Multi-faceted synthetic biology approaches towards aminoglycoside production in *Streptomyces* spp.

A thesis submitted to the University of Manchester

for the degree of Doctor of Philosophy

in the Faculty of Science and Engineering

2022

Katherine V. Baker

Department of Chemistry

School of Natural Sciences

Table of Contents

Word Count: 63499 words

List of Figures

1.

2.

3.

4.

5.

List of Tables

1.

2.

3.

4.

List of Abbreviations

2-DOI	2-deoxy-scyllo-inososе
2-DOIA	2-deoxy-scyllo-inosamine
2-DOS	2-deoxystreptamine
a.u.	Arbitrary units
Act	Actinorhodin
AME	Aminoglycoside-modifying enzymes
antiSMASH	antibiotics & Secondary Metabolite Analysis Shell
Apra, Apra^R	Apramycin, apramycin resistance
ARTS	Antibiotic Resistant Target Seeker
ATc	Anhydrotetra-cycline
ATP	Adenosine triphosphate
B. subtilis	*Bacillus subtilis*
BAGEL	Bacteriocin Genome mining tool
BGC	Biosynthetic Gene Cluster
BLAST	Basic local alignment search tool
BLASTp	Protein BLAST
bp	Base pairs
C24	Cluster 24 (from *M.* sp. DEM32671)
CDC	Centers for Disease Control and Prevention
CFU	Colony Forming Units
CRISPR	Clustered regularly interspaced short palindromic repeats
CRISPR-BEST	CRISPR-Base Editing SysTEM
dDDH	digital DNA-DNA hybridisation
ddH_2O	double-distilled water
DEM	*Micromonospora* sp. DEM32671
DMATS	dimethylallyltrypto-phan synthases
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DSB	Double-strand break
DSM	Deutsche Sammlung von Mikroorganismen (German Collection of Microorganisms)
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen (German Collection of Microorganisms and Cell Cultures)
E. coli	*Escherichia coli*
EDTA	Ethylenediamine-tetraacetic acid
Gen, Gen^R	Gentamicin, gentamicin resistance
GYM	Glucose yeast extract malt extract medium
FDA	US Food and Drug Administration
HDR	Homology-directed repair

HILIC	Hydrophilic interaction chromatography
HiTES	High-throughput elicitor screening
HPLC	High performance liquid chromatography
Hyg, Hyg^R	Hygromycin, hygromycin resistance
iTOL	Iterative Tree of Life
Kan, Kan^R	Kanamycin, kanamycin resistance
kb	Kilobase
kDa	Kilodalton
L. rhamnosus	*Lactobacillus rhamnosus*
LB	Lysogeny broth
LC-MS	Liquid chromatography-mass spectrometry
LLHR	Linear-plus-linear homologous recombination
M. abscessus	*Mycobacterium abscessus*
M. echinospora	*Micromonospora echinospora*
M. inyonensis	*Micromonospora inyonensis*
M. olivasterospora	*Micromonospora olivasterospora*
M. pallida	*Micromonospora pallida*
M. sagamiensis	*Micromonospora sagamiensis*
***M.* sp. DEM32671**	*Micromonospora* sp. DEM32671
m/z	Mass-to-charge ratio
Mb	Megabase
MDR	Multidrug-resistant
MEGA	Molecular Evolutionary Genetics Analysis
MET	Mechanotransducer
MIBiG	Minimum Information about a Biosynthetic Gene Cluster
MIC	Minimum Inhibitory Concentration
MRSA	Methicillin-resistant *Staphylococcus aureus*
MS	Mass spectrometry
N/A	Not applicable
NHEJ	Non-homologous end-joining
NRPS	Non-ribosomal peptide synthetase
OAT	Orthologous Average Nucleotide Identity Tool
OD_{450}	Optical density at λ = 450 nm
OD_{600}	Optical density at λ = 600 nm
OOPS	Output ordering and prioritization system
ORF	Open reading frame
OrthoANI	Orthologous Average Nucleotide Identity
PAC	Phage Artificial Chromosome
PAGE	Polyacrylamide gel electrophoresis
PAM	Protospacer adjacent motif

PCR	Polymerase Chain Reaction
PDB	Protein data bank
PKS	Polyketide synthase
PLP enzymes	Pyridoxal-5′-phosphate-dependent enzymes
preQ1	Pre-queuosine 1
preQ1-I	Class I preQ1 (riboswitch)
preQ1-II	Class II preQ1 (riboswitch)
PRISM	Prediction Informatics for Secondary Metabolomes
R2YE	R2 medium with yeast extract
RBS	Ribosome Binding Site
ROS	Reactive oxygen species
Red	Undecylprodigiosins
RiPP	Ribosomally synthesized and post-translationally modified peptides
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
RT-qPCR	Quantitative reverse transcription PCR
SBOL	Synthetic biology open language
S. coelicolor	*Streptomyces coelicolor*
S. lividans	*Streptomyces lividans*
SCO	*Streptomyces coelicolor*
SDS	Sodium dodecyl sulfate
SELEX	Systematic Evolution of Ligands by Exponential Enrichment
SFM	Soya flour mannitol
sgRNA	Single-stranded guide RNA
SPE	Solid phase extraction
TAE	Tris-acetate-EDTA
TAR	Transformation-associated recombination
TetR	Tetracycline repressor
TetRiS	TetR, adapted for expression in *Streptomyces*
Thio, Thio^{R}	Thiostrepton, thiostrepton resistance
Tm	Melting temperature
tRNA	Transfer RNA
TSB	Tryptic soya broth
TYGS	Type (strain) genome server
UDP-GlcNAc	Uridine diphosphate N-acetylglucosamine
UDP-Xylose	Uridine Diphosphate Xylose
UGI	Uracil glycosylase inhibitor
UNS	Unique nucleotide sequences
UV	Ultraviolet
w/v	weight by volume
YEME	Yeast extract malt extract medium

Abstract

The threat of antimicrobial resistance is regarded as a major health concern. While new therapies are in development, the rejuvenation of existing antibiotics will provide further options for treating antibiotic resistant infections. The aminoglycosides are under-utilized as antibiotics, due to ototoxic and nephrotoxic side effects. Reducing these side effects could revitalise the group and yield attractive therapeutics for treatment of resistant infections.

This PhD project focused on the heterologous expression of a predicted gentamicin biosynthetic gene cluster (Cluster 24) from a novel *Micromonospora* sp., identified by the industrial partner Demuris. After determining that the ideal *Streptomyces coelicolor* host strain was sensitive to gentamicin, we generated a series of 'superhost' strains by increasing gentamicin resistance. While we were unable to detect gentamicin production in the first trials, we identified an ion of m/z 502.2246 with proposed chemical formula $C_{18}H_{35}N_3O_{13}$, which could represent an aminoglycoside not previously linked to gentamicin biosynthesis. We further engineered Cluster 24 in the non-resistant host strain using targeted knock-outs to produce gentamicin C1a; however, the yield was too low for characterisation of the output. A cluster refactoring approach in the non-resistant host strains also encountered difficulties as it appeared the rebuilt cluster was lost over time, possibly due to product toxicity. Using the resistant strains, a constitutive promoter cassette was knocked into Cluster 24. A yield increase was visualised via antimicrobial bioassay, and we again identified the compound $C_{18}H_{35}N_3O_{13}$ which was found in previous bioactive samples.

Lastly, we identified that an incomplete pathway towards the modified nucleoside queuosine co-clustered within certain aminoglycoside clusters. As preQ1 (the final possible product) is the cognate ligand for three riboswitch classes, it was posited that cluster expression could be under riboswitch control. A putative riboswitch was identified *in silico* and tested alongside two preQ1 riboswitches from *Mycobacterium*

abscessus subsp. *abscessus* and *Lactobacillus rhamnosus* in *S. coelicolor*. While the Cluster 24 region did not respond to preQ1, the riboswitch from *M. abscessus* subsp. *abscessus* repressed fluorescence upon preQ1 induction. Deletion of three predicted queuosine biosynthesis genes from Cluster 24 appeared to increase bioactivity; however, the role of these genes in aminoglycoside biosynthesis remains unclear. The work presented in this thesis targets many aspects of heterologous production, providing a platform towards aminoglycoside production in *S. coelicolor*.

Declaration

A portion of the work detailed in Chapter 3, section 3.3.4., was carried out by an undergraduate student under the supervision of Katherine V. Baker (Amelia Kelly) in support for her degree in BSc (Hons) Microbiology at the University of Manchester. Figure S3.8, and some elements from Figure S3.9 were reformatted from Amelia Kelly's undergraduate project report "Improvement of Cluster 44 gentamicin A2 yields through CRISPR-Cas9 mediated promoter knock in" (Figure S3.9A in the thesis corresponding to Figure 5 in the report with reformatting, Figure S3.9B in the thesis corresponding to Figure 7B in the report with reformatting). No other portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

Acknowledgements

I would first like to thank my supervisory team, Prof. Eriko Takano and Prof. Rainer Breitling, who gave me the opportunity to complete this PhD work. Their advice and support, especially during the tough times we have faced in the past two years, has been invaluable. I would also like to thank Demuris and especially Dr Nick Allenby for additional advice on project direction throughout.

I would like to thank all the past and present Takano group members who have given their support during this time. In particular I would like to thank the following in no order: Suhui Ye Huang and Marc Biarnes Carrera, who summoned a vast amount of patience to answer endless questions during my first 1.5 years in the laboratory, and whose mentorship and sense of humour gave me the foundation to continue. Erik Hanko, for being a near-constant source of encouragement during my writing up period, for ensuring I kept focus even during the final weeks and for stealing my biscuits so that I didn't over-eat them. Oksana Bilyk, who has been of great help when planning *Streptomyces* experiments and who has always been there to listen when I am frustrated (as happens often, with PhD study). Kris Niño Valdehuesa, for always bringing a smile to my face with his singing in the lab. Dominic Whittall for the music recommendations, the memes, and for broadening my vocabulary substantially. Kamila Schmidt for – amongst everything else she does to keep our lab group organised – dealing with all my purchase orders, particularly in the final six months when it seemed like everything piled up at once. Gajendar Komati Reddy, who was a great encouragement in my first years in the lab. Jack Connolly, who managed patiently to fine-tune my (sometimes crazy) experimental plans later in the PhD and was always available to talk through an idea or five. Timothy Kirkwood, who made the long days in the laboratory pass by quickly with his numerous ideas for multiple Nobel (and Timbel) Prizes, and who would be very upset if he were not mentioned here. To all of you – your encouragement throughout has grounded me, and I could not have asked for a better, more supportive environment to study in. From the rest of the MIB, I also wanted to thank Francesco del Carratore and

Katherine Hollywood who have both been of great help with analytical chemistry and (especially in the case of Francesco) answering random questions whenever they come up with great patience.

I would not have been able to travel this path without the support of my parents, who have tried to understand every step of this PhD even when the long hours in the lab haven't made the most sense, and who have also rebuilt my confidence when it is at its lowest. Similarly, without my sister Jasmine supplying snacks and stealing my computer to enforce breaks, I would likely not have been able to complete this thesis.

Lastly, I would like to dedicate this work to my grandmother Brenda, who was never able to study at university but whose interest in botany and biology was an inspiration growing up. If she were still with us, I hope that she would be proud that I was able to complete this PhD, and be proud that a scientific education for women is now more possible than ever.

Thesis Format

This PhD thesis has been formatted as a 'journal format' thesis according to the University of Manchester presentation of theses policy. Each results chapter has been prepared as if it were a self-contained manuscript for publication; therefore a separate methods chapter has not been included. Each results chapter contains an abstract, introduction, results and discussion, materials and methods, references and any supplementary information required. The introductory chapter (Chapter 1) has been edited from a prior publication, and the edits made detailed prior to the main text of the chapter.

As the work within this thesis could not have been achieved without collaboration, the author contributions for each results chapter is listed immediately prior to the chapter title. As supervisors of the work, Eriko Takano, Rainer Breitling and Nicholas Allenby provided critical feedback throughout.

1. Introduction

Parts of this chapter were originally published in 2018 as part of a special Advanced Biosystems issue on synthetic biology.

> **Baker, K. V.**, Takano, E. & Breitling, R. The "Three Cs" of Novel Antibiotic Discovery and Production through Synthetic Biology: Biosynthetic Gene Clusters, Heterologous Chassis, and Synthetic Microbial Consortia. *Adv. Biosys.* **1800064**, (2018) https://doi.org/10.1002/adbi.201800064

For contribution to this thesis, the review has been rewritten in places to alter the following: **(1)** replacement of all figures throughout to ensure consistency with the rest of the thesis **(2)** removal of section '3.3. Non-Streptomyces Chassis' and '4. Consortia' to better fit the research aims of the thesis **(3)** removal of paragraphs from the original text 'Microbial antibiotics, like most bacterial natural products… useful as novel antibiotics', "Discovery of new BGCs …. often found nearby.", "Other genome mining tools …. NRPSs, and terpenoids.", "This method was mediated …. upstream of cluster genes.", "In some cases, BGCs do not… more success in increasing product titer.", "However, an approach on the rise …. of antibiotics is important." **(4)** removal of lines from the original text 'Utilizing antibiotics to treat disease …. reduced death counts worldwide.'' **(5)** removal of SMURF, CASSIS and SMIPS from Table 1. **(6)** addition of sections 1.2.4 'CRISPR-Cas-mediated strain engineering to improve heterologous production' (using "In the work of Zhang et al. … has wider-ranging capabilities" from original text) 1.3 'Aminoglycoside antibiotics', 1.4 'Riboswitches as tools for synthetic biology', 1.5 'Project Aims' **(7)** addition of text to section 1.1 to further establish the history of antibiotic discovery "A vast proportion of the antimicrobials … rarely produced under standard laboratory conditions." and "The CDC estimated in 2017 … causing 33,000 deaths yearly)." and "In recent years, priorities have shifted … now

spread worldwide." and addition of text to 1.2.1. "DeepBGC uses a novel deep-learning approach …. rule-based tools such as antiSMASH", "PRISM 4 was recently released …. secondary metabolite classes" "BAGEL was updated … RiPP domain prediction" and addition of text to 1.2.2. "However, their pipeline … true test of the approach." "One well known example … non-model organisms." and addition of text to 1.2.3. "A recent work by Enghiad … classical cosmid cloning techniques.", "In some cases … nine cryptic metabolites", "Despite advances in recent years … able to be identified) **(8)** Movement of paragraph "Traditional cosmid construction … gene cluster at present." in section 1.2.3. to later in the section to ensure clarity. **(9)** updating statistics in 1.1. regarding the current antimicrobials in development and 1.2.1 regarding the current cost of DNA sequencing and the current number of microbial genome sequences available. **(10)** minor changes to keep consistency throughout this thesis.

1.1. Antibiotics, their discovery, and the problem of resistance

An antibiotic can be defined as any small molecule which inhibits bacterial growth or causes bacterial cell death. This can occur through multiple modes of action – for example, the inhibition of cell wall, protein, or nucleic acid biosynthesis, the modification of cell membranes, or the blocking of key bacterial metabolic processes.[1–3] After antibiotics were commercialised by the mid-1900s, the mortality rate of infectious diseases sharply decreased in England to below 1%, in comparison to the earlier 25%.[4]

Antibiotic discovery was at one time incredibly prolific – following the famous discovery and characterization of penicillin (a β-lactam antibiotic from the fungus *Penicillium rubens*),[5] and establishment of a systematic screening method by Selman Waksman and colleagues in the 1940s,[6] a golden age of antibiotic discovery began. A vast proportion of the antimicrobials discovered between the 1940s and the 1960s came from the soil microbiome, largely from soil-derived actinomycetes. Waksman's platform consisted of screening these actinomycetes via antibiotic bioassay and measurement of growth inhibition via an agar overlay.[7] The discovery by Schatz and Waksman of streptomycin via this method led to Waksman being awarded the Nobel prize in 1952 and fuelled the pharmaceutical industry's screening of many antibiotics that remain therapeutically relevant to this day.[8]

After the 1960s, it became clear that the Waksman platform was becoming redundant, with many antibiotics being rediscovered through this method. Synthetic and semi-synthetic antibiotics bridged the gap for a time, the latter being modified from scaffolds identified using the Waksman platform. The first semi-synthetic antibiotic was the modified streptomycin dihydrostreptomycin in 1946, which was more stable than streptomycin and had similar antimicrobial properties. After further trials, this new version of streptomycin unfortunately

showed even stronger ototoxic effects than its predecessor.[9,10] The semi-synthesis of five β-lactam antibiotics of microbial origin has been much more successful, and this class now contributes over 60% of antibiotics currently used therapeutically.[11] Synthetic antibiotics have been available since 1949 with the use of chloramphenicol clinically – though the simultaneous discoveries of this antibiotic were derived from a *Streptomyces* sp. in soil sample from Venezuela and a *Streptomyces* sp. in soil from the Illinois Agricultural Experiment Station, the molecule was made through chemical synthesis in 1949 and was soon used widely.[12–14] The synthetic carbapenem subclass of the β-lactam class of antibiotics have been shown to have good activity against multidrug-resistant infections, as they have a higher resistance to the action of β-lactamases than naturally-derived penicillin and cephalosporin.[15]

The advent of DNA sequencing and DNA cloning led to the identification of clusters of specialised metabolite biosynthesis genes in microbial genomes, i.e. biosynthetic gene clusters (BGCs; discussed further in section **1.2.1**). Four of the specialised metabolites produced by *Streptomyces coelicolor* A3(2), a model actinomycete, were known prior to its genome being sequenced.[16–19] However, upon sequencing of the genome of *S. coelicolor* A3(2) in 2002, 15 biosynthetic gene clusters for other specialised metabolites were identified.[20] A similar story can be told about many producers of natural products – where it was thought that the capability for antibiotic production had been fully mapped out, in fact there are numerous other products which are rarely produced under standard laboratory conditions. It is estimated that 99% of microbes are currently unable to be cultivated using standard laboratory conditions – therefore, there is a likely a huge number of novel antibiotic scaffolds which we have not yet been able to access.[21,22] However, improvements in metagenomics for the capture of environmental DNA,[23] and use of heterologous production hosts (the latter of these described

further in section **1.2.3.**) is likely to assist in characterisation of a portion of the products from this group of organisms.

As of 2020, 43 new traditional antibiotics (among them, 11 beta-lactam, three tetracycline, one aminoglycoside, two macrolide and two polymyxin) and 27 non-traditional antibiotics (microbiome modulators, antibodies, immunomodulators and phage-derived treatments) are in the clinical pipeline.[24] Despite this, there is still an overall lack of antibiotics for treatment of infections caused by multi-drug resistant bacteria – only two of those currently under investigation target the critical multi-drug resistant Gram negative bacteria sub-group. As antibiotics with new modes of action are being discovered,[25,26] the question remains whether they will be clinically approved in time to stave off the crisis of antibiotic resistance.

Even as antibiotics were being rapidly discovered in the 1950s and 1960s, clinical isolates of bacteria were beginning to show resistance **(Figure 1.1)**. Despite penicillin stocks only becoming sufficient for mass-use in 1943, first reports of penicillin resistance in *Escherichia coli* and *Staphylococcus aureus* were in 1940 and 1942 respectively.[27,28] The latter of these was likely induced by the small-scale usage of the antibiotic in hospitals throughout the latter part of 1942. Methicillin-resistant *S. aureus* (MRSA) is possibly the most widely-recognized resistant bacterium, with resistance to methicillin occurring almost immediately upon clinical usage of the antibiotic;[29] in fact, recent studies suggest that the underlying transfer of the resistance genes had occurred decades earlier and was already being selected for by the application of penicillin or other beta-lactams.[30,31] Once a 'drug of last resort' for bacterial resistance to penicillins, methicillin is now no longer used in human medicine. In recent years, priorities have shifted from developing treatment for *Clostridium difficile*, once a major challenge in healthcare. After improvements in hygiene in hospital settings and reduction in use of broad-spectrum antibiotics (which allowed colonisation of the intestine by *C. difficile*),

infection with *C. difficile* is less of a clinical priority. In 2017, the WHO published a list of priorities for antibiotic research and development, where antibiotics targeting the ESKAPE pathogens (***E****nterococcus faecium*, ***S****taphylococcus aureus*, ***K****lebsiella pneumoniae*, ***A****cinetobacter baumannii*, ***P****seudomonas aeruginosa* and ***E****nterobacter* sp.) are either critical or high-level priorities.[32] These species have been implicated in most nosocomial infections and have developed resistance mechanisms against several classes of antibiotics. Carbapenem antibiotics are currently viewed as 'last resort' antibiotics for treatment of multidrug-resistant (MDR) infections – however, resistance to antibiotics in this class is becoming more common as time progresses.[33] For instance, resistance has been identified in *A. baumannii*, *P. aeruginosa*, and *Enterobacteriaceae* to the carbapenems,[32] and the resistance of *K. pneumoniae* (also implicated in nosocomial infection) to carbapenem antibiotics has now spread worldwide.[34] Resistance to colistin, another 'last resort' antibiotic, has also been seen recently: in this case, resistance is caused by a single, easily-transferrable gene called *mcr-1*, though other *mcr* genes have since been identified conveying resistance to colistin.[35–37] Colistin is often used for treatment of carbapenem-resistant *K. pneumoniae* as well as other MDR pathogens.[38]

Failure of 'last resort' antibiotics through transfer of resistance genes between bacteria has already caused deaths, a recent example being the death of a Nevada patient with *K. pneumoniae* infection resistant to all 26 antibiotics tried as treatment.[43] The 2014 O'Neill report commissioned by the UK government under David Cameron predicted that 10 million premature deaths would occur annually from 2050 onwards due to antimicrobial resistance. However, this estimate has since been criticized for a lack of empirical data and not taking uncertain mitigating factors into account, such as improvements in healthcare.[44,45] Nevertheless, antimicrobial resistance is projected to be one of the major global challenges for maintaining future healthcare systems. The Centers for Disease

Control and Prevention (CDC) estimated in 2017 that more than two million people in North America contract infections with microbes resistant to some form of antibiotic treatment annually, directly causative of 23,000 deaths.[46] Across the EU and European Economic Area in 2015, the numbers estimated via disease outcome models are similar (700,000 people with resistant infections, causing 33,000 deaths yearly).[47]

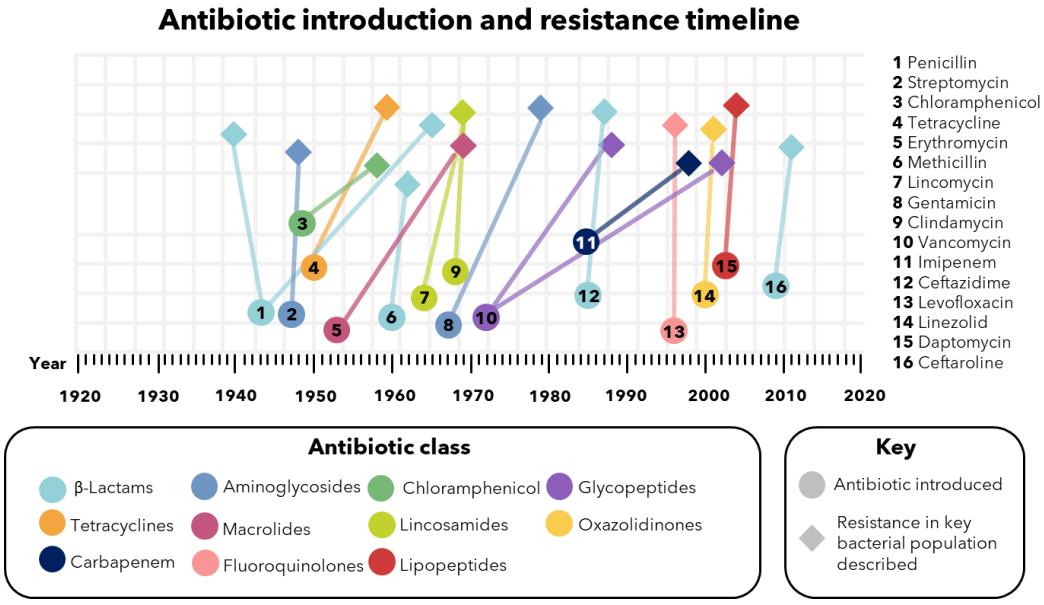


Figure 1.1: Timeline of selected antibiotic introduction and resistance. Key examples for most antibiotic classes are highlighted. Light blue = β-lactams; medium blue = aminoglycosides; dark green = chloramphenicol; orange = tetracyclines; dark pink = macrolides; light green = lincosamides; purple = glycopeptides; dark blue = carbapenem; pink = fluoroquinolones; red = lipopeptides; yellow = oxazolidinones. Where resistance has been found in two major bacterial species, two dates of resistance reported have been included. [39–42]

1.2. Synthetic biology for production of novel antimicrobials

Synthetic biology provides an interesting set of tools for the characterization of novel antibiotics identified through genome mining, as well as for improvement of antibiotic yield **(Figure 1.2)**.

