotic Production and Sporulation in Streptomyces coelicolor by Overexpression of a TetR Family Transcriptional Regulator. Applied and Environmental Microbiology. 76(23), 7741–7753.

Xu, Y., Zhou, T., Zhou, Z., Su, S., Roberts, S.A., Montfort, W.R., Zeng, J., Chen, M., Zhang, W., Lin, M., Zhan, J., Molnár, I. (2013) Rational reprogramming of fungal polyketide first-ring cyclization. Proceedings of the National Academy of Sciences. 110(14), 5398–5403.

Yadav, G., Gokhale, R.S., Mohanty, D. (2003) Computational Approach for Prediction of Domain Organization and Substrate Specificity of Modular Polyketide Synthases. Journal of Molecular Biology. 328(2), 335–363.

Yagüe, P., López-García, M.T., Rioseras, B., Sanchez, J., Manteca, A. (2013) Pre-sporulation stages of Streptomycesdifferentiation: state-of-the-art and future perspectives. FEMS Microbiology Letters. 342(2), 79–88.

Yamamoto, S., Mitobe, J., Ishikawa, T., Wai, S.N., Ohnishi, M., Watanabe, H., Izumiya, H. (2013) Regulation of natural competence by the orphan two-component system sensor kinase ChiS involves a non-canonical transmembrane regulator in Vibrio cholerae. Molecular Microbiology. 91(2), 326–347.

Yang, Y.-H., Song, E., Kim, E.-J., Lee, K., Kim, W.S., Park, S.-S., Hahn, J.-S., Kim, B.-G. (2008) NdgR, an IclR-like regulator involved in amino-acid-dependent growth, quorum sensing, and antibiotic production in Streptomyces coelicolor. Applied Microbiology and Biotechnology. 82(3), 501–511.

Yepes, A., Rico, S., Rodríguez-García, A., Santamaría, R.I., Díaz, M. (2011) Novel Two-Component Systems Implied in Antibiotic Production in Streptomyces coelicolor. PLoS ONE. 6(5), e19980.

Yin, J., Straight, P.D., Hrvatin, S., Dorrestein, P.C., Bumpus, S.B., Jao, C., Kelleher, N.L., Kolter, R., Walsh, C.T. (2007) Genome-Wide High-Throughput Mining of Natural-Product Biosynthetic Gene Clusters by Phage Display. Chemistry & Biology. 14(3), 303–312.

Yin, X., Xu, X., Wu, H., Yuan, L., Huang, X., Zhang, B. (2013) SACE_0012, a TetR-Family Transcriptional Regulator, Affects the Morphogenesis of Saccharopolyspora erythraea. Current Microbiology. 67(6), 647–651.

Ylihonko, K., Tuikkanen, J., Jussila, S., Cong, L., Mäntsälä, P. (1996) A gene cluster involved in nogalamycin biosynthesis from Streptomyces nogalater: sequence analysis and complementation of early-block mutations in the anthracycline pathway. Molecular & General Genetics: MGG. 251(2), 113–120.

Yu, L., Cao, N., Wang, L., Xiao, C., Guo, M., Chu, J., Zhuang, Y., Zhang, S. (2012) Oxytetracycline biosynthesis improvement in Streptomyces rimosus following duplication of minimal PKS genes. Enzyme and Microbial Technology. 50(6-7), 318–324.

Yu, L., Yan, X., Wang, L., Chu, J., Zhuang, Y., Zhang, S., Guo, M. (2012) Molecular cloning and functional characterization of an ATP-binding cassette transporter OtrC from Streptomyces rimosus. BMC Biotechnology. 12(1), 52.

Yu, Z., Zhu, H., Dang, F., Zhang, W., Qin, Z., Yang, S., Tan, H., Lu, Y., Jiang, W. (2012) Differential regulation of antibiotic biosynthesis by DraR-K, a novel two-component system in Streptomyces coelicolor. Molecular Microbiology. 85(3), 535–556.

Zaburannyi, N., Rabyk, M., Ostash, B., Fedorenko, V., Luzhetskyy, A. (2014) Insights into naturally minimised Streptomyces albus J1074 genome. BMC Genomics. 15(1), 1–11.

Zha, W., Rubin-Pitel, S.B., Shao, Z., Zhao, H. (2009) Improving cellular malonyl-CoA level in Escherichia coli via metabolic engineering. Metabolic Engineering. 11(3), 192–198.

Zhang, J.H., Xiao, G., Gunsalus, R.P., Hubbell, W.L. (2003) Phosphorylation Triggers Domain Separation in the DNA Binding Response Regulator NarL.

Biochemistry. 42(9), 2552–2559.

Zhang, R.-G., Pappas, K.M., Pappas, T., Brace, J.L., Miller, P.C., Oulmassov, T., Molyneaux, J.M., Anderson, J.C., Bashkin, J.K., Winans, S.C., Joachimiak, A. (2002) Structure of a bacterial quorum-sensing transcription factor complexed with pheromone and DNA. Nature. 417(6892), 971–974.