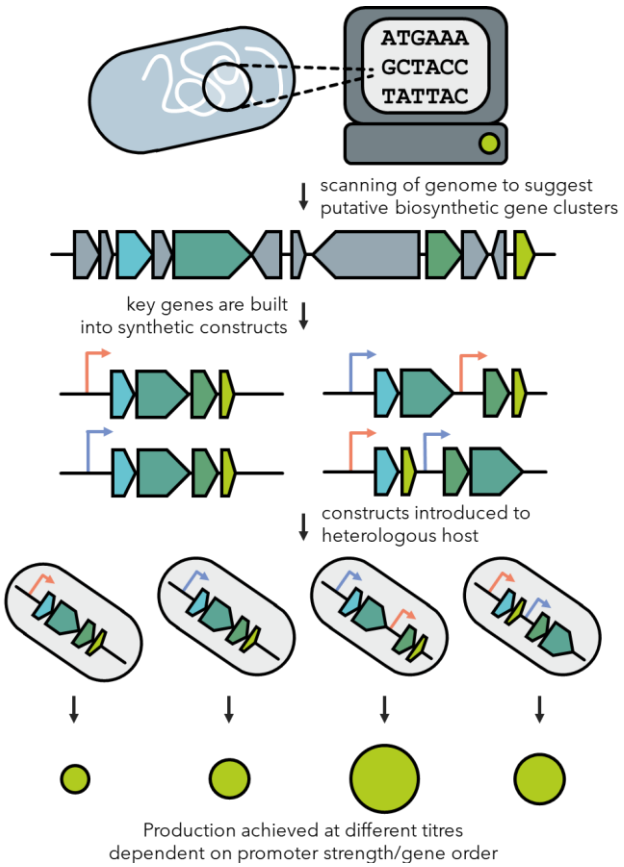


Figure 1.2: The process of biosynthetic gene cluster discovery, refactoring and expression in a heterologous host. The genome of a microbe of interest is assessed for the occurrence of BGCs using software tools. The required enzymes are rebuilt or 'refactored' into a minimal pathway required for expression. By expressing the pathway in a suitable heterologous host, expression can be achieved, and production tuned for the best possible yield of product.

1.2.1. Discovering biosynthetic gene clusters

With advances in bioinformatics tools in the genomic era, the mining of microbial genomes for antibiotic production genes – arranged in the genome as 'biosynthetic gene clusters' (BGCs) – is beginning to yield novel natural products with antimicrobial potential.[48–50] The organization of biosynthetic gene clusters allows for easy identification of biosynthesis-related genes: once a single gene in the cluster has been shown to be involved in the biosynthesis of that metabolite, the others are often found nearby.[51] Previously undiscovered due to their poor

expression in standard culture conditions, these clusters were described as 'silent' when the gene cluster has not been shown to express or 'cryptic' when the gene cluster is expressed but the resultant product is unknown.[52,53] As the cost of genome sequencing decreases rapidly (down to 454 USD per 3000 Mb as of May 2021),[54] the availability of microbial genome sequences increases: approximately 370,000 microbial genome sequences have been uploaded to NCBI (National Center for Biotechnology Information) as of the time of writing.[55] The uncharacterised gene clusters in these genomes present a huge opportunity for future natural product-based antibiotics discovery. There is an extensive range of genome mining tools available for cluster determination in both prokaryotes and eukaryotes.[51] Specialized tools are also available for the detection of clusters which may yield specific subclasses of secondary metabolites **(Table 1.1)**.

antiSMASH (antibiotics & Secondary Metabolite Analysis Shell: https://antismash.secondarymetabolites.org) is a widely used genome mining tool which allows for annotation of putative secondary metabolite gene clusters in any bacterial genome,[56–60] and more recently in fungal or plant genomes (fungiSMASH: https://fungismash.secondarymetabolites.org/; plantiSMASH: http://plantismash.secondarymetabolites.org/).[61] antiSMASH and its variants amalgamate several genome analysis tools (e.g. NCBI BLAST+, FastTree, Muscle 3, HMMer 3) into one workflow, for simultaneous comparative gene cluster analysis and predictive assignment of product output class. The capability for detection of unknown BGCs was added with the integration of ClusterFinder in 2015, relying on the idea that genomic regions rich in Pfam domains may yield interesting BGCs.[58,62] In 2017, SANDPUMA – a tool for the prediction of substrates for non-ribosomal peptide synthase adenylation domains – was included in antiSMASH.[63] CASSIS (cluster assignment by islands of sites) and SMIPS (secondary metabolites by InterProScan) have also been assimilated into the antiSMASH 4.0 workflow.[56,64] CASSIS uses the concept that the promoters of

biosynthetic gene clusters may share common sequence representative of transcription factor binding sites (assuming that all genes in a single BGC are co-regulated by the same transcription factor). SMIPS is a smaller tool which can identify 'anchor genes' – key enzymes for the synthesis of secondary metabolites i.e. polyketide synthases (PKSs), non-ribosomal peptide synthetases (NRPSs), and dimethylallyltryptophan synthases (DMATS) – in protein sequences using annotations of protein domains provided by InterProScan.[65]

Table 1.1: Cluster identification software discussed in this work. Tools range from genome mining and annotation (antiSMASH,[56–60] EvoMining,[66,67] BAGEL,[68,69] PRISM,[70,71] DeepBGC[72]) to cluster prioritization (ARTS, OOPS)[73,74]

Software	*Organisms*	*Specificity of detection*	*Described in*
antiSMASH	Prokaryotes, Eukaryotes (fungi and plants only)	N/A	56–60
BAGEL	Prokaryotes	RiPPs, bacteriocins	68,69
PRISM	Prokaryotes	Non-ribosomal peptides, type I and II polyketides, deoxygenated sugars, starter units	70,71
DeepBGC	Prokaryotes	N/A	72
EvoMining	Prokaryotes (tailored to Actinobacteria)	N/A	66,67
ARTS	Prokaryotes (tailored to Actinobacteria) based on antiSMASH output (tested on Actinobacteria)	N/A	73,75
OOPS		N/A	74

Examples of antiSMASH being utilized for antibiotic BGC discovery include Yin et al.'s analysis of the genome of *Streptomyces albolongus* YIM101047, yielding clusters with high or moderate similarity to multiple known antibiotic gene clusters for clavulanic acid, collismycin A, frontalamides, kanamycin, streptomycin and streptothricin;[49] the identification of the BGC of a two-component lantibiotic anti-*Listeria* agent in the genome of *Bacillus velezensis* GF610, hypothesised to be similar to previously described amyloliquecidin AD 2;[50,76] and the discovery of a novel circularin-like bacteriocin cluster from the genome of *Geobacillus stearothermophilus* DSM 458.[48] antiSMASH was also utilized by Nielsen and colleagues to mine the genomes of 24 *Penicillium* species to yield 1,317 putative BGCs for further analysis – a reminder that antibiotic gene cluster discovery is not solely limited to bacterial genomes.[77]

DeepBGC uses a novel deep-learning approach to mine BGCs from bacteria. While it was shown to function more accurately than ClusterFinder at discovering novel BGCs, it remains limited due to being trained on *Streptomyces* BGCs, and so may not function accurately with more diverse bacterial species. Additionally, they saw a degree of misclassification of cluster types which can be avoided in rule-based tools such as antiSMASH.[72]

Some recent genome mining tools have been focused on the analysis of actinobacterial genomes only. Actinobacteria have been widely described as being the richest producers of antimicrobial secondary metabolites [78–82] – therefore, tools directed specifically at mining their genomes may show greater accuracy in determining actinobacterial antibiotic BGCs than catch-all tools. EvoMining is a recently developed tool which takes into account three criteria from general "evolutionary concepts": i) that new enzymes evolve with the same overall reaction mechanisms as their ancestors, but increase their binding specificity for substrates;[83] ii) that families of enzymes involved in secondary metabolism are

linked to enzymes involved in primary metabolism – for example, PKSs are the evolutionary descendants of fatty acid synthetases;[84] iii) that BGCs may be an amalgamation of multiple primary-metabolism derived "sub-clusters" which have different catalytic purposes.[85,86] Taking these three criteria into account forms the general EvoMining approach: the expansion of known enzyme families to predict new outputs from clusters previously investigated by established genome mining tools. Promising results were obtained in use of EvoMining towards the identification of BGCs coding for arseno-organic metabolites – undetected by antiSMASH – from *S. coelicolor* and *Streptomyces lividans*, though the corresponding product structure must be further investigated.[67]

The Antibiotic Resistant Target Seeker (ARTS: https://arts2.ziemertlab.com/) is another recent addition to the genome-mining toolbox for Actinobacteria; it works on the hypothesis that BGCs for antibiotic production will also include a resistance gene to that antibiotic, in contrast to other functional classes of secondary metabolites.[73,75] This method of cluster determination may be a useful way to prioritise the order of gene cluster investigation: at the current time, the speed of identification of gene clusters far outweighs the speed of their characterization in the laboratory. Currently, although ARTS has been tailored to best analyse Actinobacteria genome sequences, non-Actinobacterial genome sequences can also be submitted but yield less accurate results. Another cluster prioritization tool named 'OOPS' – Output ordering and prioritization system – is available for investigation of all BGCs detected by antiSMASH, not just those which may yield novel antimicrobials. This tool prioritises BGCs based on user-defined criteria, such as optimal G+C content; the cluster's similarity to other clusters with characterised end products; presence of any internal repeats in the cluster; length of the cluster's coding sequence; phylogenetic diversity of related clusters; codon bias (to prioritise a BGC which suits an intended chassis, for example). The advantage to this tool is its tunability – whilst some using OOPS may be interested

in ease of manipulation of the DNA for refactoring of clusters, others may prioritise the discovery of completely novel products. At the current time, OOPS has been tested with Actinobacterial genome sequences, though the genomes of other microorganisms will soon follow as use of the tool increases. [74]

Other options for prokaryote genome mining are available which are limited to specific classes of natural products. BAGEL (BActeriocin GEnome mining tool: http://bagel.molgenrug.nl/) is an alternative prokaryote genome mining tool for use in explicitly identifying BGCs which produce ribosomally synthesized and post-translationally modified peptides (RiPPs) and bacteriocins.[69] Improving upon its previous iterations, BAGEL3 introduces the ability to input raw DNA sequences instead of annotated genomes, and broadens the detection of gene clusters to include RiPPs (examples described include cyanobactins, sactipeptides, and linaridins). A range of antimicrobial peptide-encoding gene clusters has been identified using BAGEL3, from *Staphylococcus capitis* TE8,[87] *Bacillus pumilus*,[88] and *Bacillus amyloliquefaciens*.[89] For example, whilst Belbahri et al.'s genome-mining work on plant-associated bacteria *B. amyloliquefaciens* describes known BGCs for fengycin, bacilysin, and mersacidin (amongst others), it also discusses hundreds of unknown secondary metabolites of multiple subclasses (e.g. microcins, bacteriocins, and NRPSs) requiring further characterization from numerous genome-sequenced *B. amyloliquefaciens* strains. This is a potential 'gold-mine' of peptide antibiotics.[89] BAGEL was updated to version 4 in 2018 to improve RiPP domain prediction.[68] Another tool for the genome annotation of prokaryotes is PRISM (Prediction Informatics for Secondary Metabolomes: http://magarveylab.ca/prism/), which, alongside genome mining, focuses upon chemical structure prediction of non-ribosomal peptides and type I and II polyketides.[70] The combined use of PRISM 3 with antiSMASH 4.0 by Qi et al. recently identified 32 uninvestigated natural products from the genome of *Streptomyces* sp. strain JV178, including those in the bacteriocin, polyketide, and

hybrid polyketide/nonribosomal peptide classes.[90] PRISM 4 was recently released, which has improved upon chemical structure prediction of 16 different secondary metabolite classes.[71]

To improve the study of BGCs and increase the chances of finding bioactive secondary metabolites, the MIBiG (Minimum Information about a Biosynthetic Gene Cluster: https://mibig.secondarymetabolites.org/) repository has been developed. This is a key database for the input of natural product cluster sequence and the annotations derived from any of the above genome mining tools. Those using the tool can input a description of the gene cluster; annotate domains of individual enzymes present; display known homologous gene clusters with direct links to their accession page; show the expected chemical structure of the cluster product along with any bioactivities; and provide details on any literature published about the specific gene cluster. MiBIG is an incredibly useful tool for both the reproducing of confirmed results and the investigation of novel gene clusters. [85]

Computational methods have enabled us to analyse the genomes of potential antibiotic-producers and identify thousands of new biosynthetic gene clusters. Further techniques must be utilized to awaken biosynthetic gene clusters, produce new antimicrobial natural products, and characterize those metabolic outputs.

1.2.2. Refactoring biosynthetic gene clusters

After discovery of putative gene clusters, the process of refactoring may be undertaken to produce a minimal construct for the generation of the secondary metabolite of interest. Refactoring aims to rebuild biosynthetic pathways minimally, excluding native regulatory elements – which may hamper production of secondary metabolites – and instead using well-characterised regulatory elements to increase product yield. Often, this refactoring is required to achieve production – by switching out promoters, or altering the expression of regulatory

genes, production of the secondary metabolite can be switched on.[91–93] A modular approach has become more popular in refactoring recently – by treating individual cluster enzymes and other DNA regulatory elements as 'building blocks', designer biosynthetic pathways can be constructed through carrying out combinatorial biosynthesis. [94,95] Manipulation of the DNA of the gene cluster through refactoring presents an interesting avenue to awaken gene clusters, and multiple different ways of doing this have been investigated in recent years.[91–96]

Plug-and-play scaffolds for BGC refactoring are perhaps the most interesting of these approaches—enabling high-throughput research into the best possible conformation of genes not only for cluster awakening, but potential improvement of product titre. In 2013, a plug-and-play scaffold for the refactoring of BGCs was introduced by Shao et al.[94] This workflow used a single vector which contained, as well as the BGC for expression, "expression host helper modules" – organism-specific origins of replication, selection markers, and integrases for best possible heterologous expression. In the cluster itself, strong heterologous host promoters were inserted upstream of cluster genes. Promoters of *gapdh* and *rpsL* from *Streptomyces griseus* were identified as being potentially strong promoters from real-time PCR experiments, and the strength of the promoters upstream of *gapdh* and *rpsL* homologues in 18 *Streptomyces* strains were investigated using a *xylE*-catechol detection assay in *S. lividans.* This yielded a suitable set of 13 candidates out of the 36. The functionality of this system was tested through the refactoring of the spectinabilin BGC in *S. lividans*, with yields of 105±21 μg L^{-1} of spectinabilin being achieved. In 2017, a different proof-of-concept approach was utilized by Ren et al. for the combinatorial biosynthesis of beta-carotene, resulting in the successful construction of 96 different pathways for carotenoid biosynthesis.[95] However, their pipeline was not tested towards awakening a cryptic biosynthetic gene cluster, which would be the true test of the approach.

One well known example of refactoring via a highly combinatorial approach is the refactoring of the *nif* cluster by Smanski et al.[97] This cluster contains 20 genes and allows for the nitrogen fixation carried out by *Klebsiella oxytoca*. In their work, they begin with a previously refactored *nif* cluster (only including all genes deemed essential and removing non-coding DNA and regulators). Across the entire approach, they tested a library varying the gene orders, orientations and locations of promoters, ribosome binding site strength, and codon usage. The authors also tested several unusual designs in their libraries, in some cases including multiple promoters or using only transcription read-through for expression. The best performer of the completely synthetic *nif* cluster library had 57% of wild-type activity, a significant increase upon prior attempts without this total reshuffling of the cluster. However, several combinations of genetic components in the library had adverse effects on growth, with some correlation between number of transcriptional units leading to poor growth. In addition, transfer of the best performing candidate from *K. oxytoca* to *E. coli* only maintained 7% activity in comparison to the wild type (and upon creating 20 new clusters by changing RBSs to be better suited to *E. coli*, a further increase of 1% was seen.) The work highlights the difficulties in rationally designing a pathway from scratch, especially in non-model organisms.

Alteration of the expression of regulatory genes also presents a target for refactoring. Disruption of global regulators for unknown activation of BGCs, and disruption of 'pathway-situated regulators' for the targeted activation of BGCs, are approaches which have been utilized in recent years.[91,96] For example, the activation of the BGC for oviedomycin production was achieved through homologous-recombination mediated disruption of the gene encoding the global regulator AdpA in *Streptomyces ansochromogenes*.[96,98] This regulator has been shown to affect the secondary metabolite metabolism in other *Streptomyces* spp., examples being the regulation of StrR and GriR for streptomycin and grixazone

biosynthesis respectively.[99,100] Alteration in pathway-specific regulatory gene expression can be combined with promoter insertion for a tandem refactoring approach, as shown by Olano et al.[91] Insertion of strong constitutive promoters through homologous recombination alongside overexpression of positive regulatory genes yielded the products of all five chosen gene clusters of *Streptomyces albus* J1074. Through insertion of promoters, blue pigment indigoidine and two new polycyclic tetramate macrolactams 6-epi-alteramides A and B were yielded. Overexpression of positive regulators caused the production of antimycin, candicidin, and paulomycins from the three remaining BGCs. It is interesting to note that this study utilized overexpression of both endogenous and heterologously expressed regulatory genes: the first being a LuxR-family gene sshg_05706 identified 7.6kb upstream of the *S. albus* J1074 cluster for a hybrid PKS-NRPS, later revealed to be antimycin. The second, PimM – a regulator for the *S. natalensis* pimaricin pathway – was inserted into the *S. albus* J1074 chromosome and activated a cluster for candicidin production as well as the nearby antimycin cluster.[91] Whilst overexpression of host regulatory genes has been successfully utilized for overproduction of secondary metabolites many times,[101–103] the use of heterologous regulators for this purpose may become more popular in the future – especially with the high levels of cross-regulation seen between different *Streptomyces* secondary metabolic pathways.[104,105]

1.2.3. *Streptomyces* spp. as heterologous expression hosts

Heterologous expression is another strategy many studies have employed when the goal is to awaken silent gene clusters, improve product titre, or produce antibiotics in a more genetically tractable or faster-growing organism. Heterologous expression in a suitable chassis circumvents the difficulty of culturing some uncharacterised microbial species which may still have unknown optimal culture conditions. Many have turned to the Actinomycetes – in particular, *Streptomyces* spp. – for heterologous expression, and in general there

seems to be a tendency towards using chassis that are closely related to the native host organism. Actinomycetes are a popular choice of host for the heterologous expression of biosynthetic gene clusters as they frequently have the same required precursor molecules, the same chemical skeleton-modifying enzymes, and a similar G+C content as other prolific antibiotic producers.[106] *Streptomyces* as a group are known to be some of the richest bacterial producers of secondary metabolites,[107] and have been shown to utilise this for symbioses across multiple kingdoms of life.[108–112] *S. coelicolor* is a common chassis for production of novel antibiotics due to the organism's high level of characterization and the availability of several genome-minimized strains. In one commonly utilized strain (*S. coelicolor* M1146), 4 large secondary metabolite BGCs have been deleted (actinorhodin, undecylprodigiosin, coelimycin, and calcium-dependent antibiotic) to reduce the metabolic burden when introducing heterologous BGCs. Point mutations in two genes (RNA polymerase subunit B *rpoB* and ribosomal S12 protein *rpsL*) were also introduced to increase the levels of endogenous secondary metabolites (*S. coelicolor* M1154).[113] As many antibiotics originate from *Streptomyces* spp., using a well-characterised *Streptomyces* strain as a heterologous host is an approach which has been shown to be successful. A review was recently published by Breitling et al. focusing on the myriad of approaches used for heterologous expression in *Streptomyces*.[114]

Multiple DNA manipulation approaches have been utilized in recent years for the introduction of heterologous BGCs into *S. coelicolor* **(Figure 1.3)**. These include linear-plus-linear homologous recombination (LLHR), Transformation-associated recombination (TAR), and a streptomyces artificial chromosomal conjugation vector (pSBAC).[115–117] A review by Nah et al. highlights selected examples of heterologous expression in *Streptomyces* over the past 30 years.[118]

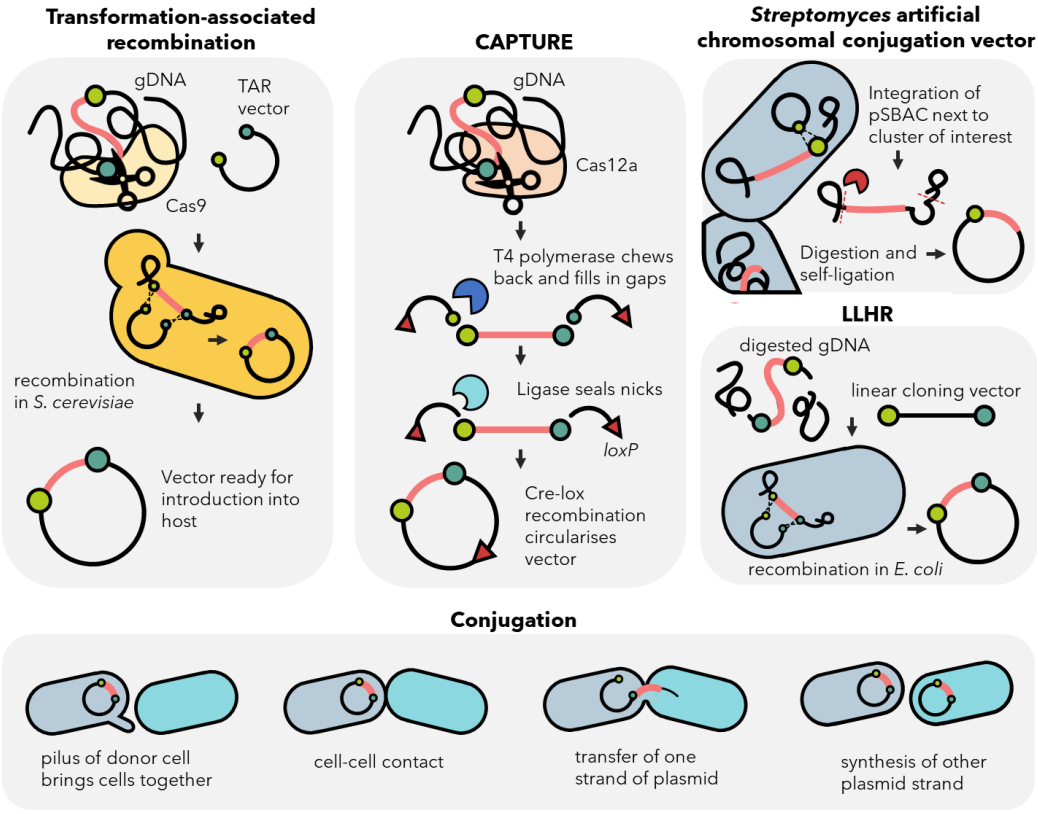


Figure 1.3: DNA manipulation approaches for heterologous expression in *Streptomyces*. Transformation-associated recombination (TAR), CAPTURE, Linear plus linear homologous recombination (LLHR), and *Streptomyces* artificial chromosomal conjugation vector are all techniques utilized for heterologous expression. **TAR:** Cas9 is used to digest genomic DNA. Digested gDNA is co-transformed into *S. cerevisiae* along with TAR vector with custom "hook" regions. Homologous recombination occurs, and plasmid can be purified for expression in heterologous host. **CAPTURE:** Cas12a digests gDNA for specific fragment. Two DNA receivers with custom overhangs can ligate to the fragment of interest after T4 polymerase chews back the DNA to form ssDNA overhangs. Upon addition of dNTPs, the gaps are filled in, and a ligase is added to seal the nicks. Recombination at *loxP* sites allows for vector circularisation for introduction to host. **LLHR:** After addition of specific flankers to a linear vector, recombination occurs between two linear DNA fragments to form a circular plasmid. ***Streptomyces* artificial chromosomal conjugation vector:** pSBAC undergoes site-specific recombination with region flanking the cluster of interest. Restriction digest of purified DNA and self-ligation occurs to form whole pSBAC containing cluster of interest which can be conjugated into strain of choice. **Conjugation:** Once constructed, plasmids can be introduced to *Streptomyces* as shown.

To create plasmids with BGCs for heterologous expression in *Streptomyces*, there are many different options. TAR cloning is a method used to capture large sections of DNA for further manipulation; here, the natural homologous recombination capability of *Saccharomyces cerevisiae* is used. By creating a vector with 'hooks' (representing identical sequence to the ends of a region of genomic DNA of interest) and transforming both digested genomic DNA and the vector into *S. cerevisiae*, up to 300 kb of DNA can be captured. This then allows for the engineering of large BGCs. TAR-based methods have been utilized to heterologously express the taromycin BGC from *Saccharomonospora sp*. CNQ-490 in *S. coelicolor* to characterize the lipopeptide antibiotic taromycin A.[119] The gene cluster for taromycin A is very similar in sequence to that of the clinically approved antibiotic daptomycin, but the two products show some slight differences in structure. Recent work has been published after a second taromycin—taromycin B—was identified from the same gene cluster which, like taromycin A, shows antibiotic activity against MRSA and vancomycin-resistant enterococci.[120] Unfortunately, neither of these new antibiotics shows greater bioactivity than daptomycin, and so further work may be required to increase the bioactivity or to carry out further genome mining for more taromycins. TAR-based methods were also used in the heterologous expression of the BGC for enterocin, captured from the genome of *Salinispora pacifica*.[121] In this case, enterocin was successfully produced by *S. coelicolor* M1146 and *S. lividans* TK23, with no significant difference in production quantities to that of the native producer.

The TAR system has been shown in many cases to be a useful method in the direct capture of antibiotic BGCs from the genomes of the microbes which naturally produce these antibiotics. However, several steps in this method—such as introducing DNA into *Saccharomyces cerevisiae*, and then into *E. coli*—mean that the process is more time-consuming than competitive approaches. LLHR is an

alternative method that was used by Yin et al. for cloning the salinomycin BGC into *S. coelicolor*.[122] In this approach, recombination can occur between two linear DNA strands to form a circular plasmid which is then suitable for conjugation into a heterologous host.[115] There is some difficulty in cloning large, G+C-rich fragments with this technique: in the case of Yin et al., the 106 kb BGC had to be divided into three fragments and assembled together to ensure the process occurred efficiently.

A non-homologous-recombination-mediated method using a *Streptomyces* artificial chromosome conjugation vector has been used for the capture of multiple antibiotic BGCs. The vector, pSBAC, has the key property of functioning as a shuttle vector between *E. coli* and *Streptomyces* spp. Integrating the vector into the genome site-specifically and excising the combined vector and desired cluster sequence allow for heterologous overexpression of large BGCs. This has been showcased with the integration of the biosynthetic gene cluster for tautomycetin production into the genome of *S. coelicolor*;[123] and, more recently, the gene cluster for pikromycin was integrated into the *S. coelicolor* genome in the same way.[124] Integration of the pikromycin BGC into the *S. coelicolor* genome only produced 10-deoxymethynolide, a major derivative compound (at yields of 346.8 and 396.2 mg L^{-1} in *S. lividans* and *S. coelicolor*, respectively); however in *S. lividans* the end product pikromycin was produced, though at varying amounts (reported yield: 333.7 mg L^{-1}).

A recent work by Enghiad et al. utilises a Cas12a-mediated cloning method for the cloning of natural product BGCs. The technology, termed 'CAPTURE', begins with Cas12a digestion of the region of interest out from the host genome. Two DNA receivers (one containing an *E. coli* origin of replication, the other an antibiotic resistance marker; both containing *loxP* sites) are then treated with T4 DNA polymerase to generate overhangs for ligation, the gaps filled in with

addition of dNTPs, and ligase added to seal the nicks. By introduction of this product into a specialised *E. coli* strain expressing Cre recombinase and the phage lambda Red Gam protein, the receiver *loxP* sites can be used to circularise the region of interest. Enghiad et al. used this new technology for the cloning of 47 BGCs with huge variation in size (the largest being 113 kb), speeding up the heterologous expression process to 3-4 days from approximately a month using classical cosmid cloning techniques.[125]

Traditional cosmid construction and conjugation approaches are still being routinely used—for example, the gene cluster for chuangxinmycin production was introduced into *S. coelicolor* M1146 recently, achieving production of chuangxinmycin and elucidating a biosynthetic pathway for this antibiotic through disruption of key pathway genes *cxnA–F*.[126] Additionally, the kocurin gene cluster was recently introduced into *S. coelicolor* through traditional conjugation methods—however, the yield of kocurin was extremely low, attributed to the difficulty previously seen when expressing thiopeptides in multiple *Streptomyces* spp. due to very strong influence of media components on production.[127–129] For example, in a previously reported study, thiopeptide GE2270 was expressed by *S. coelicolor* in only one media condition out of 35 tested.[130] The large BGC (141 kb) of the antibiotic vancoresmycin was also recently conjugated into *S. coelicolor* to elucidate its mode of action and biosynthetic pathway. This BGC is hypothesized to be the largest heterologously expressed gene cluster at the time of publication.[131]

Recently, use of other *Streptomyces* species as heterologous hosts has become more common. As with *S. coelicolor*, construction of a number of genome-minimized *Streptomyces avermitilis* hosts (SUKA2-22) has been carried out for the heterologous expression of over 20 natural product BGCs and is an option for

heterologous production of avermectin antibiotics.[132] Additionally, several *Streptomyces lividans* TK24 chassis were recently generated through sequential deletion of native BGCs by Novakova et al.[133] The production levels of the heterologous BGC for mithramycin were increased to nearly 3 g L^{-1} when the *red*, *act*, and *cda* BGCs were deleted from the genome. Further deletion of the *cpk* BGC (for the same deletions as the earlier described *S. coelicolor* M1146 strain—*Δact Δred Δcda Δcpk)*[113] did not increase production levels further. The *cpk* BGC in *S. lividans* is unexpressed or expressed at undetectable levels—consequently, the product of this BGC will not compete for precursor units and affect heterologous production as coelimycin P1 (the output of *cpk* in *S. coelicolor*) does.[133] *Streptomyces gilvosporeus* was chosen as the heterologous host for the natamycin BGC from *Streptomyces natalensis*, as it has been shown to be one of the other natural producers of this antibiotic.[134] In this case, three consecutive rounds of random mutagenesis were also carried out to eventually increase product titre to 340% of the baseline value (110% (6.64 ± 1.38 g L^{-1}) → 230% (10.49 ± 0.47 g L^{-1}) → 340% (14.11 ± 0.73 g L^{-1}) natamycin production compared to wild type in each of the rounds of mutagenesis. The approach of using a chassis organism in which the BGC is naturally present in the genome may be useful in other overexpression strategies—particularly if the pathway requires a rare metabolite as a precursor—but is only a viable method if the heterologous host is genetically tractable and well-characterised.

S. albus J1074 is another commonly used chassis for heterologous expression of secondary metabolite gene clusters due to its relatively small genome size.[135] Manderscheid et al. recently took an interesting approach to increase the number of copies of BGCs to overexpress secondary metabolites through integrating additional φC31-attachment sites into the *S. albus* J1074 genome.[136] Heterologous expression of pamamycin, mensacarcin, aranciamycin, and griseorhodin gene clusters gave higher levels of expression in correlation to

the additional number of φC31-attachment sites—and therefore, extra copies of the BGC in the genome. This work highlights a different approach for overexpression of BGCs through introduction of multiple copies and may be useful in cases where BGC expression is at too low a level for adequate product characterization.

Another recent use of *S. albus* J1074 was for the heterologous expression of malacidins, calcium-dependent antibiotics newly discovered from desert soil eDNA fragments using a metagenomic approach for production of novel compounds.[137] TAR cloning was utilized to capture the DNA sequence of three cosmids containing the 72kb malacidin gene cluster after library construction, and to collect them into a single construct for heterologous expression. The malacidins elucidated showed antibiotic activity against gram-positive bacteria resistant to antibiotics currently in clinical use—even those resistant to vancomycin, which has a similar mode of action of binding cell-wall precursor lipid II.[138] The original producer of the malacidins is still unknown, a rare case in the history of natural product discovery.

Despite advances in recent years, heterologous expression can remain challenging. Success often relies on further engineering of the gene cluster or host strain, significantly increasing the time spent on generation of compounds. In the review by Nah et al., 84 examples of heterologous expression of natural product gene clusters in actinomycetes are described – of these, only 20 examples showed higher titres in the heterologous host than in the native producer, and many increases were marginal.[118] The location of the gene cluster insertion may also have significant effect on the resultant titres – in the work by Bilyk et al. (2017), up to 8-fold change in production of target compound aranciamycin could be seen simply by changing the location of the cluster through transposon mutagenesis of the *S. albus* host strain.[139] Iqbal et al. (2016)were aware of the challenges with host strain

choice in their work to optimise metagenomic library expression, and so prior to introduction of the library, extensive experiments to choose a host strain were carried out. 39 different *Streptomyces* strains were investigated for polyketide production from 97 cosmids containing PKS from eDNA; the authors suggest that the metabolites of interest are more likely to show pigmentation, and so screening for compound production would be easier. From the original 39 strains, only four species showed production of coloured compounds, and after introduction of their 1.5 million-membered cosmid library, only one novel metabolite was able to be identified.[140]

However, synthetic biology methods facilitate the necessary engineering of both cluster and chassis that makes such long-range transfers possible: in particular, heterologous expression is often coupled with, e.g., refactoring of the cluster with new regulatory elements;[91] optimization of precursor supply in the intended chassis by metabolic engineering;[141,142] or creation of a synthetic microbial consortium for efficient substrate utilization and the best possible yield of product.[143,144] In some cases, the strategy towards cluster awakening is unclear: HiTES (high-throughput elicitor screening) has proven to be a method of interest for awakening cryptic gene clusters in their native hosts. After attempting induction of expression with a library of 502 natural products of various classes, cluster gene expression can be measured by either fluorescence readout (in the case where a reporter gene has been inserted to the cluster) or more recently by imaging mass spectrometry.[145,146] The more recent work with HiTES led to production of nine cryptic metabolites.[145]

1.2.4. CRISPR-Cas-mediated strain editing for heterologous production

1.2.4.1. Function and mechanisms of CRISPR-Cas9

CRISPR-Cas systems, as a suite of gene editing tools, come from our harnessing of an endogenous bacterial defence system. CRISPR-Cas systems have been identified in 42.3% of bacterial genome sequences as of 2020,[147] and are becoming an incredibly widespread tool in synthetic biology and biotechnology for strain engineering and better heterologous production of secondary metabolites. The most commonly used CRISPR systems are Class 2 CRISPR systems, which require only one Cas protein for full functionality; Class 1 CRISPR systems require multiple.[147] The type II Cas endonuclease Cas9 from *Streptococcus pyogenes* (spCas9) is the best characterised and most commonly used, as it is a relatively simple system to design.[148]

The two components required for double-stranded DNA cleavage are the Cas protein itself, and the guide RNA, which targets the site-specific modification of the DNA template. SpCas9 has two domains with nuclease activity – an HNH domain and a RuvC-like domain, which are positioned so that they cut the targeting and non-targeting strand of DNA respectively.[149] The RNA which associates with the Cas protein is collectively called the single guide RNA (sgRNA), and consists of a CRISPR RNA (crRNA) which has a sequence specific to that of the cut site, as well as a trans-activating CRISPR RNA (tracrRNA), which recruits the Cas9 endonuclease.[150] The versatility of CRISPR-Cas as a system comes with the ease of reprogramming the sgRNA, as there is no laborious engineering of the Cas9 nuclease required.[151]

The system works by the formation of a Cas9-sgRNA complex which detects small motifs in the genome – termed protospacer-adjacent motifs (PAMs). For spCas9, the canonical motif is 'NGG', where N is any nucleotide **(Figure 1.4)**. Upon

detection of this motif, the Cas9-sgRNA complex probes to see whether there is a complementary region between the dsDNA of the genome and the sgRNA. If this is found, the complex binds to the dsDNA of the genome site-specifically and the conformational changes in the endonuclease allow for endonuclease activity to be triggered and the subsequent generation of a double-stranded break (DSB) in the DNA. This break will then be repaired. With a donor DNA supplied, custom insertions and deletions can be made through the action of the cell's own homology directed repair (HDR) pathway.

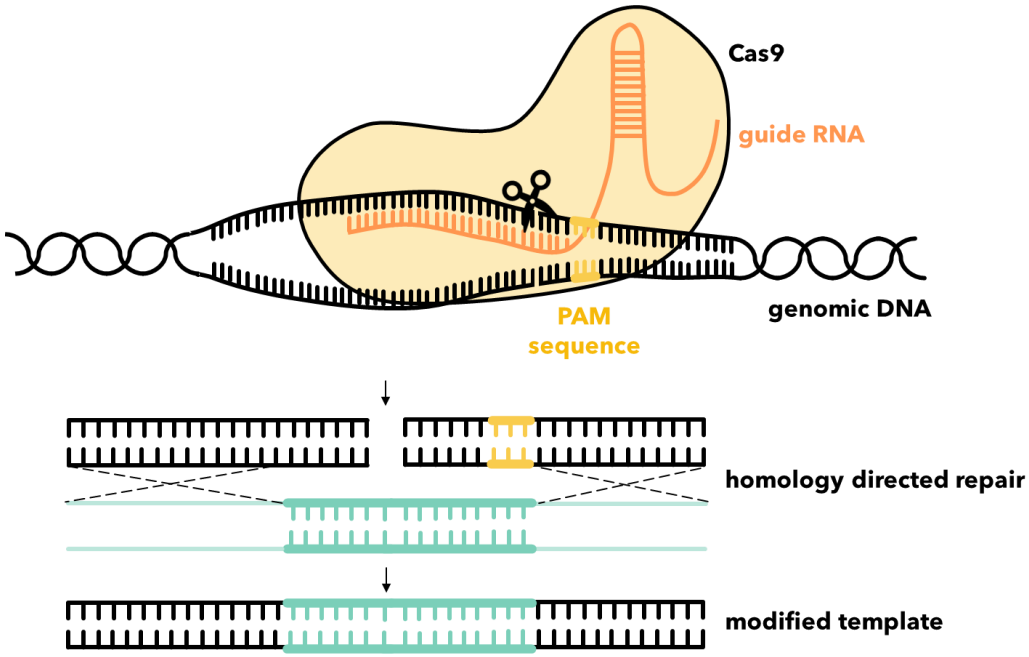


Figure 1.4: Mechanism of CRISPR-Cas9-mediated gene editing. The system consists of two main components: the Cas9 endonuclease, and a single-stranded guide RNA (sgRNA). The Cas9 endonuclease is guided to cleave DNA specific to the sequence of the sgRNA. Cleavage is guided by the position of an 'NGG' protospacer adjacent motif (PAM) directly occurring next to a 20 bp protospacer. Cas9 cleavage causes a double-strand break in the DNA which can be repaired by the error-prone non-homologous end-joining (NHEJ) pathway or the high-fidelity homology-directed repair (HDR) pathway. By supplying a repair template to the cells, deletions or insertions can be easily made.

CRISPR-Cas9-mediated errors can also be made if the cell uses non-homologous end-joining (NHEJ) to repair the DSB caused by Cas9, yielding deletion and

insertion mutations.[148] These off-target effects are the main drawback to this otherwise elegant technology. The original CRISPR-spCas9 system has been said to show off-target activity on DNA sequence with mismatches between 3 and 5 bp in the sequence of the protospacer furthest from the PAM, as well as a low tolerance of a 5′ – NAG – 3′ PAM in the genome (one-fifth of the binding efficiency of the typical 5′ – NGG – 3′ motif).[152,153]

1.2.4.2. Improvement of CRISPR-Cas in actinomycetes

Since the original harnessing of the CRISPR-Cas system, researchers have focused on ways to improve its specificity and therefore reduce deleterious off-target effects. The main relevance of this technology to this work is the extensive use towards the engineering of natural product BGCs – therefore, the focus will be upon those advances in the actinomycetes, a group rich in natural products.

There are currently four very commonly used CRISPR-Cas9 systems for modification of actinomycetes – pCRISPomyces-2,[154] pCRISPR-Cas9,[155] pKCcas9dO,[156] and pWHU2653,[157] all developed in 2015. Each of these is a single plasmid vector containing a codon-optimized spCas9 for expression in actinomycetes, the required sgRNAs, and the editing template. More recently, three other Cas nucleases have been shown to be functional in various *Streptomyces* spp. by Yeo et al., expanding this toolbox for strains where spCas9 appears to be non-functional or yields non-viable strains due to toxicity.[158–160]

Cas9 toxicity has been shown to be a problem in some *Streptomyces* strains, and various toolkits have been prepared to alleviate the problem of no exconjugants growing after introduction of Cas9 sequence into the cell. In the work carried out by Ye et al. (2020), Cas9 is expressed either under the tight control of a theophylline riboswitch, or under a weaker constitutive promoter than in the original pCRISPomyces-2 plasmid to improve viability of exconjugants carrying

CRISPR machinery.[161] In the work of Wang et al. (2020), thiostrepton, theophylline, and blue light are used for the control of Cas9 expression transcriptionally, translationally, and at the protein level. Though editing efficiency remained at ~10%, the transformation efficiency into *S. coelicolor* sharply increased, yielding viable exconjugants.[162] Wang et al. expanded on this strategy to swap the blue-light-inducible split-Cas9 with a combination of standard Cas9 and inhibitor AcrIIA4. In their more recent work, Cas9 expression was still controlled by the thiostrepton inducible promoter and theophylline riboswitch; it was hypothesised that these caused the expression of Cas9 to very low levels, and the low-level constitutive expression of AcrIIA4 inhibited what basal activity was occurring in uninduced conditions. As AcrIIA4 was expressed under a medium-strength promoter ermEp*, upon induction it would not be able to be produced in high enough quantities to inhibit the now strongly-expressed Cas9, and so editing activity would resume.[163,164]

Modification of repetitive sequences can be problematic using CRISPR-Cas, especially for natural products such as polyketides which function as a repetitive modular system. A modification to the typical CRISPR-Cas9 workflow was described in the work carried out by Najah et al. on *Streptomyces ambofaciens*. A 'bait' DNA sequence was chosen to be incorporated through homologous recombination into the target DNA which would then be targeted by spCas9, rather than target DNA being cleaved directly.[165]

To increase the speed at which strains can be generated by this technology, Wang et al. generated an improved method of screening for modifications by CRISPR-Cas9. Use of *gusA* or *idgS* systems as chromogenic reporters incorporated into the plasmid carrying the CRISPR machinery was shown in *S. coelicolor*, as well as in a rare actinomycete, *Verrucosispora* sp. MS100137. By incorporating either of these reporters into existing CRISPR-Cas9 system pWHU265, colonies which

maintained colour could be confirmed to have been successfully conjugated with CRISPR plasmid pQS-gusA/idgS, while those which were then cultured again and lost colour had successfully lost pQS-gusA/idgS.[166] This increase in speed of screening could be particularly important in rare actinomycetes which may have significantly longer growth times than commonly-used laboratory strains.

Finally, CRISPR-BEST (CRISPR-Base Editing SysTEM) is perhaps the most major modification to CRISPR technology in streptomycetes, completely removing the requirement of causing DSB for modification to occur. Two plasmids for the base-editing of cytosine to thymine (CRISPR-cBEST) and adenine to guanine (CRISPR-aBEST) were constructed. CRISPR-cBEST contains a fusion protein of cytidine deaminase rAPOBEC1, a mutated Cas9 nickase, and uracil glycosylase inhibitor (UGI). It functions through the conversion of cytosine to uracil and promotion of the cellular mismatch repair pathway for the replacement of guanine for adenine (and therefore, original cytosine for thymine). CRISPR-aBEST contains adenosine deaminase ecTadA and the same mutated Cas9 nickase; it functions through deamination of target adenine to an inosine, which is read as guanine by DNA polymerase during DNA replication. This technology was used for the introduction of stop codons in *S. coelicolor* A3(2), and non-model strains *Streptomyces collinus* Tü365, and *Streptomyces griseofuscus* DSM 40191, showing simultaneous mutation of the duplicated gene cluster kirromycin in *S. collinus* Tü365. Though this new system appears to reduce the number of off-target effects, the massive insertions the normal CRISPR-Cas9 systems are capable of are not possible.

1.2.4.3 Applications of CRISPR-Cas in natural products research

In terms of applications of CRISPR in the field of natural products research, the technology is slowly being used for the characterization of novel biosynthetic gene clusters, as well as cluster awakening. It is particularly relevant in non-model

actinomycetes which can be biosynthetically talented and had not been able to be modified using previous tools, though revision of the currently available CRISPR systems to suit non-model actinomycetes is currently an area of interest.

Streptomyces formicae was identified from worker ants of the fungus-growing *Tetraponera penzigi* plant-ants collected in Kenya, and identification of the polyketide antibiotics formicamycins soon followed. To identify the biosynthetic gene cluster responsible for production of these new antimicrobials, CRISPR-Cas9 was used for the deletion of first the full cluster and then deletion of one pathway enzyme to further elucidate the mechanisms of biosynthesis.[167] A similar approach was used in the work by Low et al. (2018) on a *Streptomyces* strain that was collected from mangrove sediment in Singapore, using a deletion of a posited gene cluster to determine whether it was responsible for production of sceliphrolactam, and fusion of two genes involved in the biosynthesis to further determine the mechanism.[159]

A general approach for increase of natural product yields has been described by Zhang et al. in 2017. In their case, knock-in of promoters into BGCs in five different *Streptomyces* spp. yielded metabolites which were previously undetected when culturing these strains. Most importantly, through doing this, they were able to awaken the cluster for production of a novel pentangular type II polyketide in *S. viridochromogenes* with a dihydrobenzo[α]naphthacenequinone core.[168] Particularly for heterologous expression, small modifications such as insertion or swapping of promoters may make expression of the products to detectable levels much easier. Use of mCRISTAR allows for the replacement of multiple native BGC promoters with designed alternatives which may be more suitable for specific applications. Transformation-associated recombination, a method used in other studies for the capture and isolation of BGCs for heterologous expression,[119–121] is here used in combination with CRISPR for the reassembly of Cas9-digested

fragments.[93] Designed promoters can be inserted site-specifically as long as they share at least 40 bp homology with the fragments on either side. Currently the system has been shown to work with a specific set of promoters for a specific BGC (tetarimycin gene cluster) expressed in *S. albus* J1074, though future work may illustrate whether this method has wider-ranging capabilities. miCASTAR was recently described by the same group, a modification of mCRISTAR with the Cas9 digestion of fragments carried out *in vitro* prior to use of TAR cloning for the reassembly of fragments. The authors suggest that this modification speeds up the process by approximately a week.[169]

Lastly, CRISPR-Cas9 has been used for the characterisation of low-producing gene clusters via the turning off of highly-expressed antibiotic gene clusters which mask production of other compounds. Culp et al. generated streptothricin and streptomycin gene knockouts in several different actinomycete strains. These antibiotics are frequently rediscovered in culture extracts and so remain a barrier for the discovery of new scaffolds. Three of their knock-out strains (one streptothricin-inactivated, and two streptomycin-inactivated) showed activity against an *E. coli* indicator strain, and were then probed via LC-MS. All three strains showed production of rarer compounds.[170]

1.3. Aminoglycoside antibiotics

The class of aminoglycosides originated with the discovery and isolation of streptomycin in 1944 from *Streptomyces griseus,* with many of the other naturally-derived aminoglycosides soon following during the golden age of antibiotic discovery.[171,172] Aminoglycosides have activity against Gram-negative bacteria and some Gram-positive bacteria, showing specifically strong activity to members of the Enterobacteriaceae family (*E. coli*, *Klebsiella* spp., *Serratia marcescens*), other Gram-negatives such as *Yersinia pestis*, as well as some gram-positive bacteria including vancomycin- and methicillin-resistant *Staphylococcus*.[172–174] As such, this class of compounds remains therapeutically relevant for the treatment of serious infections with Gram-negative bacteria.[172] Typically, aminoglycosides produced by members of the *Streptomyces* species end in -mycin,[175,176] while those isolated from *Micromonospora* spp. will end in -micin.[177] This nomenclature allows for identification of bacterial origin at first glance.