Zhang, W., Ames, B.D., Tsai, S.C., Tang, Y. (2006) Engineered biosynthesis of a novel amidated polyketide, using the malonamyl-specific initiation module from the oxytetracycline polyketide synthase. Applied and Environmental Microbiology. 72(4), 2573–2580.

Zhang, W., Watanabe, K., Cai, X., Jung, M.E., Tang, Y., Zhan, J. (2008) Identifying the Minimal Enzymes Required for Anhydrotetracycline Biosynthesis. Journal of the American Chemical Society. 130(19), 6068–6069.

Zhao, W., Zhong, Y., Yuan, H., Wang, J., Zheng, H., Wang, Y., Cen, X., Xu, F., Bai, J., Han, X., Lu, G., Zhu, Y., Shao, Z., Yan, H., Li, C., Peng, N., Zhang, Z., Zhang, Y., Lin, W., Fan, Y., Qin, Z., Hu, Y., Zhu, B., Wang, S., Ding, X., Zhao, G.-P. (2010) Complete genome sequence of the rifamycin SV-producing Amycolatopsis mediterranei U32 revealed its genetic characteristics in phylogeny and metabolism. Nature Publishing Group. 20(10), 1096–1108.

Zheng, H., Wisedchaisri, G., Gonen, T. (2013) Crystal structure of a nitrate/nitrite exchanger. Nature. 497(7451), 547–651.

Zheng, J., Gay, D.C., Demeler, B., White, M.A., Keatinge-Clay, A.T. (2012) Divergence of multimodular polyketide synthases revealed by a didomain structure. Nature Chemical Biology. 8(7), 615–621.

Zhou, H., Gao, Z., Qiao, K., Wang, J., Vederas, J.C., Tang, Y. (2012) A fungal ketoreductase domain that displays substrate-dependent stereospecificity. Nature Chemical Biology. 8(4), 331–333.

Zhou, Y., Liang, Y., Lynch, K.H., Dennis, J.J., Wishart, D.S. (2011) PHAST: A Fast Phage Search Tool. Nucleic Acids Research, 1–6.

Zhou, Z., Gu, J., Li, Y.Q., Wang, Y. (2012) Genome plasticity and systems evolution in Streptomyces. BMC Bioinformatics. 13(Suppl 10), S8.

11 Appendix – Publications

genomeA
Journals.ASM.org

Draft Genome Sequence of the Oxytetracycline-Producing Bacterium *Streptomyces rimosus* ATCC 10970

Florence E. Pethick,[a] Alison C. MacFadyen,[a] Zhenyu Tang,[a,b] Vartul Sangal,[a] Tze-Tze Liu,[c] Ju Chu,[b] Gregor Kosec,[d] Hrvoje Petkovic,[d] Meijin Guo,[b] Ralph Kirby,[e] Paul A. Hoskisson,[a] Paul R. Herron,[a] Iain S. Hunter[a]

Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, Glasgow, United Kingdom[a]; State Key Laboratory of Bioreactor Engineering, East China University of Science and Technology, Shanghai, China[b]; Genome Research Center, National Yang-Ming University, Taipei, Taiwan[c]; Acies Bio, Ltd., Ljubljana, Slovenia[d]; Department of Life Sciences, Institute of Genome Science, National Yang-Ming University, Taipei, Taiwan[e]

We report the draft genome of *Streptomyces rimosus* (ATCC 10970), a soil isolate that produces oxytetracycline, a commercially important and clinically useful antibiotic.

Received 28 January 2013 **Accepted** 6 February 2013 **Published** 7 March 2013

Citation Pethick FE, MacFadyen AC, Tang Z, Sangal V, Liu T-T, Chu J, Kosec G, Petkovic H, Guo M, Kirby R, Hoskisson PA, Herron PR, Hunter IS. 2013. Draft genome sequence of the oxytetracycline-producing bacterium *Streptomyces rimosus* ATCC 10970. Genome Announc. 1(2):e00063-13. doi:10.1128/genomeA.00063-13.

Copyright © 2013 Pethick et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported license.

Address correspondence to Iain S. Hunter, i.s.hunter@strath.ac.uk.

Streptomyces rimosus is a Gram-positive, aerobic, filamentous actinobacterium. It was reported in 1950 and was patented (1) as the founder strain for the production of the antibiotic oxytetracycline (OTC) (2). Recombination was recognized (3) and a rudimentary genetic map was deduced (4). The pharmaceutical industry set out to improve the commercial productivity of OTC, deriving higher-titer genealogies that led to the contemporary industrial strains that are now responsible for the production of $>10^8$ kg of OTC annually. Stimulated by commercial interest over the years, a sophisticated genetics and molecular biology has been developed for *S. rimosus* (5). This announcement reports the derivation of data to enable the generation of a scaffold genome sequence for this important industrial species.