Aminoglycosides share a common core structure – a dibasic aminocyclitol, usually 2-deoxystreptamine (2-DOS; or streptidine in the case of streptomycin) as the central sugar, connected with glycosidic bonds to amino sugars usually at the 4th and 5th/6th carbon **(Figure 1.5)**. These amino sugars are frequently modified with different hydroxyl, amine, or amino groups to create a wide diversity in antimicrobial activity and mode of action.[172] A general structure-toxicity correlation has been observed for the aminoglycosides, with deamination and deoxygenation across the whole molecule causing a decrease and increase to toxicity respectively.[179]

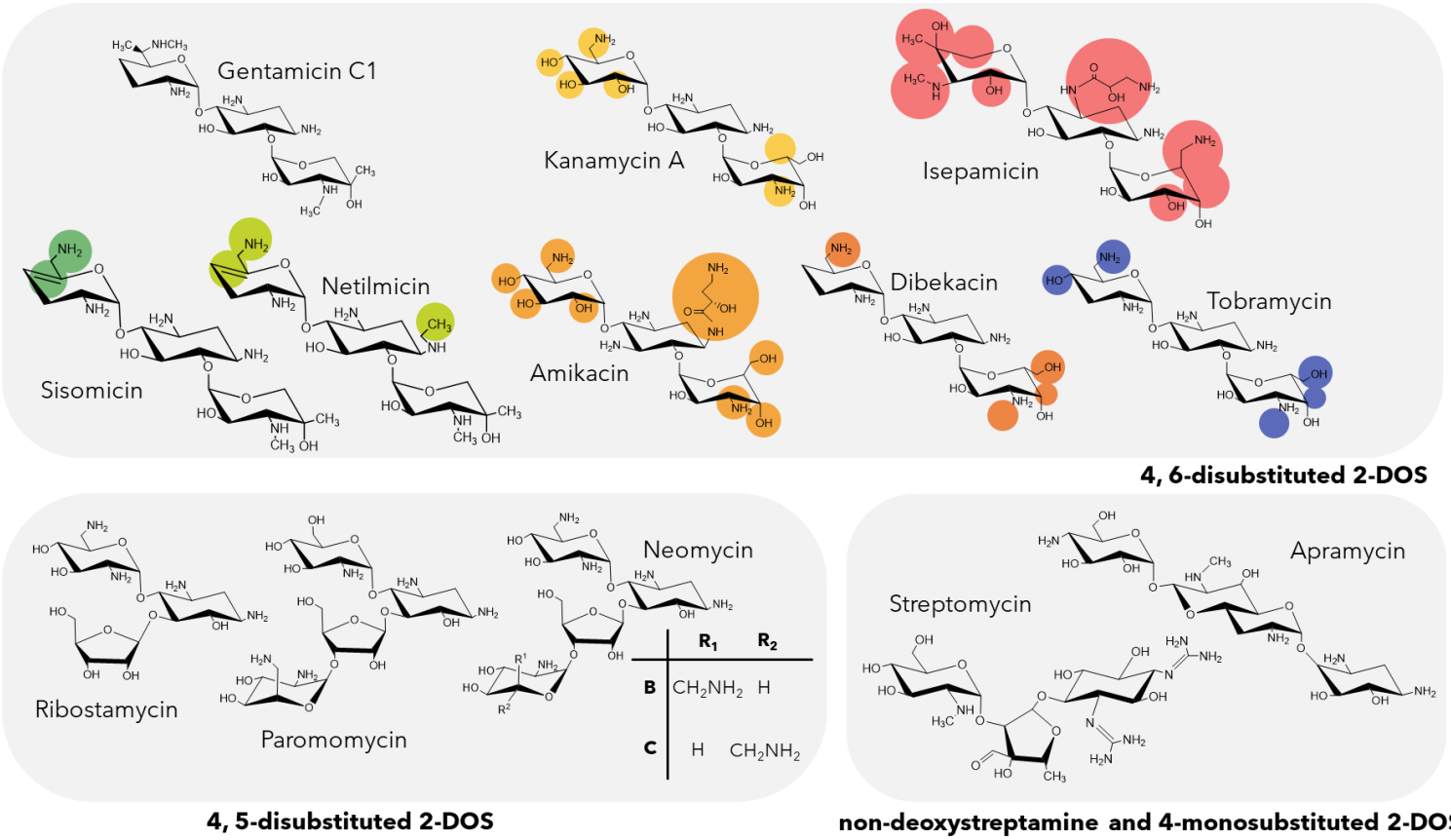


Figure 1.5: Example aminoglycosides. These fall into four main categories: 4,6-disubstituted 2-deoxystreptamine, 4,5-disubstituted 2-deoxystreptamine, 4-monosubstituted 2-deoxystreptamine, non-deoxystreptamine. Of the 4,6-disubstituted 2-deoxystreptamines, where they differ from gentamicin C1 has been highlighted. Kanamycin-derived species are in yellow and orange, sisomicin-derived species are in green.[172,178]

Most aminoglycosides have a bactericidal effect on susceptible microbes, binding to the A-site in the 16S rRNA **(Figure 1.6)**. Upon binding, the conformational change in the A site can allow for the misreading of an mRNA codon.[183,184] The production and accumulation of incorrect polypeptides has been shown to alter membrane permeability to allow for an increase in intracellular aminoglycoside concentrations, further damaging the cell.[185] Some aminoglycosides also act by halting the movement of peptidyl-tRNA from the A to the P site, effectively immobilizing it and stalling protein synthesis entirely.[180,186] Gentamicin, kanamycin and neomycin B also show a tertiary effect, binding to the 23S rRNA of the 50S subunit to affect the movement of the ribosomal subunits, further impacting on the efficiency of translation.[187–189] As a non-2-DOS-derived aminoglycoside, streptomycin has a different primary activity, as it binds to a different area of the 16S rRNA, inhibiting tRNA selection and also disrupting the 70S initiation complex.[190–192]

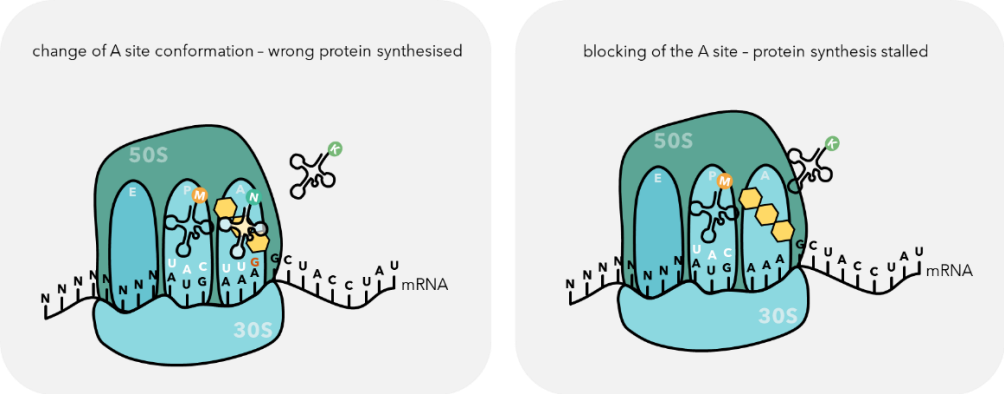


Figure 1.6: Main action of 2-DOS-derived aminoglycosides on the bacterial ribosome. Aminoglycoside molecule binds with high affinity to the A site of the 16S rRNA of the 30S ribosomal subunit. Though each aminoglycoside has specificity towards different regions of the A site, all cause a change in conformation. Misreading of the codon on delivery of aminoacyl-tRNA causes error-prone synthesis. The incorrect polypeptide chain formed is released and causes damage to the cell, particularly to the cell membrane. This can lead to an increase in membrane permeability and increased intracellular aminoglycoside concentrations. Some aminoglycosides also show an inhibition of ribosomal translocation i.e. by blocking the movement of aminoacyl-tRNA from A to the P site.[172,180–184]

Adaptive and acquired resistance to aminoglycosides can occur through mutation and modification to the ribosome, particularly in mutations to the *rrs* gene (coding for the 16S rRNA).[193] Though most mutations in this gene are non-viable, a viable mutant A1408G (and the equivalents) have been found in clinical isolates of resistant *M. tuberculosis*.[194] Resistance via enzymatic modification of the aminoglycoside binding site (through methylation of a single nucleotide) may also be carried out by 16S ribosomal RNA methyltransferases.[193] The most common mechanism of resistance to aminoglycosides is chemical modification of the aminoglycoside scaffold by aminoglycoside-modifying enzymes (AMEs), which can be categorised into N-acetyltransferases, O-nucleotidyltransferases, and O-phosphotransferases dependent on the position of the scaffold which is altered.[195] This modification of the aminoglycoside scaffold inhibits binding to the 16S rRNA and consequently the bacteria is able to avoid cell death.

Aminoglycosides are well known for their toxic side effects: specifically, nephro- and ototoxicity **(Figure 1.7)**. Ototoxicity was one of the first side effects described, three years after the discovery of streptomycin.[196] The nephrotoxicity of aminoglycosides can be attributed to approximately 5% of the administered dose accumulating in the epithelial cells of the proximal tubule of the kidney after glomerular filtration, causing oxidative stress to cells and leading to inflammation and apoptosis.[197,198] Ototoxic damage (damage to the auditory or vestibular system) caused by aminoglycosides is the main irreversible side effect. Though members of this class of antibiotics are all capable of causing damage to both the cochlear and vestibular systems, each aminoglycoside will affect one of the two more strongly – with cochlear damage causing hearing loss, and vestibular system damage able to cause unsteadiness, dizziness, ataxia, and nystagmus.[199,200] In both cases, damage is caused to the lysosomes of the proximal tubule epithelial cells and outer hair cells for the kidney and ear respectively, causing phospholipidosis and rupturing of the lysosomes. Once aminoglycosides are released into the

cytosol, they bind to iron salts which cause the formation of reactive oxygen species (ROS). Consequent cytochrome C release and later caspase recruitment eventually leads to apoptosis and necrosis.[200] Concurrently, aminoglycosides have been shown to inhibit protein synthesis in the endoplasmic reticulum, as well as affecting the accuracy and post-translational folding of proteins.[201,202]

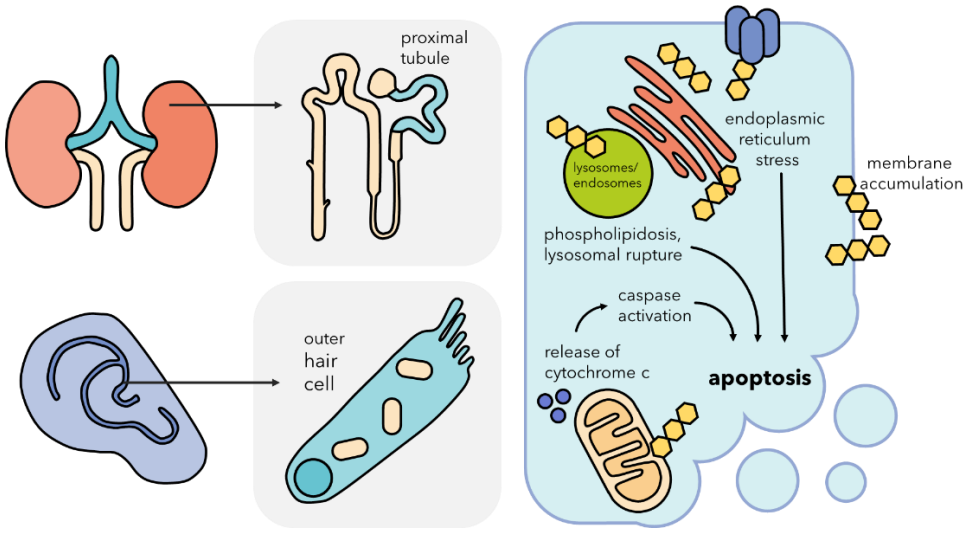


Figure 1.7: Action of aminoglycosides to cause nephro- and ototoxicity. Aminoglycosides are retained in the endosomes, lysosomes, and Golgi apparatus of both the epithelial cells lining the proximal tubules after glomerular filtration and the outer hair cells. Accumulation of aminoglycosides in these organelles leads to phospholipidosis of the lysosomes and endosomes, causing lysosomal rupture. Aminoglycosides have high affinity for iron salts, and chelate with these causing the formation of reactive oxygen species (ROS). Massive increase in ROS causes cytochrome C release and caspase activation, alongside concurrent endoplasmic reticulum stress, leading to apoptosis and cell death.[200–204]

Sensitivity to the ototoxic side effects of aminoglycosides is inheritable. Some mutations in the mitochondrial DNA (in particular, m.1555A>G, m.1494C>T) in the 12S rRNA gene give rise to ribosomes which have an aminoglycoside binding pocket resembling that of a bacterial ribosome. A 2010 study described the comparison of equivalent positions in the human 12S rRNA and the *E. coli* 16S rRNA – the secondary structure of the human 12S rRNA became much more similar to the bacterial 16S rRNA upon mutation of position 1555 from A to G, and therefore it was suggested to convey greater sensitivity to

aminoglycosides.[205] Despite this known toxicity, aminoglycosides remain especially relevant for newborns with suspected sepsis and for the treatment of multi-drug resistant tuberculosis.[206–208]

To combat bacterial resistance and the toxic side effects of aminoglycosides, many of the members of this class have been modified in some way from an original scaffold. Most engineered aminoglycosides have been generated semi-synthetically. Efforts have been made for 50 years towards the semi-synthesis of aminoglycosides, since the rational design of dibekacin from the kanamycin scaffold –[209] one 2018 review by Thamban Chandrika and Garneau-Tsodikova compiled all available methods to chemically modify aminoglycosides.[210] Historically, most semi-synthetic aminoglycosides of note (dibekacin, amikacin, netilmicin, isepamicin) have been derived from 4,6-disubstituted 2-deoxystreptamines kanamycin, sisomicin, or gentamicin.[209,211–213] Currently, it is generally thought that the engineering of aminoglycosides is more effective on those belonging to the 4,5-disubstituted 2-deoxystreptamine subclass, as these are not affected by modifications to the aminoglycoside binding pocket by the action of ribosomal methyltransferases.[214] One of the more recent attempts towards engineering a 4,6-disubstituted 2-deoxystreptamine aminoglycoside was the production of N1MS from sisomicin. One of nine compounds designed during the study, N1MS showed an ototoxicity 17 times lower than sisomicin – however, it did result in a lower antimicrobial activity.[215] In 2019, a modification of paromomycin was generated by replacing of the UDP-glucosamine-derived-ring's 4'-hydroxyl group for an n-propyl group (propylamicin). This also showed reduced ototoxcity (and better selectivity for bacterial than human mitochondrial and cytosolic ribosomes than other therapeutically relevant aminoglycosides), but with no loss of antimicrobial activity in comparison to the parent compound when tested against *E. coli* in mouse thigh and septicaemia models. A different approach to making new aminoglycosides has been to generate hybrid antimicrobials – as many of the aminoglycosides are derived from the same 2-DOS precursor,

combining enzymes to make new scaffolds has been recently studied. The patent for genkamicin, a hybrid antibiotic using enzymes from the gentamicin and kanamycin pathway, was filed in 2015.[216]

1.3.1. Gentamicin

Gentamicin is an aminoglycoside antibiotic which was first identified as a product from *Micromonospora echinospora* DSM 43816 (then *Micromonospora purpurea*) in 1963 by Weinstein and colleagues.[177] The structure of gentamicin was first described by researchers at pharmaceutical company Schering, and is a pseudo-trisaccharide comprising of a 2-deoxysteptamine ring flanked by a purpurosamine ring at the C4 position and a garosamine ring at the C6 position **(Figure 1.8)**.

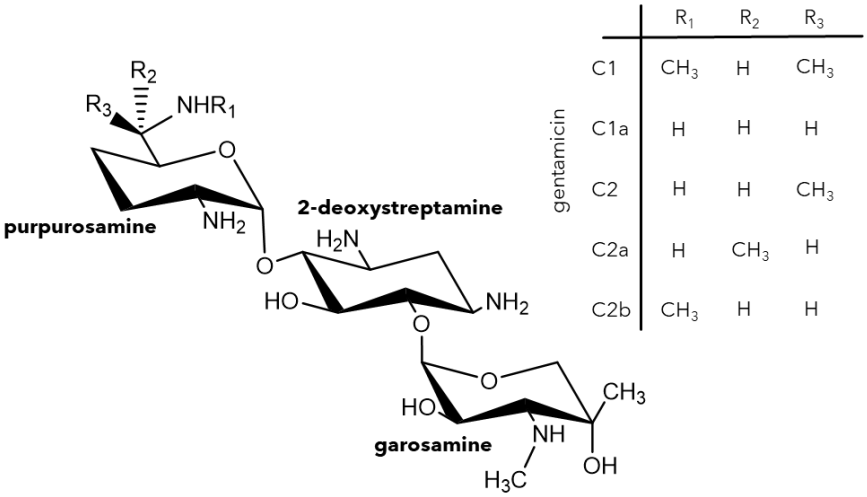


Figure 1.8: Structure of gentamicin C congeners. By variance at the 5′ carbon of the purpurosamine ring, five different congeners with varying antimicrobial bioactivity and cytotoxicity are generated.

Gentamicin's mode of action is similar to the other aminoglycosides, as it binds to the bacterial 16S rRNA at the A site of the 30S subunit to inhibit protein synthesis **(Figure 1.9)**.[217] Gentamicin is one of the aminoglycosides (along with paromomycin) which acts by causing significant miscoding of nascent polypeptides, as well as inhibiting translocation of peptidyl-tRNAs from the A

to the P site of the ribosome.[180] Gentamicin not able to be generated through total chemical synthesis, as **i)** a large library of compounds is required to be generated **ii)** it remains challenging to ensure that the multiple saccharide rings are orientated properly when protecting the compound **iii)** all pathway steps are inefficient and have low product yields, causing the starting material to be quickly depleted.[218] Production of this antibiotic at industrial scale is instead carried out through culturing *M. echinospora*.[219]

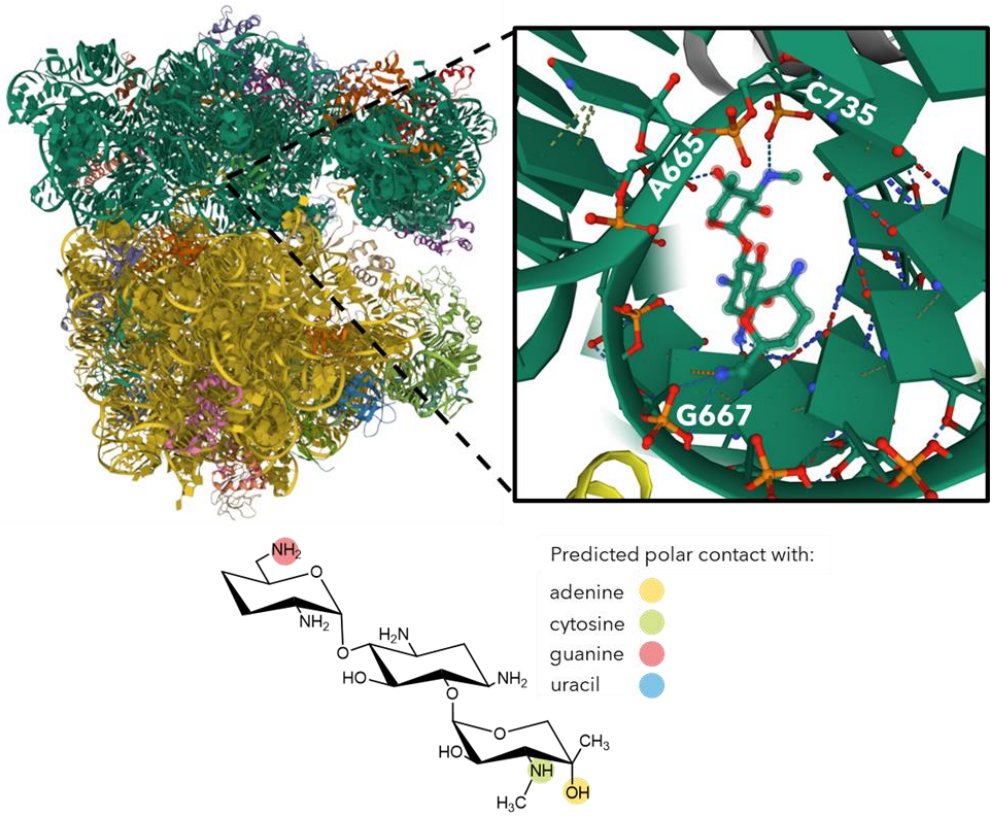


Figure 1.9: Three-dimensional structure of the *E. coli* ribosome in complex with gentamicin C1a (PDB ID 4V53) visualised by Mol* Viewer.[187,220,221] Nucleotides predicted to interact with the gentamicin C1a structure are highlighted on both the ribosome structure and the gentamicin C1a structure.

Five congeners of gentamicin C accounts for approximately 80% of the complex mixture of commercial gentamicin: gentamicin C1 (25-50%), C1a (10-35%), C2 and C2a (25-55% combined) – all major components – and C2b as a minor component.[222–224] The other ~20% of gentamicin is composed of by-products and pathway intermediates, including gentamicin A, B, X, and others. Generally, these congeners have been shown to have lower antibiotic

activity than the gentamicin C congeners.[224] It is difficult to determine which gentamicin C congener has the overall stronger antibiotic activity: for example, against *S. pyogenes*, gentamicin C1a showed a two-fold higher antibiotic activity than the other congeners, whilst against *Bacillus subtilis*, all congeners showed similar minimum inhibitory concentrations (MICs).[225] Importantly, as gentamicin is derived from microbial cultures, the ratio of these congeners in any one batch can vary massively.[226]

Over the past 14 years, the biosynthetic pathway of gentamicin has been characterised **(Figure 1.10)**. Glucose-6-phosphate is converted through the action of six enzymes in seven steps to gentamicin A2, which is the first pseudo-trisaccharide of the gentamicin pathway.[227,228] Once the common precursor 2-DOS is generated via the first four enzymes of the pathway, glycosyltransferases add N-acetyl-d-glucosamine and xylose to form the classical three-sugar gentamicin structure. From this point, modification yields gentamicin A and gentamicin X2.

The gentamicin pathway branches after gentamicin X2.[233] Conversion to G418 leads to the production of gentamicin C components gentamicin C1, gentamicin C2, and gentamicin C2a, while the other major branch through JI-20A leads to production of gentamicin C1a and gentamicin C2b.[232,234–239] The steps between gentamicin A2 and G418 have been elucidated more recently **(Figure 1.11)** – as the methyltransferases GenN, GenD1 and GenK can each methylate many of the pathway intermediates between these two precursors, it gives rise to a much higher variety of gentamicin by-products than was previously known. The major route to G418 still takes place through production of gentamicin A by the action of GenN, the transformation to gentamicin X2 by GenD1, and finally G418 production by GenK, though all routes to the biosynthesis of G418 are viable.[230]

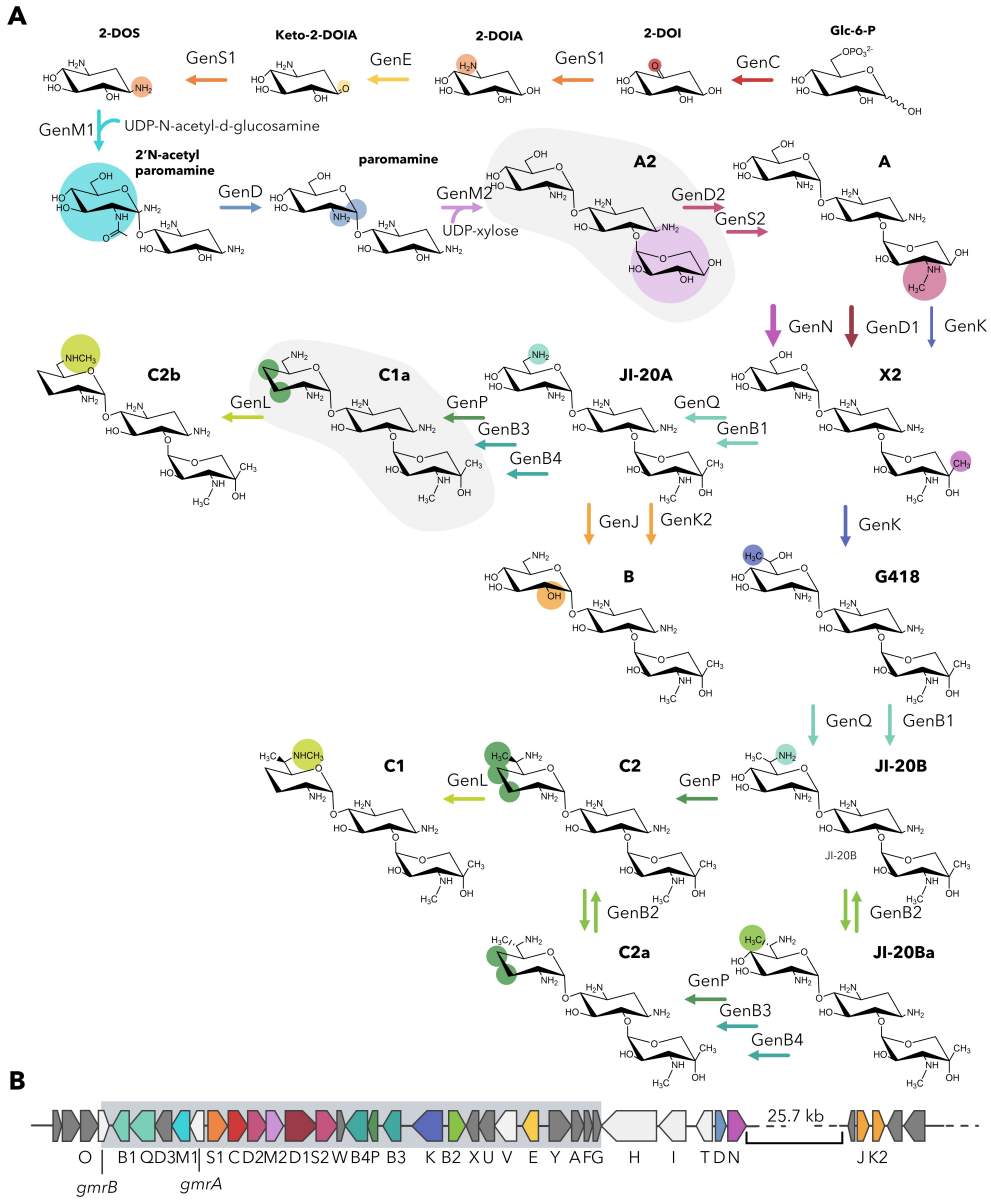


Figure 1.10: Entire known biosynthesis pathway for production of gentamicin as of 2021. Enzyme names are derived from the scheme for *M. echinospora* DSM 43036 GenBank Accession No.: AJ628149. Figure adapted from descriptions in Park et al. (2008),[227] Huang et al. (2015),[229], Li et al. (2018),[230], Yu et al. (2017),[231] Li et al. (2021),[232] Reva (2017).[233] **A** – Entire known biosynthesis pathway. **B** – Entire known gentamicin cluster map. Gene names are included minus the '*gen*' prefix where present, to maintain clarity.

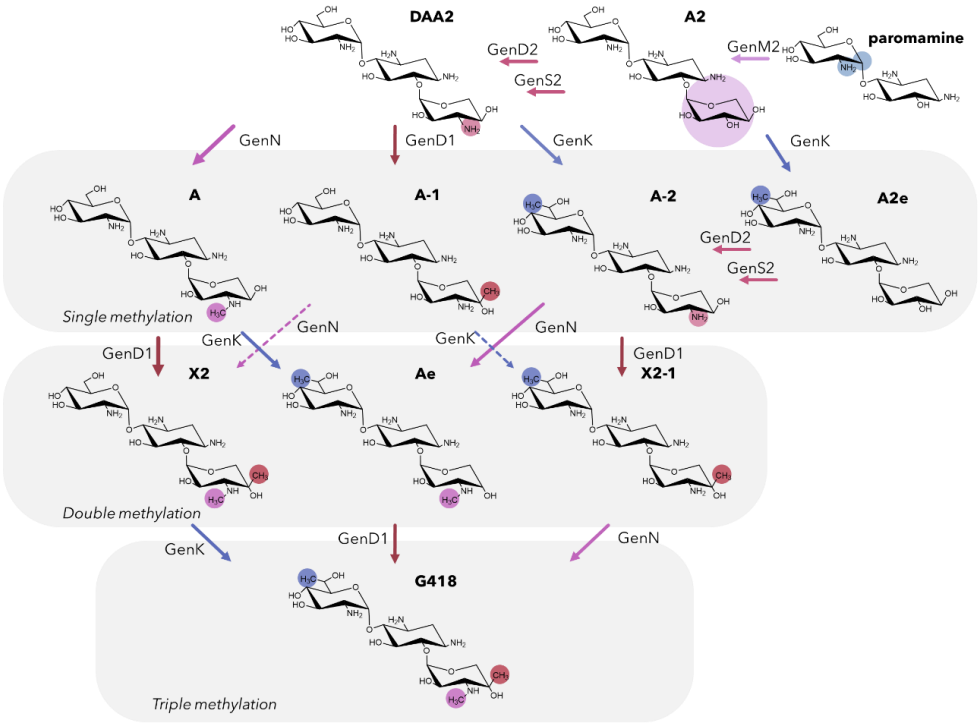


Figure 1.11: Methylation of gentamicin pathway intermediates between gentamicin A2 and G418. Major and minor pathways are highlighted with thicker and thinner arrows respectively; dashed lines show reactions with little experimental evidence. GenN is highlighted with light purple, GenK with dark purple, and GenD1 with dark red; each modification caused by each reaction is highlighted on the product with a circle. Figure adapted from Figure 1 of Li et al.[230]

The full reaction to gentamicin C1 was only recently described within the same work, with the discovery of *genL*, a methyltransferase located 2.54 Mbp from the previously identified gentamicin gene cluster in the genome of *M. echinospora* ATCC15835.[230] Though gentamicin B had previously been detected as a minor component of gentamicin mixtures assayed in the past, the natural biosynthetic route to this particular form of gentamicin was first described in 2019, through both *in vitro* and *in vivo* research.[240,241] However, Ni et al. established production of gentamicin B from JI-20A through introducing kanamycin biosynthesis genes *kanJ* and *kanK*, and it was these which were used to probe for the missing enzymes required for gentamicin B synthesis *in vivo*.[241,242]

There are also several genes within the gentamicin gene cluster which have been attributed non-biosynthetic functions. Two resistance genes are present, *gmrA* and *gmrB*, which are proposed to be 16S rRNA methyltransferases. *gmrA* has been used as a resistance gene in multiple works towards elucidating the steps of the gentamicin biosynthesis pathway, and therefore the function appears to be correctly assigned.[230,232] Limited evidence towards *gmrB* being necessary exists; a study by Wan et al (2018) where *gmrB* was knocked out in *Micromonospora purpurea* G1008 led to no discernible difference in either resistance to gentamicin or congener production.[243] The cluster also contains several candidate transporters. GenV was first identified by Unwin et al (2004) who proposed its function as a transporter based on sequence homology.[244] Park et al (2008) was the first to attempt rebuilding of the gentamicin pathway from scratch, and in this work *genV* is included as a resistance gene for the production of gentamicin precursors.[227] In their work, a plasmid was constructed containing both proposed resistance genes (*gmrA* and *gmrB*) along with *genV*, which together increased gentamicin resistance of host strain *Streptomyces venezuelae* from 1 μg/mL to 100 μg/mL gentamicin. These three genes were only analysed as one group, and so it is difficult to clarify the extent to which the presence of *genV* is responsible for this increased strain resistance. To the best of our knowledge, there has been no specific experimental evidence towards the function of GenV as an efflux pump; however, as efflux pumps are a common resistance mechanism within antibiotic gene clusters,[245] it remains likely that there is one within the gentamicin gene cluster, and *genV* remains the best candidate of those genes without attributed biosynthetic function. Unwin et al. (2004) and Piepersberg et al. (2007) also suggest that *genH* and *genI* could encode transporters, however it does not appear that any experimental characterisation of these has been published as of the writing of this work.[244,246]

Gentamicin is used as a therapeutic in rare cases, due to its potential for severe nephrotoxic and ototoxic side effects, similar to other antibiotics in its

class.[201,247] Generally, gentamicin will be utilised for treatment of serious bacterial infections where antimicrobial resistance may cause first-line treatments to fail: in the USA, the US Food and Drug Administration (FDA) has approved the use of gentamicin to treat bacterial meningitis, bacterial septicaemia, and infective endocarditis.[248]

The toxicity of gentamicin congeners has been widely studied. Kohlhepp et al. (1984) found that the gentamicin C2 congener is the most nephrotoxic;[249] Sandoval et al. (2006) were able to further separate the C2 and C2a diastereomers from each other and show that gentamicin C2 exhibits very low amounts of nephrotoxicity in both proximal tubule cell culture and a rat model. Their work suggests that the nephrotoxicity of commercial gentamicin is due to the presence of the gentamicin C2a congener.[250]

With regards to ototoxicity, gentamicin shows primarily vestibulotoxic behaviour, affecting the vestibule and causing unsteadiness and dizziness as its main side effects.[200] However, it can also damage the cochlea, though less so than other aminoglycosides.[200] In clinical studies, patients do not generally realise any hearing loss from gentamicin-caused cochlear injury – it is much easier to be aware of vestibular damage which conveys symptoms such as dizziness.[251] In 2008, Kobayashi et al. investigated the cochleotoxicity of gentamicin C1, C2, and C1a, and found C1a and C2 to be the least and most damaging to the cochlear outer hair cells of the rats tested, respectively. This study did not separate C2 and C2a and so further testing may be required.[252] A 2019 study by Ishikawa et al. again suggested that gentamicin C1a had lower cochleotoxicity than gentamicin as a mixture in both *in vitro* and *in vivo* models, and thus if this could be separated cleanly from its congeners, it could be more therapeutically useful.[253] A recent report by O'Sullivan et al. in 2020 suggests that gentamicin C2b shows a lower ototoxicity than other gentamicin C congeners while retaining equivalent antimicrobial activity.[254]

The revitalisation of gentamicin is of importance due to its known nephro- and ototoxicity, and therefore analogues have been generated with an attempt to reduce this. Patent WO 2011/143497 A1 describes a diverse modification of side chains leading to generation of several gentamicin analogues. The most promising compound showed no nephrotoxicity in rats after 14 days; however, residual cytotoxicity to HK-2 cells was observed which may still be problematic.[255] As the biosynthetic pathway towards gentamicin is now likely close-to fully, or fully known, generation of gentamicin analogues is likely to become much easier. By generating a scaffold which retains antimicrobial activity but with reduced or no nephrotoxic and ototoxic side effects, gentamicin could be rejuvenated for use clinically.

1.4. Riboswitches as tools for synthetic biology

Riboswitches are a non-coding region of RNA which function to grant either transcriptional or translational control over gene expression through inducible secondary structure formation. They mostly function independently, without assistive proteins.[256] Riboswitches contain two main domains. The aptamer is responsible for sensing of a ligand, whilst the expression region enables access to the ribosome binding site in translationally-controlling riboswitches, or forms an anti-terminator sequence in transcriptionally-controlling riboswitches.[256,257]

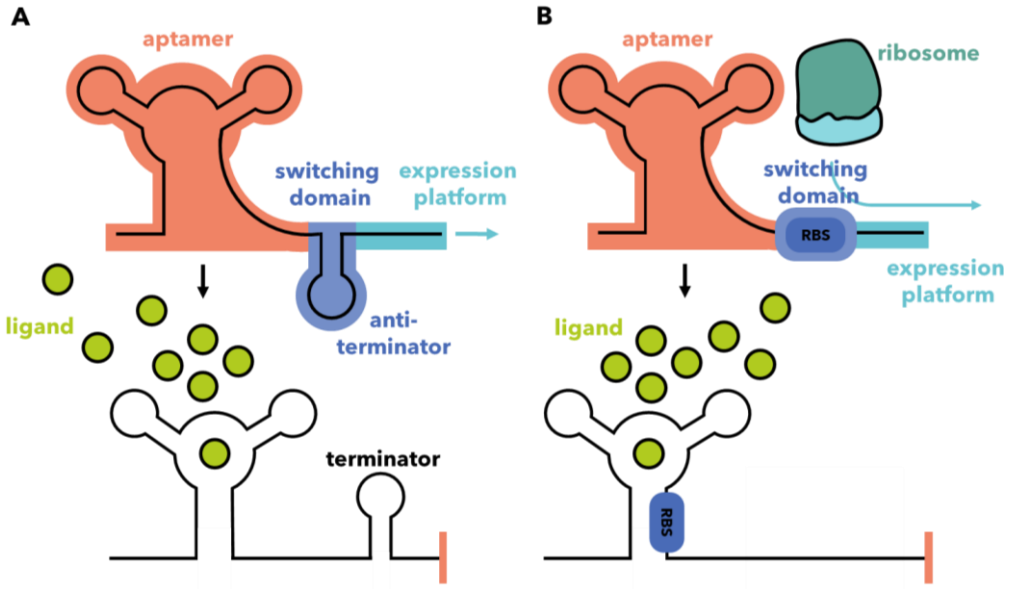


Figure 1.12: Regulation of gene expression by riboswitches. A – transcriptionally controlled riboswitches function by formation of an anti-terminator when ligand is not available. When ligand is introduced, terminator region forms which inhibits binding of RNA polymerase and therefore gene expression is not possible. **B** – translationally controlled riboswitches function by a conformational change causing sequestration of ribosome binding site upon binding of ligand. The ribosome can no longer bind and therefore gene expression is not possible.[256]

The definition of riboswitches has expanded slightly since their inception, to include temperature-sensitive and metal-ion-binding regions.[258–261] Classically, riboswitches are defined by their cognate ligand, which tend to be RNA-derived compounds.[262] As of 2017, 38 different classes of riboswitches had been described.[262] The most abundant riboswitches in nature are thiamine pyrophosphate and adenosylcobalamin or coenzyme B_{12} riboswitches.[262] It has been predicted that many more classes of riboswitch are undiscovered, but that these may represent rare classes not widespread in eubacteria.[263]

Currently, expanding the number of riboswitches available in model or industrially relevant strains is a main focus of the field. The theophylline-dependent riboswitch has been shown to function in diverse bacterial strains, including model actinomycete *S. coelicolor*,[264] *B. subtilis*,[265] pathogenic microorganisms *S. pyogenes*, *Acinetobacter baylyi*, and *Acinetobacter baumannii*,[265] *E. coli*,[265] cyanobacteria,[266] and mycobacteria,[265,267] Born et al. have recently

described expanding the use of the theophylline riboswitch to archaeon *Haloferax volcanii*; eubacterial riboswitches are rarely used in this domain. This represented a leap forward in engineering gene circuits in archaebacteria, becoming of more interest as industrial bioproduction hosts due to their salt-tolerance.[268] As already described in section **1.2.3.4.**, the theophylline riboswitch has been used multiple times for attenuating the toxic effects of Cas9 expression when using CRISPR/Cas9 systems in *S. coelicolor*. When used alone, or in combination with other elements of genetic control such as an inducible promoter, the toxicity of Cas9 overexpression was reduced significantly.[161,162,164]

The main advances in riboswitch design have perhaps come with the improvements to SELEX (Systematic Evolution of Ligands by Exponential Enrichment) technology. Over the past three decades, SELEX has allowed for the design of riboswitches specific to a target ligand. This has enabled the creation of a palette of biosensors regulated by molecules such as flavonoid naringenin,[269] nylon monomer caprolactam,[270] tetracycline,[271] ciprofloxacin,[272] and aminoglycosides paromomycin and neomycin.[273,274] These biosensors can be used to monitor production of these in culture set-ups, reducing the time and expense required to run analytical instruments. Increasing the diversity of riboswitches is advantageous, especially due to the fairly low variety of known cognate ligands found in nature. The advances of SELEX have been recently described in a review by Zhu et al.[275] However, while SELEX is capable of generating pools of novel aptamers, the translation of these aptamers to functional riboswitches can be challenging. Historically, SELEX has focused on generating sequences with high affinity to target compounds, but there have been difficulties selecting for the required rapid conformational change to switch expression on and off. Kaiser et al. described an approach for the widening of the aptamer selection to include aptamers which can associate and dissociate with the RNA sequence of interest, and not simply irreversibly bind.[273] Riboswitch-based biosensors remain behind transcription-factor-based

biosensors due to the currently limited number of molecules which can be utilised, but further advances in SELEX may begin to yield more functional riboswitches for use in synthetic biology applications.[276]

1.5. Project Aims

The revitalisation of our current antibiotic portfolio remains a strategy which is under-utilized, but this will complement the generation of new technologies as we move towards the predicted post-antibiotic era. This thesis concentrates on the expression of a gene cluster from novel *Micromonospora* species *Micromonospora* sp. DEM32671. *M.* sp. DEM32671 was identified in the Demuris actinomycete strain library and was genome sequenced prior to the beginning of this work. A predicted aminoglycoside was of interest for production in a heterologous host due to long culture times required for *M.* sp. DEM32671. This gene cluster was shown through *in silico* analysis to have highest similarity to the gentamicin gene cluster from *M. echinospora* DSM 43816, the industrial gentamicin producer.

The primary aim of the PhD is to achieve robust and reproducible production of aminoglycoside by *S. coelicolor* carrying Cluster 24, to aid in further characterisation of the product compound(s). Multiple *Streptomyces* host strains will be tested, and a cohort of gentamicin-resistant 'superhost' strains will be generated for improvement of production. These 'superhosts' will be characterised through both susceptibility tests against other aminoglycosides and whole-genome sequencing. The antibiotic activity of the strains will analysed and LC-MS used towards characterisation of the output of Cluster 24.

The secondary aim is to carry out edits to Cluster 24 for the production of specific aminoglycoside congeners. Gentamicin is well-known for its toxicity, and therefore the selective and specific production of non-toxic gentamicin congeners is of value if it is to be revitalised as a therapeutic. CRISPR-Cas9 will be used for the knock-out of key genes within the cluster towards production of gentamicin A2 and gentamicin C1a. Additionally, a total rebuilding and refactoring strategy towards construction of the minimal gene set towards gentamicin A2 and gentamicin C1a will be carried out. Towards further

improving upon yield of antibiotic, CRISPR-Cas9 will again be used for the introduction of a strong promoter cassette within the gene cluster.

Finally, the tertiary aim is to investigate whether any cluster regulation is present, and whether the alteration of this could have a positive impact on aminoglycoside production titres. After identifying queuosine biosynthesis genes within Cluster 24, the effect of these on production will be investigated through deletion and subsequent antibiotic bioassays. The presence of a putative queuosine-precursor-linked riboswitch sequence within the cluster will also be investigated. In addition, alternative candidate riboswitches of the same class will be tested to establish whether these are functional in *S. coelicolor and* could be used in creation of new gene circuits tailored for this species. It is hoped that the work detailed in this thesis will represent a pipeline towards heterologous aminoglycoside production, beginning from first introduction into heterologous host through to cluster editing towards improvement of product titre.

1.6. References

1. M. A. Kohanski, D. J. Dwyer, and J. J. Collins, "How antibiotics kill bacteria: from targets to networks.," *Nature Reviews Microbiology*, vol. 8, no. 6, pp. 423–35, 2010, doi: 10.1038/nrmicro2333.

2. M. Sebastián *et al.*, "Discovery of antimicrobial compounds targeting bacterial type FAD synthetases," *Journal of Enzyme Inhibition and Medicinal Chemistry*, vol. 33, no. 1, pp. 241–254, 2018, doi: 10.1080/14756366.2017.1411910.

3. C. Fu *et al.*, "The natural product carolacton inhibits folate-dependent C1 metabolism by targeting FolD/MTHFD," *Nature Communications*, vol. 8, no. 1, p. 1529, 2017, doi: 10.1038/s41467-017-01671-5.

4. P. W. Smith, K. Watkins, and A. Hewlett, "Infection control through the ages," *American Journal of Infection Control*, vol. 40, pp. 35–42, 2012, doi: 10.1016/j.ajic.2011.02.019.

5. J. Houbraken, J. C. Frisvad, and R. A. Samson, "Fleming's penicillin producing strain is not *Penicillium chrysogenum* but *P. rubens*.," *IMA Fungus*, vol. 2, no. 1, pp. 87–95, 2011, doi: 10.5598/imafungus.2011.02.01.12.

6. S. A. Waksman and H. B. Woodruff, "The Soil as a Source of Microorganisms Antagonistic to Disease-Producing Bacteria.," *Journal of Bacteriology*, vol. 40, no. 4, pp. 581–600, 1940.

7. L. Valiquette and K. B. Laupland, "Digging for new solutions," *Canadian Journal of Infectious Diseases and Medical Microbiology*, vol. 26, no. 6. pp. 289–290, 2015. doi: 10.1155/2015/971858.

8. W. Kingston, "Streptomycin, Schatz v. Waksman, and the Balance of Credit for Discovery." *Journal of the History of Medicine and Allied Sciences*, vol. 59, no. 3, pp. 441-62, 2004, doi: 10.1093/jhmas/jrh091.

9. W. H. Harrison, "Ototoxicity of dihydrostreptomycin.," *Quarterly Bulletin Northwest University Medical School*, vol. 28, no. 3, pp. 271–3, 1954.

10. R. L. Peck, C. E. Hoffhixe, and K. Folker, "*Streptomyces* antibiotics. IX. Dihydrostreptomycin," *Journal of the American Chemical Society*, vol. 68, no. 7, pp. 1390–1391, 1946, doi: 10.1021/ja01211a513.

11. C. A. Townsend, "Convergent biosynthetic pathways to β-lactam antibiotics," *Current Opinion in Chemical Biology*, vol. 35. Elsevier Ltd, pp. 97–108, 2016. doi: 10.1016/j.cbpa.2016.09.013.

12. J. Ehrlich, Q. R. Bartz, R. M. Smith, D. A. Joslyn, and P. R. Burkholder, "Chloromycetin, a New Antibiotic From a Soil Actinomycete," *Science*, vol. 106, no. 2757, pp. 417–417, 1947, doi: 10.1126/science.106.2757.417.

13. H. E. Carter, D. Gottlieb, and H. W. Anderson, "Chloromycetin and streptothricin," *Science*, vol. 107, no. 2770. p. 113, 1948. doi: 10.1126/science.107.2770.113-b.

14. J. Controulis, M. C. Rebstock, and H. M. Crooks, "Chloramphenicol (Chloromycetin). V. Synthesis," *Journal of the American Chemical Society*, vol. 71, no. 7, pp. 2463–2468, 1949, doi: 10.1021/ja01175a066.

15. M. I. El-Gamal, I. Brahim, N. Hisham, R. Aladdin, H. Mohammed, and A. Bahaaeldin, "Recent updates of carbapenem antibiotics," *European Journal of Medicinal Chemistry*, vol. 131, pp. 185–195, 2017. doi: 10.1016/j.ejmech.2017.03.022.

16. S. W. Tsao, B. A. M. Rudd, X. G. He, C. J. Chang, and H. G. Floss, "Identification of a red pigment from *Streptomyces coelicolor* A3(2) as a mixture of prodigiosin derivatives," *Journal of Antibiotics*. 1985. doi: 10.7164/antibiotics.38.128.

17. L. F. Wright and D. A. Hopwood, "Actinorhodin is a chromosomally determined antibiotic in *Streptomyces coelicolor* A3(2)," *Journal of General Microbiology*, 1976, doi: 10.1099/00221287-96-2-289.

18. P. P. Chong *et al.*, "Physical identification of a chromosomal locus encoding biosynthetic genes for the lipopeptide calcium-dependent antibiotic (CDA) of *Streptomyces coelicolor* A3(2)," *Microbiology (N Y)*, 1998, doi: 10.1099/00221287-144-1-193.

19. R. Kirby and D. A. Hopwood, "Genetic determination of methylenomycin synthesis by the SCP1 plasmid of *Streptomyces coelicolor* A3(2)," *Journal of General Microbiology*, 1977, doi: 10.1099/00221287-98-1-239.

20. S. D. Bentley *et al.*, "Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2)," *Nature*, vol 417, pp. 141-147, 2002, doi: 10.1038/417141a.

21. P. D. Schloss and J. Handelsman, “Status of the microbial census.,” *Microbiology and Molecular Biology Reviews*, vol. 68, no. 4, pp. 686–91, 2004, doi: 10.1128/MMBR.68.4.686-691.2004.

22. M. S. Rappé and S. J. Giovannoni, “The Uncultured Microbial Majority,” *Annual Review of Microbiology*, vol. 57, no. 1, pp. 369–394, 2003, doi: 10.1146/annurev.micro.57.030502.090759.

23. F. Hemmerling and J. Piel, “Strategies to access biosynthetic novelty in bacterial genomes for drug discovery,” *Nature Reviews Drug Discovery*, vol. 21, no. 5, pp. 359–378, 2022, doi: 10.1038/s41573-022-00414-6.

24. World Health Organization, “2020 Antibacterial agents in clinical and preclinical development: an overview and analysis,” Geneva, 2021. Accessed: May 03, 2022. [Online]. Available: https://www.who.int/publications-detail-redirect/9789240021303

25. E. J. Culp *et al.*, “Evolution-guided discovery of antibiotics that inhibit peptidoglycan remodelling,” *Nature*, vol. 578, no. 7796, pp. 582–587, 2020, doi: 10.1038/s41586-020-1990-9.

26. J. M. Stokes *et al.*, “A Deep Learning Approach to Antibiotic Discovery,” *Cell*, vol. 180, no. 4, pp. 688-702.e13, 2020, doi: 10.1016/j.cell.2020.01.021.

27. E. Abraham and E. Chain, “An enzyme from bacteria able to destroy penicillin,” *Reviews of Infectious Diseases*, vol. 10, no. 4, pp. 677–678, 1988.

28. C. H. Rammelkamp and T. Maxon, “Resistance of *Staphylococcus aureus* to the Action of Penicillin.,” *Experimental Biology and Medicine*, vol. 51, no. 3, 1942, doi: 10.3181/00379727-51-13986.

29. M. P. Jevons, “Celbenin” - resistant Staphylococci,” *BMJ*, vol. 1, no. 5219, pp. 124–125, 1961, doi: 10.1136/bmj.1.5219.124-a.

30. C. P. Harkins *et al.*, “Methicillin-resistant *Staphylococcus aureus* emerged long before the introduction of methicillin into clinical practice,” *Genome Biology*, vol. 18, no. 1, p. 130, 2017, doi: 10.1186/s13059-017-1252-9.

31. J. Larsen *et al.*, “Emergence of methicillin resistance predates the clinical use of antibiotics,” *Nature*, vol. 602, no. 7895, pp. 135–141, 2022, doi: 10.1038/s41586-021-04265-w.

32. WHO. Media Centre. News Release., "WHO publishes list of bacteria for which new antibiotics are urgently needed.," Feb. 27, 2017. http://www.who.int/mediacentre/news/releases/2017/bacteria-antibiotics-needed/en/ (accessed May 09, 2022).

33. G. Meletis, "Carbapenem resistance: overview of the problem and future perspectives.," *Therapeutic Advances in Infectious Diseases*, vol. 3, no. 1, pp. 15–21, 2016, doi: 10.1177/2049936115621709.

34. World Health Organization, "Antibacterial agents in clinical development: an analysis of the antibacterial clinical development pipeline, including tuberculosis.," *Who/Emp/Iau/2017.12*, p. 48, 2017.

35. B. B. Xavier *et al.*, "Identification of a novel plasmid-mediated colistin-resistance gene, *mcr-2*, in *Escherichia coli*, Belgium, June 2016," *Eurosurveillance*, vol. 21, no. 27, p. 30280, 2016, doi: 10.2807/1560-7917.ES.2016.21.27.30280.

36. Y.-Y. Liu *et al.*, "Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study," *The Lancet Infectious Diseases*, vol. 16, no. 2, pp. 161–168, 2016, doi: 10.1016/S1473-3099(15)00424-7.

37. W. Yin *et al.*, "Novel Plasmid-Mediated Colistin Resistance Gene *mcr-3* in *Escherichia coli*," *mBio*, vol. 8, no. 3, pp. e00543-17, 2017, doi: 10.1128/mBio.00543-17.

38. S. Biswas, J. M. Brunel, J. C. Dubus, M. Reynaud-Gaubert, and J. M. Rolain, "Colistin: An update on the antibiotic of the 21st century," *Expert Review of Anti-Infective Therapy*, vol. 10, no. 8. pp. 917–934, 2012. doi: 10.1586/eri.12.78.

39. R. McGehee RF, F. F. Barre, and M. Finland, "Resistance of *Staphylococcus aureus* to lincomycin, clinimycin, and erythromycin.," *Antimicrobial Agents and Chemotherapy*, vol. 8, pp. 392–397, 1968.

40. C. P. Miller and M. Bohnhoff, "Two Streptomycin-resistant Variants of *Meningococcus*.," *Journal of Bacteriology*, vol. 54, no. 4, pp. 467–81, 1947.

41. T. Frieden, "Antibiotic resistance threats in the United States," Centers for Disease Control and Prevention, 2013. doi: CS239559-B.

42. A. E. Clatworthy, E. Pierson, and D. T. Hung, "Targeting virulence: A new paradigm for antimicrobial therapy," *Nature Chemical Biology*, vol. 3, no. 9, pp. 541–548, 2007. doi: 10.1038/nchembio.2007.24.