Genome sequencing of the *S. rimosus* type strain ATCC 10970 was carried out using a whole-genome shotgun sequencing approach performed on a Roche 454 GS Junior apparatus. Using three single-ended runs, we obtained 440,253 reads. The reads were assembled using Genome Sequencer *de novo* Assembler (version 2.7, Roche), which led to a final assembly of 453 contigs of >500 bp. The total size of the assembly was 9.5 Mbp, with a mean contig size of 21 kbp (average, 17× coverage) and a G+C content of 71.88%. Automatic functional annotation results were obtained using the NCBI Prokaryotic Genome Annotation Pipeline (http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html).

The draft genome of *S. rimosus* is estimated to have a total of 8,416 protein-coding genes, along with 66 tRNAs. The overall genome size is consistent with the physical map of *S. rimosus*, derived by pulsed-field gel electrophoresis (6). It is also within the range of other published streptomycete genome sequences. Three gene clusters associated with the production of previously reported secondary metabolites [OTC (7), rimocidin (8), and desferrioxamine C (9)] are readily identified within the sequence, in addition to around 45 other putative secondary metabolite biosynthetic clusters whose functions remain to be elucidated (10). The strain is lysogenic for phages RP2 and RP3, with the DNA sequences corresponding to the restriction maps published previously (11). The genome sequence is also likely to include at least one giant (387 kb) linear plasmid known to be present in *S. rimosus* (12).

Sequencing of *S. rimosus* ATCC 10970 provides a scaffold from which to investigate and understand how mutation results in higher titers of OTC.

Nucleotide sequence accession numbers. The *S. rimosus* ATCC 10970 Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. ANSJ00000000. The version described in this paper is the first version, ANSJ01000000.

ACKNOWLEDGMENTS

We acknowledge financial support from the National Yang-Ming University, Taiwan, for the sequencing project (R.K.), a British Council for Research Projects Network grant (P.A.H. and H.P.), and the British Council/CUC for a Scotland-China Higher Education Research Partnership for PhD Studies (Z.T.).

REFERENCES

1. **Sobin BA, Finlay AC.** 1950. Terramycin and its production. US patent 2,516,080.
2. **Finlay AC, Hobby GL, P'an SY, Regna PP, Routien JB, Seeley DB, Shull GM, Sobin BA, Solomons IA, Vinson JW, Kane JH.** 1950. Terramycin, a new antibiotic. Science **111**:85.
3. **Alikhanian SI, Mindlin SZ.** 1957. Recombinations in *Streptomyces rimosus*. Nature **180**:1208–1209.
4. **Friend EJ, Hopwood DA.** 1971. The linkage map of *Streptomyces rimosus*. J. Gen. Microbiol. **68**:187–197.
5. **Petković H, Cullum J, Hranueli D, Hunter IS, Perić-Concha N, Pigac J, Thamchaipenet A, Vujaklija D, Long PF.** 2006. Genetics of *Streptomyces rimosus*, the oxytetracycline producer. Microbiol. Mol. Biol. Rev. **70**: 704–728.
6. **Pandza K, Pfalzer G, Cullum J, Hranueli D.** 1997. Physical mapping shows that the unstable oxytetracycline gene cluster of *Streptomyces rimosus* lies close to one end of the linear chromosome. Microbiology **143**: 1493–1501.
7. **Zhang W, Ames BD, Tsai SC, Tang Y.** 2006. Engineered biosynthesis of a novel amidated polyketide, using the malonamyl-specific initiation module from the oxytetracycline polyketide synthase. Appl. Environ. Microbiol. **72**:2573–2580.
8. **Seco EM, Perez-Zuniga FJ, Rolon MS, Malpartida F.** 2004. Starter unit

choice determines the production of two tetraene macrolides, rimocidin and CE-108, in *Streptomyces diastaticus* var. Chem. Biol. **11**:357–366.
9. **Barona-Gómez F, Wong U, Giannakopulos AE, Derrick PJ, Challis GL.** 2004. Identification of a cluster of genes that directs desferrioxamine biosynthesis in *Streptomyces coelicolor* M145. J. Am. Chem. Soc. **126**: 16282–16283.
10. **Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R.** 2011. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters. Nucleic Acids Res. **29**:W339–W346.
11. **Rausch H, Vesligaj M, Pocta D, Biuković G, Pigac J, Cullum J, Schmieger H, Hranueli D.** 1993. The temperate phages RP2 and RP3 of *Streptomyces rimosus*. J. Gen. Microbiol. **139**:2517–2524.
12. **Gravius B, Glocker D, Pigac J, Pandza K, Hranueli D, Cullum J.** 1994. The 387 kb linear plasmid pPZG101 of *Streptomyces rimosus* and its interactions with the chromosome. Microbiology **140**:2271–2277.