43. L. Chen, R. Todd, J. Kiehlbauch, M. Walters, and A. Kallen, "*Notes from the Field* : Pan-Resistant New Delhi Metallo-Beta-Lactamase-Producing *Klebsiella pneumoniae* — Washoe County, Nevada, 2016," *MMWR. Morbidity and Mortality Weekly Report*, vol. 66, no. 1, p. 33, 2017, doi: 10.15585/mmwr.mm6601a7.

44. J. O'Neill, "The Review on Antimicrobial Resistance. Antimicrobial Resistance: Tackling a Crisis for the Health and Wealth of Nations. UK Government Report," 2014. Accessed: Jan. 03, 2018. [Online]. Available: https://amr-review.org/sites/default/files/AMR Review Paper - Tackling a crisis for the health and wealth of nations_1.pdf

45. M. E. A. de Kraker, A. J. Stewardson, and S. Harbarth, "Will 10 Million People Die a Year due to Antimicrobial Resistance by 2050?," *PLOS Medicine*, vol. 13, no. 11, p. e1002184, 2016, doi: 10.1371/journal.pmed.1002184.

46. CDC, "Antibiotic use in the United States, 2017: Progress and Opportunities," *US Department of Health and Human Services*, pp. 1–37, 2017.

47. A. Cassini *et al.*, "Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis," *The Lancet Infectious Diseases*, vol. 19, no. 1, pp. 56–66, 2019, doi: 10.1016/S1473-3099(18)30605-4.

48. K. Egan *et al.*, "Genome Sequence of *Geobacillus stearothermophilus* DSM 458, an Antimicrobial-Producing Thermophilic Bacterium, Isolated from a Sugar Beet Factory.," *Genome Announcements*, vol. 5, no. 43, 2017, doi: 10.1128/genomeA.01172-17.

49. M. Yin *et al.*, "The complete genome sequence of *Streptomyces albolongus* YIM 101047, the producer of novel bafilomycins and odoriferous sesquiterpenoids," *Journal of Biotechnology*, vol. 262, pp. 89–93, 2017, doi: 10.1016/j.jbiotec.2017.09.018.

50. M. M. Gerst, E. G. Dudley, L. Xiaoli, and A. E. Yousef, "Draft Genome Sequence of *Bacillus velezensis* GF610, a Producer of Potent Anti-Listeria

Agents.," *Genome Announcements*, vol. 5, no. 41, 2017, doi: 10.1128/genomeA.01046-17.

51. M. H. Medema and M. A. Fischbach, "Computational approaches to natural product discovery," *Nature Chemical Biology*, vol. 11, no. 9, pp. 639–648, 2015, doi: 10.1038/nchembio.1884.

52. H. Gross, "Strategies to unravel the function of orphan biosynthesis pathways: recent examples and future prospects," *Applied Microbiology and Biotechnology*, vol. 75, no. 2, pp. 267–277, May 2007, doi: 10.1007/s00253-007-0900-5.

53. Y.-M. Chiang, S.-L. Chang, B. R. Oakley, and C. C. C. Wang, "Recent advances in awakening silent biosynthetic gene clusters and linking orphan clusters to natural products in microorganisms.," *Current Opinion in Chemical Biology*, vol. 15, no. 1, pp. 137–43, 2011, doi: 10.1016/j.cbpa.2010.10.011.

54. K. A. Wetterstrand, "DNA Sequencing Costs: Data from the NHGRI Large-Scale Genome Sequencing Program," *www.genome.gov/Sequencingcostsdata*, 2017. https://www.genome.gov/sequencingcostsdata/ (accessed Jan. 07, 2018).

55. National Center for Biotechnology Information, "Microbial Genomes, National Center for Biotechnology Information (NCBI)," Nov. 17, 2021. https://www.ncbi.nlm.nih.gov/genome/microbes/ (accessed Nov. 17, 2021).

56. K. Blin *et al.*, "antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification," *Nucleic Acids Research*, vol. 45, 2017, doi: 10.1093/nar/gkx319.

57. K. Blin *et al.*, "antiSMASH 2.0—a versatile platform for genome mining of secondary metabolite producers," *Nucleic Acids Research*, vol. 41, no. W1, pp. W204–W212, 2013, doi: 10.1093/nar/gkt449.

58. T. Weber *et al.*, "antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters," *Nucleic Acids Research*, vol. 43, 2015, doi: 10.1093/nar/gkv437.

59. M. H. Medema *et al.*, "AntiSMASH: Rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences," *Nucleic Acids Research*, vol. 39, 2011, doi: 10.1093/nar/gkr466.

60. K. Blin *et al.*, "antiSMASH 6.0: improving cluster detection and comparison capabilities," *Nucleic Acids Research*, vol. 49, no. W1, p. W29, 2021, doi: 10.1093/NAR/GKAB335.

61. S. A. Kautsar, H. G. Suarez Duran, K. Blin, A. Osbourn, and M. H. Medema, "plantiSMASH: automated identification, annotation and expression analysis of plant biosynthetic gene clusters," *Nucleic Acids Research*, vol. 45, no. W1, pp. W55–W63, 2017, doi: 10.1093/nar/gkx305.

62. P. Cimermancic *et al.*, "Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters.," *Cell*, vol. 158, no. 2, pp. 412–21, 2014, doi: 10.1016/j.cell.2014.06.034.

63. M. G. Chevrette, F. Aicheler, O. Kohlbacher, C. R. Currie, and M. H. Medema, "SANDPUMA: ensemble predictions of nonribosomal peptide chemistry reveal biosynthetic diversity across Actinobacteria," *Bioinformatics*, vol. 33, no. 20, pp. 3202–3210, 2017, doi: 10.1093/bioinformatics/btx400.

64. T. Wolf, V. Shelest, N. Nath, and E. Shelest, "CASSIS and SMIPS: promoter-based prediction of secondary metabolite gene clusters in eukaryotic genomes," *Bioinformatics*, vol. 32, no. 8, pp. 1138–1143, 2016, doi: 10.1093/bioinformatics/btv713.

65. P. Jones *et al.*, "InterProScan 5: genome-scale protein function classification.," *Bioinformatics*, vol. 30, no. 9, pp. 1236–40, 2014, doi: 10.1093/bioinformatics/btu031.

66. N. Séelem-Mojica, C. Aguilar, K. Gutiéerrez-García, C. E. Martínez-Guerrero, and F. Barona-Gómez, "EvoMining reveals the origin and fate of natural product biosynthetic enzymes," *Microbial Genomics*, 2019, doi: 10.1099/MGEN.0.000260.

67. P. Cruz-Morales *et al.*, "Phylogenomic Analysis of Natural Products Biosynthetic Gene Clusters Allows Discovery of Arseno-Organic Metabolites in Model Streptomycetes," *Genome Biology and Evolution*, vol. 8, no. 6, pp. 1906–1916, 2016, doi: 10.1093/gbe/evw125.

68. A. J. van Heel, A. de Jong, C. Song, J. H. Viel, J. Kok, and O. P. Kuipers, "BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins," *Nucleic Acids Research*, vol. 46, 2018, doi: 10.1093/nar/gky383.

69. A. J. van Heel, A. de Jong, M. Montalbán-López, J. Kok, and O. P. Kuipers, "BAGEL3: automated identification of genes encoding bacteriocins and (non-) bactericidal posttranslationally modified peptides," *Nucleic Acids Research*, vol. 41, no. W1, pp. W448-453, Jul. 2013, doi: 10.1093/nar/gkt391.

70. M. A. Skinnider, N. J. Merwin, C. W. Johnston, and N. A. Magarvey, "PRISM 3: expanded prediction of natural product chemical structures from microbial genomes.," *Nucleic Acids Research*, vol. 45, no. Web Server issue, p. W49, 2017, doi: 10.1093/nar/gkx320.

71. M. A. Skinnider *et al.*, "Comprehensive prediction of secondary metabolite structure and biological activity from microbial genome sequences," *Nature Communications*, vol. 11, no. 1, pp. 1–9, 2020, doi: 10.1038/s41467-020-19986-1.

72. G. D. Hannigan *et al.*, "A deep learning genome-mining strategy for biosynthetic gene cluster prediction," *Nucleic Acids Research*, vol. 47, no. 18, p. e110, 2019, doi: 10.1093/NAR/GKZ654.

73. M. Alanjary *et al.*, "The Antibiotic Resistant Target Seeker (ARTS), an exploration engine for antibiotic cluster prioritization and novel drug target discovery," *Nucleic Acids Research*, vol. 45, no. W1, pp. W42–W48, 2017, doi: 10.1093/nar/gkx360.

74. A. Peña, F. del Carratore, M. Cummings, E. Takano, and R. Breitling, "Output ordering and prioritisation system (OOPS): ranking biosynthetic gene clusters to enhance bioactive metabolite discovery," *Journal of Industrial Microbiology & Biotechnology*, vol. 45, no. 7, pp. 615–619, 2017, doi: 10.1007/s10295-017-1993-1.

75. M. D. Mungan, M. Alanjary, K. Blin, T. Weber, M. H. Medema, and N. Ziemert, "ARTS 2.0: feature updates and expansion of the Antibiotic Resistant Target Seeker for comparative genome mining," *Nucleic Acids Research*, vol. 48, no. W1, pp. W546-552, 2020, doi: 10.1093/NAR/GKAA374.

76. A. Du Preez van Staden, "*In vitro* and *In vivo* Characterization of Amyloliquecidin, a Novel Two-Component Lantibiotic Produced by *Bacillus amyloliquefaciens*," PhD thesis, Stellenbosch University, Stellenbosch, 2015.

77. J. C. Nielsen *et al.*, "Global analysis of biosynthetic gene clusters reveals vast potential of secondary metabolite production in *Penicillium* species," *Nature Microbiology*, vol. 2, no. 6, p. 17044, 2017, doi: 10.1038/nmicrobiol.2017.44.

78. O. Genilloud, "The re-emerging role of microbial natural products in antibiotic discovery," *Antonie Van Leeuwenhoek*, vol. 106, no. 1, pp. 173–188, 2014, doi: 10.1007/s10482-014-0204-6.

79. O. Genilloud, "Actinomycetes: still a source of novel antibiotics.," *Nat Prod Rep*, vol. 34, no. 10, pp. 1203–1232, 2017, doi: 10.1039/c7np00026j.

80. R. H. Baltz, "Antimicrobials from Actinomycetes: Back to the future," *Microbiol.*, vol. 2, no. 3, pp. 125–131, 2007.

81. R. H. Baltz, "Gifted microbes for genome mining and natural product discovery," *Journal of Industrial Microbiology & Biotechnology*, vol. 44, no. 4–5, pp. 573–588, 2017, doi: 10.1007/s10295-016-1815-x.

82. K. Tiwari and R. K. Gupta, "Rare actinomycetes: a potential storehouse for novel antibiotics," *Critical Reviews in Biotechnology*, vol. 32, no. 2, pp. 108–132, 2012, doi: 10.3109/07388551.2011.562482.

83. J. A. Gerlt and P. C. Babbitt, "Divergent Evolution of Enzymatic Function: Mechanistically Diverse Superfamilies and Functionally Distinct Suprafamilies," *Annual Review of Biochemistry*, vol. 70, no. 1, pp. 209–246, 2001, doi: 10.1146/annurev.biochem.70.1.209.

84. H. Jenke-Kodama, A. Sandmann, R. Müller, and E. Dittmann, "Evolutionary Implications of Bacterial Polyketide Synthases," *Molecular Biology and Evolution*, vol. 22, no. 10, pp. 2027–2039, 2005, doi: 10.1093/molbev/msi193.

85. M. H. Medema *et al.*, "Minimum Information about a Biosynthetic Gene cluster.," *Nature Chemical Biology*, vol. 11, no. 9, pp. 625–31, 2015, doi: 10.1038/nchembio.1890.

86. M. H. Medema, P. Cimermancic, A. Sali, E. Takano, and M. A. Fischbach, "A Systematic Computational Analysis of Biosynthetic Gene Cluster Evolution: Lessons for Engineering Biosynthesis," *PLoS Computational Biology*, vol. 10, no. 12, p. e1004016, 2014, doi: 10.1371/journal.pcbi.1004016.

87. R. Kumar, P. K. Jangir, J. Das, B. Taneja, and R. Sharma, "Genome Analysis of *Staphylococcus capitis* TE8 Reveals Repertoire of Antimicrobial Peptides and Adaptation Strategies for Growth on Human Skin.," *Scientific Reports*, vol. 7, no. 1, p. 10447, 2017, doi: 10.1038/s41598-017-11020-7.

88. A. J. van Heel, M. Montalban-Lopez, Q. Oliveau, and O. P. Kuipers, "Genome-guided identification of novel head-to-tail cyclized antimicrobial peptides, exemplified by the discovery of pumilarin.," *Microb Genom*, vol. 3, no. 10, p. e000134, 2017, doi: 10.1099/mgen.0.000134.

89. L. Belbahri *et al.*, "Comparative Genomics of *Bacillus amyloliquefaciens* Strains Reveals a Core Genome with Traits for Habitat Adaptation and a Secondary Metabolites Rich Accessory Genome.," *Front Microbiol*, vol. 8, p. 1438, 2017, doi: 10.3389/fmicb.2017.01438.

90. Y. Qi, J. M. D'Alessandro, and J. A. V Blodgett, "Draft Genome Sequence of *Streptomyces* sp. Strain JV178, a Producer of Clifednamide-Type Polycyclic Tetramate Macrolactams.," *Genome Announcements*, vol. 6, no. 1, 2018, doi: 10.1128/genomeA.01401-17.

91. C. Olano *et al.*, "Activation and identification of five clusters for secondary metabolites in *Streptomyces albus* J1074.," *Microbial Biotechnology*, vol. 7, no. 3, pp. 242–56, 2014, doi: 10.1111/1751-7915.12116.

92. H.-S. Kang, Z. Charlop-Powers, and S. F. Brady, "Multiplexed CRISPR/Cas9- and TAR-Mediated Promoter Engineering of Natural Product Biosynthetic Gene Clusters in Yeast," *ACS Synthetic Biology*, vol. 5, no. 9, pp. 1002–1010, 2016, doi: 10.1021/acssynbio.6b00080.

93. M. M. Zhang *et al.*, "CRISPR-Cas9 strategy for activation of silent Streptomyces biosynthetic gene clusters.," *Nature Chemical Biology*, 2017, doi: 10.1038/nchembio.2341.

94. Z. Shao, G. Rao, C. Li, Z. Abil, Y. Luo, and H. Zhao, "Refactoring the Silent Spectinabilin Gene Cluster Using a Plug-and-Play Scaffold," *ACS Synthetic Biology*, vol. 2, no. 11, pp. 662–669, 2013, doi: 10.1021/sb400058n.

95. H. Ren, P. Hu, and H. Zhao, "A plug-and-play pathway refactoring workflow for natural product research in *Escherichia coli* and *Saccharomyces cerevisiae*," *Biotechnology and Bioengineering*, vol. 114, no. 8, pp. 1847–1854, 2017, doi: 10.1002/bit.26309.

96. J. Xu, J. Zhang, J. Zhuo, Y. Li, Y. Tian, and H. Tan, "Activation and mechanism of a cryptic oviedomycin gene cluster via the disruption of a global regulatory gene, *adpA*, in *Streptomyces ansochromogenes*.," *The Journal of Biological Chemistry*, vol. 292, no. 48, pp. 19708–19720, 2017, doi: 10.1074/jbc.M117.809145.

97. M. J. Smanski *et al.*, "Functional optimization of gene clusters by combinatorial design and assembly," *Nature Biotechnology 2014 32:12*, vol. 32, no. 12, pp. 1241–1249, 2014, doi: 10.1038/nbt.3063.

98. Y. Pan, G. Liu, H. Yang, Y. Tian, and H. Tan, "The pleiotropic regulator AdpA-L directly controls the pathway-specific activator of nikkomycin biosynthesis in *Streptomyces ansochromogenes*," *Molecular Microbiology*, vol. 72, no. 3, pp. 710–723, 2009, doi: 10.1111/j.1365-2958.2009.06681.x.

99. T. Higashi, Y. Iwasaki, Y. Ohnishi, and S. Horinouchi, "A-factor and phosphate depletion signals are transmitted to the grixazone biosynthesis genes via the pathway-specific transcriptional activator GriR.," *Journal of Bacteriology*, vol. 189, no. 9, pp. 3515–24, 2007, doi: 10.1128/JB.00055-07.

100. Y. Ohnishi, S. Kameyama, H. Onaka, and S. Horinouchi, "The A-factor regulatory cascade leading to streptomycin biosynthesis in *Streptomyces griseus* : identification of a target gene of the A-factor receptor.," *Mol Microbiol*, vol. 34, no. 1, pp. 102–11, 1999.

101. G. Liu, Y. Tian, H. Yang, and H. Tan, "A pathway-specific transcriptional regulatory gene for nikkomycin biosynthesis in *Streptomyces ansochromogenes* that also influences colony development," *Molecular Microbiology*, vol. 55, no. 6, pp. 1855–1866, 2005, doi: 10.1111/j.1365-2958.2005.04512.x.

102. T. V. Hung, S. Malla, B. C. Park, K. Liou, H. C. Lee, and J. K. Sohng, "Enhancement of clavulanic acid by replicative and integrative expression of *ccaR* and *cas2* in *Streptomyces clavuligerus* NRRL3585.," *Journal of Microbiology and Biotechnology*, vol. 17, no. 9, pp. 1538–45, 2007.

103. J. Garcia-Bernardo, A. F. Braña, C. Méndez, and J. A. Salas, "Insertional inactivation of *mtrX* and *mtrY* genes from the mithramycin gene cluster affects production and growth of the producer organism *Streptomyces argillaceus*.," *FEMS Microbiology Letters*, vol. 186, no. 1, pp. 61–5, 2000.

104. A. de la Fuente, L. M. Lorenzana, J. F. Martín, and P. Liras, "Mutants of *Streptomyces clavuligerus* with disruptions in different genes for clavulanic acid biosynthesis produce large amounts of holomycin: possible cross-regulation of two unrelated secondary metabolic pathways.," *Journal of Bacteriology*, vol. 184, no. 23, pp. 6559–65, 2002, doi: 10.1128/JB.184.23.6559-6565.2002.

105. J. Huang *et al.*, "Cross-regulation among disparate antibiotic biosynthetic pathways of *Streptomyces coelicolor*," *Molecular Microbiology*, vol. 58, no. 5, pp. 1276–1287, 2005, doi: 10.1111/j.1365-2958.2005.04879.x.

106. G.-Y. Tan and T. Liu, "Rational synthetic pathway refactoring of natural products biosynthesis in actinobacteria," *Metabolic Engineering*, vol. 39, pp. 228–236, 2017, doi: 10.1016/J.YMBEN.2016.12.006.

107. H. U. van der Heul, B. L. Bilyk, K. J. McDowall, R. F. Seipke, and G. P. van Wezel, "Regulation of antibiotic production in Actinobacteria: new perspectives from the post-genomic era," *Natural Product Reports*, vol. 35, no. 6, pp. 575–604, 2018, doi: 10.1039/C8NP00012C.

108. H. S. Park, H. J. Nah, S. H. Kang, S. S. Choi, and E. S. Kim, "Screening and Isolation of a Novel Polyene-Producing *Streptomyces* Strain Inhibiting Phytopathogenic Fungi in the Soil Environment," *Frontiers in Bioengineering and Biotechnology*, vol. 9, p. 692340, 2021, doi: 10.3389/FBIOE.2021.692340/FULL.

109. R. I. Santamaría *et al.*, "Characterization of Actinomycetes Strains Isolated from the Intestinal Tract and Feces of the Larvae of the Longhorn Beetle *Cerambyx welensii*," *Microorganisms*, vol. 8, no. 12, p. 2013, Dec. 2020, doi: 10.3390/MICROORGANISMS8122013.

110. S. Haeder, R. Wirth, H. Herz, and D. Spiteller, "Candicidin-producing Streptomyces support leaf-cutting ants to protect their fungus garden against the pathogenic fungus *Escovopsis*," *PNAS*, vol. 106, no. 12, pp. 4742–4746, 2009, doi: 10.1073/PNAS.0812082106.

111. K. J. Grubbs *et al.*, "Pollen *Streptomyces* Produce Antibiotic That Inhibits the Honey Bee Pathogen *Paenibacillus* larvae," *Frontiers in Microbiology*, vol. 12, p. 115, 2021, doi: 10.3389/FMICB.2021.632637.

112. D. R. Kim *et al.*, "A mutualistic interaction between *Streptomyces* bacteria, strawberry plants and pollinating bees," *Nature Communications*, vol. 10, no. 1, pp. 1–10, 2019, doi: 10.1038/s41467-019-12785-3.

113. J. P. Gomez-Escribano and M. J. Bibb, "Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters," *Microbial Biotechnology*, vol. 4, no. 2, pp. 207–215, 2011, doi: 10.1111/j.1751-7915.2010.00219.x.

114. R. Breitling *et al.*, "Synthetic biology approaches to actinomycete strain improvement," *FEMS Microbiology Letters*, vol. 368, no. 10, 2021, doi: 10.1093/FEMSLE/FNAB060.

115. J. Fu *et al.*, "Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting," *Nature Biotechnology*, vol. 30, no. 5, pp. 440–446, 2012, doi: 10.1038/nbt.2183.

116. N. Kouprina and V. Larionov, "Transformation-associated recombination (TAR) cloning for genomics studies and synthetic biology," *Chromosoma*, vol. 125, no. 4, pp. 621–632, 2016, doi: 10.1007/s00412-016-0588-3.

117. H. Liu *et al.*, "Rapid Cloning and Heterologous Expression of the Meridamycin Biosynthetic Gene Cluster Using a Versatile *Escherichia coli–Streptomyces* Artificial Chromosome Vector, pSBAC ⊥," *Journal of Natural Products*, vol. 72, no. 3, pp. 389–395, 2009, doi: 10.1021/np8006149.

118. H.-J. Nah, H.-R. Pyeon, S.-H. Kang, S.-S. Choi, and E.-S. Kim, "Cloning and Heterologous Expression of a Large-sized Natural Product Biosynthetic Gene Cluster in *Streptomyces* Species," *Frontiers in Microbiology*, vol. 8, no. Mar, p. 394, 2017, doi: 10.3389/fmicb.2017.00394.

119. K. Yamanaka *et al.*, "Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A," *PNAS*, vol. 111, no. 5, pp. 1957–1962, 2014, doi: 10.1073/pnas.1319584111.

120. K. A. Reynolds *et al.*, "Isolation and structure elucidation of lipopeptide antibiotic taromycin B from the activated taromycin biosynthetic gene cluster," *The Journal of Antibiotics*, 2017, doi: 10.1038/ja.2017.146.

121. B. Bonet, R. Teufel, M. Crüsemann, N. Ziemert, and B. S. Moore, "Direct capture and heterologous expression of Salinispora natural product genes for the biosynthesis of enterocin.," *Journal of Natural Products*, vol. 78, no. 3, pp. 539–42, 2015, doi: 10.1021/np500664q.

122. J. Yin *et al.*, "Direct cloning and heterologous expression of the salinomycin biosynthetic gene cluster from *Streptomyces albus* DSM41398 in *Streptomyces coelicolor* A3(2)," *Scientific Reports*, vol. 5, no. 1, p. 15081, 2015, doi: 10.1038/srep15081.

123. H.-J. Nah, M.-W. Woo, S.-S. Choi, and E.-S. Kim, "Precise cloning and tandem integration of large polyketide biosynthetic gene cluster using *Streptomyces* artificial chromosome system.," *Microbial Cell Factories*, vol. 14, p. 140, Sep. 2015, doi: 10.1186/s12934-015-0325-2.

124. H.-R. Pyeon, H.-J. Nah, S.-H. Kang, S.-S. Choi, and E.-S. Kim, "Heterologous expression of pikromycin biosynthetic gene cluster using *Streptomyces* artificial chromosome system.," *Microbial Cell Factories*, vol. 16, no. 1, p. 96, May 2017, doi: 10.1186/s12934-017-0708-7.

125. B. Enghiad *et al.*, "Cas12a-assisted precise targeted cloning using in vivo Cre-lox recombination," *Nature Communications*, vol. 12, no. 1, pp. 1–11, 2021, doi: 10.1038/s41467-021-21275-4.

126. Y. Shi *et al.*, "Biosynthesis of antibiotic chuangxinmycin from *Actinoplanes tsinanensis*," *Acta Pharmaceutica Sinica B*, 2017, doi: 10.1016/J.APSB.2017.07.005.

127. L. Linares-Otoya *et al.*, "Identification and heterologous expression of the kocurin biosynthetic gene cluster," *Microbiology*, vol. 163, no. 10, pp. 1409–1414, 2017, doi: 10.1099/mic.0.000538.

128. T. S. Young and C. T. Walsh, "Identification of the thiazolyl peptide GE37468 gene cluster from *Streptomyces* ATCC 55365 and heterologous expression in *Streptomyces lividans*.," *PNAS*, vol. 108, no. 32, pp. 13053–8, 2011, doi: 10.1073/pnas.1110435108.

129. A. Tocchetti *et al.*, "Capturing Linear Intermediates and C-Terminal Variants during Maturation of the Thiopeptide GE2270," *Chemistry & Biology*, vol. 20, no. 8, pp. 1067–1077, 2013, doi: 10.1016/J.CHEMBIOL.2013.07.005.

130. K. Flinspach, C. Kapitzke, A. Tocchetti, M. Sosio, and A. K. Apel, "Heterologous Expression of the Thiopeptide Antibiotic GE2270 from *Planobispora rosea* ATCC 53733 in *Streptomyces coelicolor* Requires Deletion of Ribosomal Genes from the Expression Construct," *PLoS ONE*, vol. 9, no. 3, p. e90499, 2014, doi: 10.1371/journal.pone.0090499.

131. B. Kepplinger *et al.*, "Mode of Action and Heterologous Expression of the Natural Product Antibiotic Vancoresmycin," *ACS Chemical Biology*, p. acschembio.7b00733, 2017, doi: 10.1021/acschembio.7b00733.

132. H. Ikeda, K. Shin-ya, and S. Omura, "Genome mining of the *Streptomyces avermitilis* genome and development of genome-minimized hosts for heterologous expression of biosynthetic gene clusters," *Journal of Industrial Microbiology & Biotechnology*, vol. 41, no. 2, pp. 233–250, 2014, doi: 10.1007/s10295-013-1327-x.

133. R. Novakova *et al.*, "Increased heterologous production of the antitumoral polyketide mithramycin A by engineered *Streptomyces lividans* TK24 strains," *Applied Microbiology and Biotechnology*, vol. 102, no. 2, pp. 857–869, 2018, doi: 10.1007/s00253-017-8642-5.

134. Y. Wang *et al.*, "Iteratively improving natamycin production in *Streptomyces gilvosporeus* by a large operon-reporter based strategy," *Metabolic Engineering*, vol. 38, pp. 418–426, 2016, doi: 10.1016/j.ymben.2016.10.005.

135. N. Zaburannyi, M. Rabyk, B. Ostash, V. Fedorenko, and A. Luzhetskyy, "Insights into naturally minimised *Streptomyces albus* J1074 genome," *BMC Genomics*, vol. 15, no. 1, p. 97, 2014, doi: 10.1186/1471-2164-15-97.

136. N. Manderscheid *et al.*, "An influence of the copy number of biosynthetic gene clusters on the production level of antibiotics in a heterologous host," *Journal of Biotechnology*, vol. 232, pp. 110–117, 2016, doi: 10.1016/j.jbiotec.2016.05.038.

137. B. M. Hover *et al.*, "Culture-independent discovery of the malacidins as calcium-dependent antibiotics with activity against multidrug-resistant Gram-positive pathogens," *Nature Microbiology*, p. 1, 2018, doi: 10.1038/s41564-018-0110-1.

138. E. Breukink and B. de Kruijff, "Lipid II as a target for antibiotics," *Nature Reviews Drug Discovery*, vol. 5, no. 4, pp. 321–323, 2006, doi: 10.1038/nrd2004.

139. B. Bilyk, L. Horbal, and A. Luzhetskyy, "Chromosomal position effect influences the heterologous expression of genes and biosynthetic gene clusters in *Streptomyces albus* J1074," *Microb Cell Fact*, vol. 16, no. 1, 2017, doi: 10.1186/S12934-016-0619-Z.

140. H. A. Iqbal, L. Low-Beinart, J. U. Obiajulu, and S. F. Brady, "Natural Product Discovery through Improved Functional Metagenomics in *Streptomyces*," *Journal of the American Chemical Society*, vol. 138, no. 30, p. 9341, 2016, doi: 10.1021/JACS.6B02921.

141. B. A. Pfeifer, S. J. Admiraal, H. Gramajo, D. E. Cane, and C. Khosla, "Biosynthesis of Complex Polyketides in a Metabolically Engineered Strain of *E. coli*," *Science*, vol. 291, no. 5509, pp. 1790–1792, 2001, doi: 10.1126/science.1058092.

142. Y.-G. Ryu, M. J. Butler, K. F. Chater, and K. J. Lee, "Engineering of primary carbohydrate metabolism for increased production of actinorhodin in *Streptomyces coelicolor*," *Appl Environ Microbiol*, vol. 72, no. 11, pp. 7132–9, 2006, doi: 10.1128/AEM.01308-06.

143. J. A. Jones *et al.*, "Complete Biosynthesis of Anthocyanins Using *E. coli* Polycultures," *mBio*, vol. 8, no. 3, pp. e00621-17, 2017, doi: 10.1128/mBio.00621-17.

144. V. Ganesan, Z. Li, X. Wang, and H. Zhang, "Heterologous biosynthesis of natural product naringenin by co-culture engineering," *Synthetic and Systems Biotechnology*, vol. 2, no. 3, pp. 236–242, 2017, doi: 10.1016/J.SYNBIO.2017.08.003.

145. F. Xu *et al.*, "A Genetics-Free Method for High-Throughput Discovery of Cryptic Microbial Metabolites," *Nature Chemical Biology*, vol. 15, no. 2, p. 161, 2019, doi: 10.1038/S41589-018-0193-2.

146. F. Xu, B. Nazari, K. Moon, L. B. Bushin, and M. R. Seyedsayamdost, "Discovery of a Cryptic Antifungal Compound from *Streptomyces albus* J1074 Using High-Throughput Elicitor Screens," *Journal of the American Chemical Society*, vol. 139, no. 27, p. 9203, 2017, doi: 10.1021/JACS.7B02716.

147. K. S. Makarova *et al.*, "Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants," *Nature Reviews Microbiology*. 2020. doi: 10.1038/s41579-019-0299-x.

148. Y. Tong, T. Weber, and S. Y. Lee, "CRISPR/Cas-based genome engineering in natural product discovery," *Natural Product Reports*, vol. 36, no. 9., pp. 1262–1280, 2019. doi: 10.1039/c8np00089a.

149. H. Nishimasu *et al.*, "Crystal structure of Cas9 in complex with guide RNA and target DNA," *Cell*, 2014, doi: 10.1016/j.cell.2014.02.001.

150. M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier, "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity," *Science*, 2012, doi: 10.1126/science.1225829.

151. W. Jiang, D. Bikard, D. Cox, F. Zhang, and L. A. Marraffini, "RNA-guided editing of bacterial genomes using CRISPR-Cas systems," *Nature Biotechnology*, 2013, doi: 10.1038/nbt.2508.

152. Y. Fu *et al.*, "High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells," *Nature Biotechnology*, vol. 31, no. 9, pp. 822–826, 2013, doi: 10.1038/nbt.2623.

153. P. D. Hsu *et al.*, "DNA targeting specificity of RNA-guided Cas9 nucleases," *Nature Biotechnology*, 2013, doi: 10.1038/nbt.2647.

154. R. E. Cobb, Y. Wang, and H. Zhao, "High-Efficiency Multiplex Genome Editing of *Streptomyces* Species Using an Engineered CRISPR/Cas System," *ACS Synthetic Biology*, 2015, doi: 10.1021/sb500351f.

155. Y. Tong, P. Charusanti, L. Zhang, T. Weber, and S. Y. Lee, "CRISPR-Cas9 Based Engineering of Actinomycetal Genomes," *ACS Synthetic Biology*, vol. 4, no. 9, pp. 1020–1029, 2015, doi: 10.1021/acssynbio.5b00038.

156. H. Huang, G. Zheng, W. Jiang, H. Hu, and Y. Lu, "One-step high-efficiency CRISPR/Cas9-mediated genome editing in *Streptomyces*," *Acta Biochimica et Biophysica Sinica*, vol. 47, no. 4, pp. 231–243, 2015, doi: 10.1093/abbs/gmv007.

157. H. Zeng *et al.*, "Highly efficient editing of the actinorhodin polyketide chain length factor gene in *Streptomyces coelicolor* M145 using CRISPR/Cas9-CodA(sm) combined system," *Applied Microbiology and Biotechnology*, vol. 99, no. 24, pp. 10575–10585, 2015, doi: 10.1007/s00253-015-6931-4.

158. W. L. Yeo *et al.*, "Characterization of Cas proteins for CRISPR-Cas editing in streptomycetes," *Biotechnology and Bioengineering*, vol. 116, no. 9, pp. 2330–2338, 2019, doi: 10.1002/bit.27021.

159. Z. J. Low *et al.*, "Identification of a biosynthetic gene cluster for the polyene macrolactam sceliphrolactam in a *Streptomyces* strain isolated from mangrove sediment," *Scientific Reports*, 2018, doi: 10.1038/s41598-018-20018-8.

160. F. Alberti, D. J. Leng, I. Wilkening, L. Song, M. Tosin, and C. Corre, "Triggering the expression of a silent gene cluster from genetically intractable bacteria results in scleric acid discovery," *Chemical Science*, vol. 10, no. 2, pp. 453–463, 2019, doi: 10.1039/c8sc03814g.

161. S. Ye, B. Enghiad, H. Zhao, and E. Takano, "Fine-tuning the regulation of Cas9 expression levels for efficient CRISPR-Cas9 mediated recombination in *Streptomyces*," *Journal of Industrial Microbiology & Biotechnology*, vol. 47, no. 4–5, pp. 413–423, 2020, doi: 10.1007/s10295-020-02277-5.

162. K. Wang *et al.*, "Multi-Layer Controls of Cas9 Activity Coupled With ATP Synthase Over-Expression for Efficient Genome Editing in *Streptomyces*," *Frontiers in Bioengineering and Biotechnology*, vol. 7, p. 304, 2019, doi: 10.3389/fbioe.2019.00304.

163. M. Myronovskyi and A. Luzhetskyy, "Native and engineered promoters in natural product discovery," *Natural Product Reports*, vol. 33, no. 8, pp. 1006–19, 2016, doi: 10.1039/c6np00002a.

164. Y.-H. Jiang *et al.*, "Fine-Tuning Cas9 Activity with a Cognate Inhibitor AcrIIA4 to Improve Genome Editing in *Streptomyces*," *ACS Synthetic Biology*, p. acssynbio.1c00141, 2021, doi: 10.1021/ACSSYNBIO.1C00141.

165. S. Najah, C. Saulnier, J. L. Pernodet, and S. Bury-Moné, "Design of a generic CRISPR-Cas9 approach using the same sgRNA to perform gene editing at distinct loci," *BMC Biotechnology*, vol. 19, no. 1, p. 18, 2019, doi: 10.1186/s12896-019-0509-7.

166. Q. Wang *et al.*, "Dual-function chromogenic screening-based CRISPR/Cas9 genome editing system for actinomycetes," *Applied Microbiology and Biotechnology*, vol. 104, no. 1, pp. 225–239, 2020, doi: 10.1007/s00253-019-10223-4.

167. Z. Qin *et al.*, "Formicamycins, antibacterial polyketides produced by *Streptomyces formicae* isolated from African Tetraponera plant-ants," *Chemical Science*, 2017, doi: 10.1039/c6sc04265a.

168. M. M. Zhang *et al.*, "CRISPR-Cas9 strategy for activation of silent *Streptomyces* biosynthetic gene clusters," *Nature Chemical Biology*, vol. 13, no. 6, pp. 607–609, 2017, doi: 10.1038/NCHEMBIO.2341.

169. S. H. Kim, W. Lu, M. K. Ahmadi, D. Montiel, M. A. Ternei, and S. F. Brady, "Atolypenes, tricyclic bacterial sesterterpenes discovered using a multiplexed in vitro Cas9-TAR gene cluster refactoring approach," *ACS Synthetic Biology*, vol. 8, no. 1, p. 109, 2019, doi: 10.1021/ACSSYNBIO.8B00361.

170. E. J. Culp, G. Yim, N. Waglechner, W. Wang, A. C. Pawlowski, and G. D. Wright, "Hidden antibiotics in actinomycetes can be identified by inactivation of gene clusters for common antibiotics," *Nature Biotechnology*, vol. 37, no. 10, pp. 1149–1154, 2019, doi: 10.1038/s41587-019-0241-9.

171. A. Schatz, E. Bugle, and S. A. Waksman, "Streptomycin, a Substance Exhibiting Antibiotic Activity Against Gram-Positive and Gram-Negative Bacteria," *Proceedings of the Society for Experimental Biology and Medicine*, vol. 55, no. 1, pp. 66–69, 1944, doi: 10.3181/00379727-55-14461.

172. K. M. Krause, A. W. Serio, T. R. Kane, and L. E. Connolly, "Aminoglycosides: An overview," *Cold Spring Harbor Perspectives in Medicine*, vol. 6, no. 6, Jun. 2016, doi: 10.1101/cshperspect.a027029.

173. E. F. Goljan, *Rapid Review Pathology*. 2017.

174. J. K. Aronson, *Gentamicin*. 2016. doi: 10.1016/c2009-1-62477-1.

175. H. Umezawa *et al.*, "Production and isolation of a new antibiotic: kanamycin.," *Journal of Antibiotics (Tokyo)*, vol. 10, no. 5, pp. 181–188, 1957.

176. S. A. Waksman and H. A. Lechevalier, "Neomycin, a new antibiotic active against streptomycin-resistant bacteria, including tuberculosis organisms," *Science (1979)*, vol. 109, no. 2830, pp. 305–307, 1949, doi: 10.1126/science.109.2830.305.

177. M. J. Weinstein *et al.*, "Gentamicin, a New Antibiotic Complex from *Micromonospora*," *Journal of Medicinal Chemistry*, vol. 6, no. 4. pp. 463–464, Jul. 01, 1963. doi: 10.1021/jm00340a034.

178. Y. Takahashi and M. Igarashi, "Destination of aminoglycoside antibiotics in the 'post-antibiotic era,'" *Nature Publishing Group*, vol. 71, pp. 4–14, 2017, doi: 10.1038/ja.2017.117.

179. M. Hainrichson, I. Nudelman, and T. Baasov, "Designer aminoglycosides: the race to develop improved antibiotics and compounds for the treatment of human genetic diseases", doi: 10.1039/b712690p.

180. A. Tsai *et al.*, "The Impact of Aminoglycosides on the Dynamics of Translation Elongation," *Cell Reports*, vol. 3, no. 2, pp. 497–508, 2013, doi: 10.1016/j.celrep.2013.01.027.

181. D. N. Wilson, "Ribosome-targeting antibiotics and mechanisms of bacterial resistance," *Nature Reviews Microbiology*, vol. 12, no. 1. Nat Rev Microbiol, pp. 35–48, 2014. doi: 10.1038/nrmicro3155.

182. B. D. Davis, L. Chen, and P. C. Tai, "Misread protein creates membrane channels: An essential step in the bactericidal action of aminoglycosides," *PNAS*, vol. 83, no. 16, pp. 6164–6168, 1986, doi: 10.1073/pnas.83.16.6164.

183. M. Kaul, C. M. Barbieri, and D. S. Pilch, "Fluorescence-Based Approach for Detecting and Characterizing Antibiotic-Induced Conformational Changes in Ribosomal RNA: Comparing Aminoglycoside Binding to Prokaryotic and Eukaryotic Ribosomal RNA Sequences," *Journal of the American Chemical Society*, 2004, doi: 10.1021/ja030568i.

184. S. Shandrick *et al.*, "Monitoring molecular recognition of the ribosomal decoding site," *Angewandte Chemie - International Edition*, 2004, doi: 10.1002/anie.200454217.

185. H. J. Busse, C. Wostmann, and E. P. Bakker, "The bactericidal action of streptomycin: Membrane permeabilization caused by the insertion of mistranslated proteins into the cytoplasmic membrane of *Escherichia coli* and subsequent caging of the antibiotic inside the cells due to degradation of these pro," *Journal of General Microbiology*, 1992, doi: 10.1099/00221287-138-3-551.

186. M. Misumi and N. Tanaka, "Mechanism of inhibition of translocation by kanamycin and viomycin: A comparative study with fusidic acid," *Biochemical and Biophysical Research Communications*, 1980, doi: 10.1016/0006-291X(80)90382-4.

187. M. A. Borovinskaya *et al.*, "Structural basis for aminoglycoside inhibition of bacterial ribosome recycling," *Nature Structural and Molecular Biology*, 2007, doi: 10.1038/nsmb1271.

188. L. Wang *et al.*, "Allosteric control of the ribosome by small-molecule antibiotics," *Nature Structural and Molecular Biology*, 2012, doi: 10.1038/nsmb.2360.

189. M. R. Wasserman *et al.*, "Chemically related 4,5-linked aminoglycoside antibiotics drive subunit rotation in opposite directions," *Nature Communications*, 2015, doi: 10.1038/ncomms8896.

190. A. Okuyama, T. Watanabe, and N. Tanaka, "Effects of aminoglycoside antibiotics on initiation of viral RNA-directed protein synthesis," *The Journal of Antibiotics*, 1972, doi: 10.7164/antibiotics.25.212.

191. A. P. Carter, W. M. Clemons, D. E. Brodersen, R. J. Morgan-Warren, B. T. Wimberly, and V. Ramakrishnan, "Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics," *Nature*, 2000, doi: 10.1038/35030019.

192. T. Hermann, "Aminoglycoside antibiotics: Old drugs and new therapeutic approaches," *Cellular and Molecular Life Sciences*. 2007. doi: 10.1007/s00018-007-7034-x.

193. S. Garneau-Tsodikova and K. J. Labby, "Mechanisms of resistance to aminoglycoside antibiotics: Overview and perspectives," *MedChemComm*, vol. 7, no. 1, pp. 11–27, 2016. doi: 10.1039/c5md00344j.

194. C. E. Maus, B. B. Plikaytis, and T. M. Shinnick, "Molecular analysis of cross-resistance to capreomycin, kanamycin, amikacin, and viomycin in Mycobacterium tuberculosis," *Antimicrobial Agents and Chemotherapy*, vol. 49, no. 8, pp. 3192–3197, Aug. 2005, doi: 10.1128/AAC.49.8.3192-3197.2005.

195. M. S. Ramirez and M. E. Tolmasky, "Aminoglycoside modifying enzymes," *Drug Resistance Updates*, vol. 13, no. 6, pp. 151–171, 2010, doi: 10.1016/j.drup.2010.08.003.

196. H. C. Hinshaw, M. M. Pyle, and W. H. Feldman, "Streptomycin in tuberculosis," *The American Journal of Medicine*, vol. 2, no. 5, pp. 429–435, 1947, doi: 10.1016/0002-9343(47)90087-9.

197. J. Fabre, M. Rudhardt, P. Blanchard, and C. Regamey, "Persistence of sisomicin and gentamicin in renal cortex and medulla compared with other organs and serum of rats.," *Kidney International*, vol. 10, no. 6, pp. 444–9, 1976.

198. A. Vandewalle, N. Farman, J. P. Morin, J. P. Fillastre, P. Y. Hatt, and J. P. Bonvalet, "Gentamicin incorporation along the nephron: autoradiographic study on isolated tubules.," *Kidney International*, vol. 19, no. 4, pp. 529–39, 1981.

199. J. A. Segal and P. Skolnick, "Polyamine-like actions of aminoglycosides and aminoglycoside derivatives at NMDA receptors," *European Journal of Pharmacology*, 1998, doi: 10.1016/S0014-2999(98)00108-3.

200. M. D. Rizzi and K. Hirose, "Aminoglycoside ototoxicity," *Current Opinion in Otolaryngology & Head and Neck Surgery*, vol. 15, no. 5, pp. 352–357, 2007, doi: 10.1097/MOO.0b013e3282ef772d.

201. S. J. McWilliam, D. J. Antoine, R. L. Smyth, and M. Pirmohamed, "Aminoglycoside-induced nephrotoxicity in children.," *Pediatric Nephrology*, vol. 32, no. 11, pp. 2015–2025, 2017, doi: 10.1007/s00467-016-3533-z.

202. J. M. Lopez-Novoa, Y. Quiros, L. Vicente, A. I. Morales, and F. J. Lopez-Hernandez, "New insights into the mechanism of aminoglycoside nephrotoxicity: An integrative point of view," *Kidney International*, vol. 79, no. 1. pp. 33–45, 2011. doi: 10.1038/ki.2010.337.

203. J. Fabre, M. Rudhardt, P. Blanchard, and C. Regamey, "Persistence of sisomicin and gentamicin in renal cortex and medulla compared with other organs and serum of rats.," *Kidney International*, vol. 10, no. 6, pp. 444–9, 1976.

204. A. Vandewalle, N. Farman, J. P. Morin, J. P. Fillastre, P. Y. Hatt, and J. P. Bonvalet, "Gentamicin incorporation along the nephron: autoradiographic study on isolated tubules.," *Kidney International*, vol. 19, no. 4, pp. 529–39, 1981.

205. E. C. Böttger, "Mutant A1555G mitochondrial 12S rRNA and aminoglycoside susceptibility," *Antimicrobial Agents and Chemotherapy*. 2010. doi: 10.1128/AAC.01819-09.

206. L. A. Grohskopf, W. C. Huskins, R. L. Sinkowitz-Cochran, G. L. Levine, D. A. Goldmann, and W. R. Jarvis, "Use of antimicrobial agents in United States neonatal and pediatric intensive care patients," *Pediatric Infectious Disease Journal*, vol. 24, no. 9, pp. 766–773, 2005, doi: 10.1097/01.INF.0000178064.55193.1C.

207. J. A. Caminero, G. Sotgiu, A. Zumla, and G. B. Migliori, "Best drug treatment for multidrug-resistant and extensively drug-resistant tuberculosis," *The*

Lancet Infectious Diseases, vol. 10, no. 9, pp. 621–629, 2010, doi: 10.1016/S1473-3099(10)70139-0.

208. G. M. Pacifici, "Clinical pharmacokinetics of aminoglycosides in the neonate: A review," *European Journal of Clinical Pharmacology*, vol. 65, no. 4, pp. 419–427, 2009, doi: 10.1007/S00228-008-0599-Y.

209. S. Umezawa, T. Tsuchiya, R. Muto, Y. Nishimura, and H. Umezawa, "Synthesis of 3′-deoxykanamycin effective against kanamygin-resistant *Escherichia coli* and *Pseudomonas aeruginosa*," *The Journal of Antibiotics*, 1971, doi: 10.7164/antibiotics.24.274.

210. N. Thamban Chandrika and S. Garneau-Tsodikova, "Comprehensive review of chemical strategies for the preparation of new aminoglycosides and their biological activities," *Chemical Society Reviews*, vol. 47, no. 4, pp. 1189–1249, 2018. doi: 10.1039/c7cs00407a.

211. H. Kawaguchi, T. Naito, S. Nakagawa, and K. Ichi Fujisawa, "BB-K 8, a new semisynthetic aminoglycoside antibiotic," *The Journal of Antibiotics*, 1972, doi: 10.7164/antibiotics.25.695.

212. J. J. Wright, "Synthesis of 1-N-ethylsisomicin: A broad-spectrum semisynthetic aminoglycoside antibiotic," *Journal of the Chemical Society, Chemical Communications*, 1976, doi: 10.1039/C39760000206.

213. T. L. Nagabhushan, A. B. Cooper, H. Tsai, P. J. L. Daniels, and G. H. Miller, "The syntheses and biological properties of l-n-(s-4-amino-2-hydroxybutyryl)-gentamicin b and 1-n-(s-3-amino-2-hydroxypropionyl)-gentamicin b," *The Journal of Antibiotics*, vol. 31, no. 7, pp. 681–687, 1978, doi: 10.7164/antibiotics.31.681.

214. Y. Doi, J. ichi Wachino, and Y. Arakawa, "Aminoglycoside Resistance: The Emergence of Acquired 16S Ribosomal RNA Methyltransferases," *Infectious Disease Clinics of North America*. 2016. doi: 10.1016/j.idc.2016.02.011.

215. M. E. Huth *et al.*, "Designer aminoglycosides prevent cochlear hair cell loss and hearing loss," *Journal of Clinical Investigation*, 2015, doi: 10.1172/JCI77424.

216. Y. Sun, X. Jian, and Z. Deng, "Preparation and applications of a series of novel aminoglycoside compounds, and construction method of high-yield

engineering strain for producing novel aminoglycoside compounds," Chinese Patent CN106432382B, Apr. 01, 2019

217. C. M. T. Spahn and C. D. Prescott, "Throwing a spanner in the works: antibiotics and the translation apparatus," *Journal of Molecular Medicine*, vol. 74, no. 8, pp. 423–439, 1996, doi: 10.1007/BF00217518.

218. J. L. Houghton , K. D. Green, W. Chen , and S. Garneau-Tsodikova, "The Future of Aminoglycosides: The End or Renaissance?," *ChemBioChem*, vol. 11, no. 7, pp. 880–902, 2010, doi: 10.1002/cbic.200900779.

219. H. Meenavilli, R. Potumarthi, and A. Jetty, "Gentamicin production by *Micromonospora echinospora* (Me-22) in stirred tank reactor: effect of various parameters," *Journal of Basic Microbiology*, vol. 48, pp. 53–58, 2008, doi: 10.1002/jobm.200700116.

220. D. Sehnal *et al.*, "Mol* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures," *Nucleic Acids Research*, vol. 49, no. W1, pp. W431–W437, Jul. 2021, doi: 10.1093/NAR/GKAB314.

221. M. A. Borovinskaya *et al.*, "RCSB PDB - 4V53: Crystal structure of the bacterial ribosome from Escherichia coli in complex with gentamicin.," Jul. 09, 2014. https://www.rcsb.org/structure/4v53 (accessed Nov. 18, 2021).

222. P. J. Claes, R. Busson, and H. Vanderhaeghe, "Determination of the component ratio of commercial gentamicins by high-performance liquid chromatography using pre-column derivatization.," *J Chromatogr*, vol. 298, no. 3, pp. 445–57, 1984.

223. C. Kumar, M. Himabindu, and A. Jetty, "Microbial Biosynthesis and Applications of Gentamicin: A Critical Appraisal," *Critical Reviews in Biotechnology*, vol. 28, no. 3, pp. 173–212, 2008, doi: 10.1080/07388550802262197.

224. A. F. Vydrin, I. v Shikhaleev, V. L. Makhortov, N. N. Shcherenko, and N. v Kolchanova, "Component Compositions of Gentamicin Sulfate Preparations," *Phamaceutical Chemistry Journal*, vol. 37, no. 8, pp. 52–54, 2003, doi: 10.1023/A:1027372416983.

225. L. Verbist, P. J. Claes, R. Busson, and H. Vanderhaeghe, "Activity of C 2a gentamicin," *Journal of Antimicrobial Chemotherapy*, vol. 15, no. 2, pp. 250–251, 1985, doi: 10.1093/jac/15.2.250.

226. L. O. White, A. Lovering, and D. S. Reeves, "Variations in Gentamicin C1, C1a, C2, and C2a Content of Some Preparations of Gentamicin Sulphate Used Clinically as Determined by High-Performance Liquid Chromatography," *Therapeutic Drug Monitoring*, vol. 5, no. 1, pp. 123–126, 1983, doi: 10.1097/00007691-198303000-00014.

227. J. W. Park *et al.*, "Genetic dissection of the biosynthetic route to gentamicin A2 by heterologous expression of its minimal gene set.," *PNAS*, vol. 105, no. 24, pp. 8399–404, 2008, doi: 10.1073/pnas.0803164105.

228. F. Kudo and T. Eguchi, "Biosynthetic genes for aminoglycoside antibiotics," *The Journal of Antibiotics*, vol. 62, no. 9, pp. 471–481, 2009, doi: 10.1038/ja.2009.76.

229. C. Huang *et al.*, "Delineating the Biosynthesis of Gentamicin X2, the Common Precursor of the Gentamicin C Antibiotic Complex," *Chemistry & Biology*, vol. 22, no. 2, pp. 251–261, 2015, doi: 10.1016/j.chembiol.2014.12.012.

230. S. Li *et al.*, "Methyltransferases of gentamicin biosynthesis," *PNAS*, vol. 115, no. 6, pp. 1340–1345, 2018, doi: 10.1073/pnas.1711603115.

231. Y. Yu, Q. Zhang, and Z. Deng, "Parallel pathways in the biosynthesis of aminoglycoside antibiotics.," *F1000Res*, vol. 6, 2017, doi: 10.12688/f1000research.11104.1.

232. S. Li *et al.*, "Mechanistic Insights into Dideoxygenation in Gentamicin Biosynthesis," *ACS Catalysis*, vol. 11, no. 19, pp. 12274–12283, Oct. 2021, doi: doi.org/10.1021/acscatal.1c03508.

233. A. Reva, "Enzymology of gentamicin biosynthesis", PhD thesis, University of Cambridge, Cambridge, 2019, doi: 10.17863/CAM.25238.

234. D. Li, H. Li, X. Ni, H. Zhang, and H. Xia, "Construction of a gentamicin C1a-overproducing strain of *Micromonospora purpurea* by inactivation of the *gacD* gene," *Microbiological Research*, vol. 168, no. 5, pp. 263–267, 2013, doi: 10.1016/j.micres.2012.12.006.

235. H. J. Kim *et al.*, "GenK-Catalyzed C-6′ Methylation in the Biosynthesis of Gentamicin: Isolation and Characterization of a Cobalamin-Dependent Radical SAM Enzyme," *Journal of the American Chemical Society*, vol. 135, no. 22, pp. 8093–8096, 2013, doi: 10.1021/ja312641f.

236. W. Hong and L. Yan, "Identification of *gntK*, a gene required for the methylation of purpurosamine C-6′ in gentamicin biosynthesis.," *Journal of General Applied Microbiology*, vol. 58, no. 5, pp. 349–56, 2012.

237. J. Guo *et al.*, "Specificity and promiscuity at the branch point in gentamicin biosynthesis," *Chemistry and Biology*, vol. 21, no. 5, pp. 608–618, 2014, doi: 10.1016/j.chembiol.2014.03.005.

238. Y. Gu, X. Ni, J. Ren, H. Gao, D. Wang, and H. Xia, "Biosynthesis of Epimers C2 and C2a in the Gentamicin C Complex," *ChemBioChem*, vol. 16, no. 13, pp. 1933–1942, 2015, doi: 10.1002/cbic.201500258.

239. L. Shao *et al.*, "Characterization of a key aminoglycoside phosphotransferase in gentamicin biosynthesis," *Bioorganic & Medicinal Chemistry Letters*, vol. 23, no. 5, pp. 1438–1441, 2013, doi: 10.1016/j.bmcl.2012.12.064.

240. Y. H. Ban *et al.*, "Complete reconstitution of the diverse pathways of gentamicin B biosynthesis," *Nature Chemical Biology*, vol. 15, pp. 295–303, 2019, doi: 10.1038/s41589-018-0203-4.

241. Y. Chang *et al.*, "Overproduction of gentamicin B in industrial strain *Micromonospora echinospora* CCTCC M 2018898 by cloning of the missing genes *genR* and *genS*," *Metabolic Engineering Communications*, vol. 9, p. e00096, 2019, doi: 10.1016/j.mec.2019.e00096.

242. X. Ni, Z. Sun, Y. Gu, H. Cui, and H. Xia, "Assembly of a novel biosynthetic pathway for gentamicin B production in *Micromonospora echinospora*," *Microbial Cell Factories*, vol. 15, no. 1, p. 1, 2016, doi: 10.1186/s12934-015-0402-6.

243. Y-.F. Wan, R. Lian, W-.R. Hong, and X-.A. Shi, "Study on the function of resistant genes *gmrB* and *gmrA* in gentamicin biosynthetic gene cluster," *Chinese Journal of Antibiotics*, vol. 43, no. 6, pp. 688–695, 2018.

244. J. Unwin, S. Standage, D. Alexander, T. Hosted, A. C. Horan, and E. M. H. Wellington, "Gene cluster in *Micromonospora echinospora* ATCC15835 for the biosynthesis of the gentamicin C complex.," *The Journal of Antibiotics*, vol. 57, no. 7, pp. 436–45, 2004, doi: 10.7164/antibiotics.57.436.

245. E. Severi and G. H. Thomas, "Antibiotic export: Transporters involved in the final step of natural product production," *Microbiology (United Kingdom)*, vol. 165, no. 8, pp. 805–818, 2019, doi: 10.1099/MIC.0.000794

246. W. Piepersberg, K. M. Aboshanab, H. Schmidt-Beißner, and U. F. Wehmeier, "The Biochemistry and Genetics of Aminoglycoside Producers," in *Aminoglycoside Antibiotics: From Chemical Biology to Drug Discovery*, D. P. Arya, Ed. John Wiley & Sons, Ltd, 2007, pp. 15–118. doi: 10.1002/9780470149676.CH2.

247. E. Selimoglu, "Aminoglycoside-induced ototoxicity.," *Curr Pharm Des*, vol. 13, no. 1, pp. 119–26, 2007.

248. C. Chen, Y. Chen, P. Wu, and B. Chen, "Update on new medicinal applications of gentamicin: Evidence-based review," *Journal of the Formosan Medical Association*, vol. 113, no. 2, pp. 72–82, 2014, doi: 10.1016/J.JFMA.2013.10.002.

249. S. J. Kohlhepp, M. O. Loveless, P. W. Kohnen, D. C. Houghton, W. M. Bennett, and D. N. Gilbert, "Nephrotoxicity of the constituents of the gentamicin complex.," *The Journal of Infectious Diseases*, vol. 149, no. 4, pp. 605–14, 1984, doi: 10.1093/infdis/149.4.605

250. R. M. Sandoval *et al.*, "A non-nephrotoxic gentamicin congener that retains antimicrobial efficacy.," *J Am Soc Nephrol*, vol. 17, no. 10, pp. 2697–705, Oct. 2006, doi: 10.1681/ASN.2005101124.

251. R. M. Ahmed, I. P. Hannigan, H. G. MacDougall, R. C. Chan, and G. Michael Halmagyi, "Gentamicin ototoxicity: A 23-year selected case series of 103 patients," *Medical Journal of Australia*, 2012, doi: 10.5694/mja11.10850.

252. M. Kobayashi, M. Sone, M. Umemura, T. Nabeshima, T. Nakashima, and S. Hellström, "Comparisons of cochleotoxicity among three gentamicin compounds following intratympanic application," *Acta Oto-Laryngologica*, vol. 128, no. 3, pp. 245–249, 2008, doi: 10.1080/00016480701558948.

253. M. Ishikawa *et al.*, "Lower ototoxicity and absence of hidden hearing loss point to gentamicin C1a and apramycin as promising antibiotics for clinical use," *Scientific Reports*, vol. 9, no. 1, p. 2410, 2019, doi: 10.1038/s41598-019-38634-3.

254. M. E. O'Sullivan *et al.*, "Dissociating antibacterial from ototoxic effects of gentamicin C-subtypes," *PNAS*, vol. 117, no. 51, pp. 32423–32432, 2020, doi: 10.1073/pnas.2013065117.

255. T. W. Glinka and O. Rodny, "Aminoglycoside derivatives." Google Patents, 2011.

256. R. R. Breaker, "Riboswitches and the RNA World," *Cold Spring Harbor Perspectives in Biology*, vol. 4, no. 2, pp. 3566–3567, 2012, doi: 10.1101/CSHPERSPECT.A003566.

257. B. JE and B. RR, "The distributions, mechanisms, and structures of metabolite-binding riboswitches," *Genome Biology*, vol. 8, no. 11, Nov. 2007, doi: 10.1186/GB-2007-8-11-R239.

258. M. J. Cromie, Y. Shi, T. Latifi, and E. A. Groisman, "An RNA sensor for intracellular $Mg^{(2+)}$," *Cell*, vol. 125, no. 1, pp. 71–84, 2006, doi: 10.1016/J.CELL.2006.01.043.

259. B. Klinkert and F. Narberhaus, "Microbial thermosensors," *Cellular and Molecular Life Sciences*, vol. 66, no. 16, pp. 2661–2676, 2009, doi: 10.1007/S00018-009-0041-3.

260. C. E. Dann, C. A. Wakeman, C. L. Sieling, S. C. Baker, I. Irnov, and W. C. Winkler, "Structure and mechanism of a metal-sensing regulatory RNA," *Cell*, vol. 130, no. 5, pp. 878–892, 2007, doi: 10.1016/J.CELL.2007.06.051.

261. J. Johansson, "RNA thermosensors in bacterial pathogens," *Contrib Microbiol*, vol. 16, pp. 150–160, 2009, doi: 10.1159/000219378.

262. P. J. McCown, K. A. Corbino, S. Stav, M. E. Sherlock, and R. R. Breaker, "Riboswitch diversity and distribution," *RNA*, vol. 23, no. 7, pp. 995–1011, 2017, doi: 10.1261/RNA.061234.117.

263. R. M. Atilho, G. Mirihana Arachchilage, E. B. Greenlee, K. M. Knecht, and R. R. Breaker, "A bacterial riboswitch class for the thiamin precursor HMP-PP employs a terminator-embedded aptamer", eLife, vol. 8, no. e45210, 2019, doi: 10.7554/eLife.45210.001.

264. M. M. Rudolph, M. P. Vockenhuber, and B. Suess, "Synthetic riboswitches for the conditional control of gene expression in *Streptomyces coelicolor*," *Microbiology (Reading)*, vol. 159, no. Pt 7, pp. 1416–1422, 2013, doi: 10.1099/MIC.0.067322-0.

265. S. Topp *et al.*, "Synthetic riboswitches that induce gene expression in diverse bacterial species," *Applied Environmental Microbiology*, vol. 76, no. 23, pp. 7881–7884, 2010, doi: 10.1128/AEM.01537-10.

266. A. T. Ma, C. M. Schmidt, and J. W. Golden, "Regulation of gene expression in diverse cyanobacterial species by using theophylline-responsive riboswitches," *Applied Environmental Microbiology*, vol. 80, no. 21, pp. 6704–6713, 2014, doi: 10.1128/AEM.01697-14.

267. J. C. Seeliger, S. Topp, K. M. Sogi, M. L. Previti, J. P. Gallivan, and C. R. Bertozzi, "A riboswitch-based inducible gene expression system for mycobacteria," *PLoS One*, vol. 7, no. 1, 2012, doi: 10.1371/JOURNAL.PONE.0029266.

268. J. Born, K. Weitzel, B. Suess, and F. Pfeifer, "A Synthetic Riboswitch to Regulate Haloarchaeal Gene Expression," *Frontiers in Microbiology*, vol. 12, 2021, doi: 10.3389/fmicb.2021.696181.

269. S. Jang *et al.*, "Development of Artificial Riboswitches for Monitoring of Naringenin in Vivo," *ACS Synthetic Biology*, vol. 6, no. 11, pp. 2077–2085, 2017, doi: 10.1021/ACSSYNBIO.7B00128.

270. S. Jang, S. Jang, D. K. Im, T. J. Kang, M. K. Oh, and G. Y. Jung, "Artificial Caprolactam-Specific Riboswitch as an Intracellular Metabolite Sensor," *ACS Synthetic Biology*, vol. 8, no. 6, pp. 1276–1283, 2019, doi: 10.1021/ACSSYNBIO.8B00452/.

271. C. Berens, A. Thain, and R. Schroeder, "A tetracycline-binding RNA aptamer," *Bioorganic & Medicinal Chemistry*, vol. 9, no. 10, pp. 2549–2556, 2001, doi: 10.1016/S0968-0896(01)00063-3.

272. F. Groher *et al.*, "Riboswitching with ciprofloxacin-development and characterization of a novel RNA regulator," *Nucleic Acids Research*, vol. 46, no. 4, pp. 2121–2132, 2018, doi: 10.1093/NAR/GKX1319.

273. A. Boussebayle *et al.*, "Next-level riboswitch development-implementation of Capture-SELEX facilitates identification of a new synthetic riboswitch," *Nucleic Acids Research*, vol. 47, no. 9, pp. 4883–4895, 2019, doi: 10.1093/nar/gkz216.

274. J. E. Weigand *et al.*, "Mechanistic insights into an engineered riboswitch: a switching element which confers riboswitch activity," *Nucleic Acids Research*, vol. 39, no. 8, pp. 3363–3372, 2011, doi: 10.1093/NAR/GKQ946.

275. C. Zhu, G. Yang, M. Ghulam, L. Li, and F. Qu, "Evolution of multi-functional capillary electrophoresis for high-efficiency selection of aptamers," *Biotechnology Advances*, vol. 37, no. 8, Dec. 2019, doi: 10.1016/J.BIOTECHADV.2019.107432.

276. G. S. Hossain, M. Saini, R. Miyake, H. Ling, and M. W. Chang, "Genetic Biosensor Design for Natural Product Biosynthesis in Microorganisms," *Trends in Biotechnology*, vol. 38, no. 7, pp. 797–810, Jul. 2020, doi: 10.1016/J.TIBTECH.2020.03.013.

2. Author contributions for Chapter 2

Katherine V. Baker was responsible for the *in silico* analysis of *M.* sp. DEM32671, characterisation of antibiotic production from *Streptomyces* hosts carrying Cluster 24, generation of resistant *Streptomyces* strains, characterisation of resistant strains via antibiotic susceptibility tests, sample preparation for whole-genome sequencing, analysis of whole-genome sequencing data, *Streptomyces* culturing, antibiotic bioassays, LC-MS sample preparation and data interpretation.

Nick Allenby provided the genome sequence and strain of *Micromonospora* sp. DEM32671. Victoria Jackson was responsible for the early genome sequencing analysis of *M.* sp. DEM32671, and the insertion of Cluster 24 into *Streptomyces* hosts (detailed in section 2.3.1., content in paragraph "*Micromonospora* sp. DEM32671 … 88.5% of the genome (Figure 2.2)."; 2.5.3 and 2.5.5.)

Eriko Takano and Olga Genilloud oversaw the collaboration between Medina and the University of Manchester. Jesús Martín Serrano carried out all running of LC-MS samples, and data interpretation (detailed in sections 2.3.8 and 2.5.14). Jesús Martín Serrano and Francesco Del Carratore carried out LC-MS data processing (detailed in sections 2.3.8 and 2.5.14, information used to construct Figures 2.19, 2.20, 2.21, 2.22).

2. Establishing *M.* sp. DEM32671 as a producer of aminoglycosides through heterologous expression in *Streptomyces* spp.

Katherine V. Baker,[1] Victoria Jackson, [1] Jesús Martín Serrano,[2] Francesco del Carratore,[1] Olga Genilloud,[2] Nicholas E. E. Allenby,[3] Rainer Breitling,[1] Eriko Takano[1]

[1] Manchester Institute of Biotechnology, Manchester Synthetic Biology Research Centre SYNBIOCHEM, Department of Chemistry, The University of Manchester, Manchester, M1 7DN, United Kingdom

[2] Fundación Medina, Av. Conocimiento, Parque Tecnológico de Ciencias de la Salud, 34 - 18016 Granada, Spain

[3] Demuris Limited, The Biosphere, Draymans Way, Newcastle Helix, Newcastle upon Tyne NE4 5BX, United Kingdom

2.1. Abstract

As the antibiotic crisis worsens, the diversification of our antibiotic arsenal is becoming more and more important. Synthetic biology has given many options for antibiotic production from less amenable microorganisms, with heterologous expression of entire gene clusters being a common technique to awaken production. However, on many occasions, a simple transplantation of a gene cluster into a heterologous host requires further fine-tuning, for example by removing regulatory systems or insertion of heterologous-host specific sequences to allow proper genetic function. In this work we detail a pipeline towards improvement of aminoglycoside yield in heterologous host *Streptomyces coelicolor*. We inserted Cluster 24 from *Micromonospora* sp. DEM32671, predicted *in silico* to produce gentamicin, into *S. coelicolor* M1146. After firstly seeing low antibiotic yield from liquid cultures, we generated more resistant strains via exposure to increasing concentrations of gentamicin sulfate, producing a strain that was resistant to a 5-fold increase of gentamicin than the parental strain. This approach led to a yield improvement allowing for detection of gentamicin intermediates and an interesting, predicted aminoglycoside of mass 501.2246. This provides a platform for further aminoglycoside production in *S. coelicolor*.

2.2. Introduction

The resistance of microorganisms to clinically used antibiotics remains a growing problem and represents one of the largest global health challenges faced by current generations.[1] Currently, most paths towards treating antibiotic resistant infections are focused on i) generation of new antibiotic scaffolds, either through high-throughput discovery efforts underpinned by *in silico* advances, or semi-synthesis ii) alternative therapies such as phage therapies and immunotherapies.[2] Though these avenues are key in ensuring we keep ahead of antibiotic resistance, diversifying our current collection of antibiotics will be essential in the coming years.

The genus *Micromonospora* was first described by Ørskov in 1923 and has since grown to include 110 different *Micromonospora* spp. with valid published names in the LPSN database (https://www.bacterio.net/genus/*Micromonospora*).[3,4] They are Gram-positive Actinobacteria which have been discovered in many diverse environments, including the soil,[5–7] plant tissues,[8,9] aquatic environments (with extreme environments including deep-sea sediments and hypersaline water, and from aquatic organisms such as sponges),[10–13] and rock and soil samples collected through mining.[14–16] *Micromonospora* as a genus are regarded as being perhaps the second most rich in natural products, with *Streptomyces* being the first. They are well known for their production of aminoglycoside antibiotics, especially gentamicin.[17] As of 2005, 740 bioactive secondary metabolites isolated from *Micromonospora* had been identified.[18] With the exception of *Micromonospora echinospora* for gentamicin and *Micromonospora* spp. for production of vitamin B12, there are few examples of *Micromonospora* being used for industrial-scale fermentation.[19,20] There is a history of *Micromonospora* secondary metabolites being heterologously produced in *Streptomyces* species: intermediate compound gentamicin A2 was previously produced by *Streptomyces venezuelae*,[21] antitumor thiocoraline from *Micromonospora* sp. ACM2 and *M.* sp. ML1 was produced by *Streptomyces albus*,[22] and fluostatin derivatives from *Micromonospora rosaria* SCSION160 were produced by

Streptomyces coelicolor.[23] There are many reasons to do this, namely characterisation of biosynthetic pathways, awakening silent gene clusters, or to achieve a higher yield of production of the metabolite of interest.[24] The wider array of tools for genetic manipulation available for Streptomycetes make them a more amenable host in some cases where an intermediate or by-product of a gene cluster may be the favoured product, where native hosts can be slow growing, or where the compound of interest is not produced in high enough yields.[24]

Advances in the field of synthetic biology have allowed for the improved production of many secondary metabolites, including antibiotics.[25] Recent examples include the production of Type IV-V glycopeptide antibiotics in *S. coelicolor* by generation of a specialised host strain carrying precursor biosynthesis genes,[26] and the production of bioactive, previously cryptic streptophenazines in heterologous host *S. coelicolor* by cluster refactoring.[27] As our tools and approaches for genetic engineering increase in variety, and are adapted for an increasing number of strains, standard heterologous expression experiments can be more easily further improved upon.

Whilst it remains an approach more commonly used for producing the endogenous secondary metabolites from bacterial strains, introducing mutations into a host strain may aid in increasing production yield of a compound of interest. Estévez et al. have recently described an approach for increasing the yield of a fredericamycin through cumulative mutations to *Streptomyces albus* subsp. *chlorinus* by exposure to streptomycin.[28] This antibiotic-exposure approach has additionally been used multiple times in *Streptomyces diastatochromogenes*, where Fan et al. used streptomycin as a pressure for increasing tetraene macrolides (between 8.7 and 25-fold increases concentrations of tetramycin A/P and tetrin B in the strain carrying five mutations) and Shentu et al. used multiple antibiotics to increase yields of fungicide toyocamycin (24-fold higher than the wild type strain).[29,30] Wang et al. also used multiple antibiotics for mutagenesis, and increased yields of ε -

poly-L-lysine by *Streptomyces albulus* 2.79-fold higher than the parental strain.[31] Despite a diversity in structure, aminoglycosides such as streptomycin function generally to bind the 16S rRNA to cause misreading of tRNAs, and can also block or cause error-prone amino acid translocation directly **(Figure 2.1).**[32–42] Therefore, evolving resistance by mutation of ribosomal components can have knock-on effects leading to overproduction of antibiotics.[43] These works highlight that mutation in ribosomal and RNA polymerase components is a viable strategy towards overproduction of useful secondary metabolites, and may also solve one of the many challenges of producing antimicrobials by a bacterial host: the potential for toxicity to the heterologous host itself.

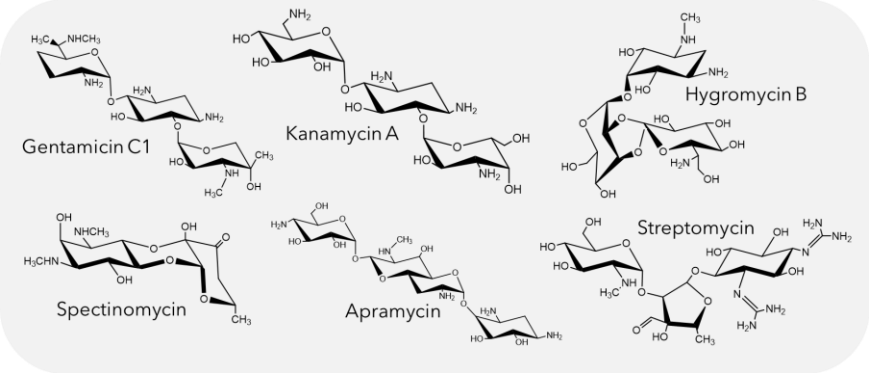


Figure 2.1: Structures of selected aminoglycoside and aminocyclitol antibiotics mentioned in this work.

Here we describe strain *M.* sp. DEM32671 as a novel *Micromonospora* strain identified in the Demuris actinomycete strain library. In this work, we describe a pipeline from initial genome sequencing and phylogenetic analysis of *M.* sp. DEM32671 towards heterologous production of a predicted aminoglycoside from a target biosynthetic gene cluster (BGC). To improve upon first production tests, a cohort of resistant *Streptomyces* strains was generated, and the potential resistance mechanisms investigated via whole-genome sequencing. The resultant awakening of Cluster 24 in a resistant production host allowed for first characterisation of compound from *M.* sp. DEM32671 Cluster 24. This work provides a foundation for production of aminoglycosides in *Streptomyces* heterologous hosts.

2.3. Results and Discussion

2.3.1. Placing *Micromonospora* sp. DEM32671 in the genus *Micromonospora*

Micromonospora sp. DEM32671 was obtained from marine sediment and later catalogued as part of the Demuris actinomycete strain library. After isolates from early experiments showed antibiotic bioactivity against *Bacillus subtilis* (Nick Allenby, personal communication) the strain was determined to be an interesting candidate for production of novel antimicrobials. As the strain showed slow growth in liquid cultures, the aim was to carry out full-genome sequencing to determine which gene clusters were present and to express these clusters in a heterologous host. A draft genome sequence of strain *M.* sp. DEM32671 was obtained prior to this work using PacBio SMRT (Nick Allenby, unpublished). The draft genome consisted of a 7.52 Mbp circular chromosome with a G+C content of 72.62%. 6363 coding sequences were predicted, making up 88.5% of the genome **(Figure 2.2)**. Three larger regions were present, marked with asterisks, where G+C content is below average (reaching as low as 65% G+C, ~7% below average). This could be indicative of regions picked up through horizontal gene transfer from species with lower average G+C%.

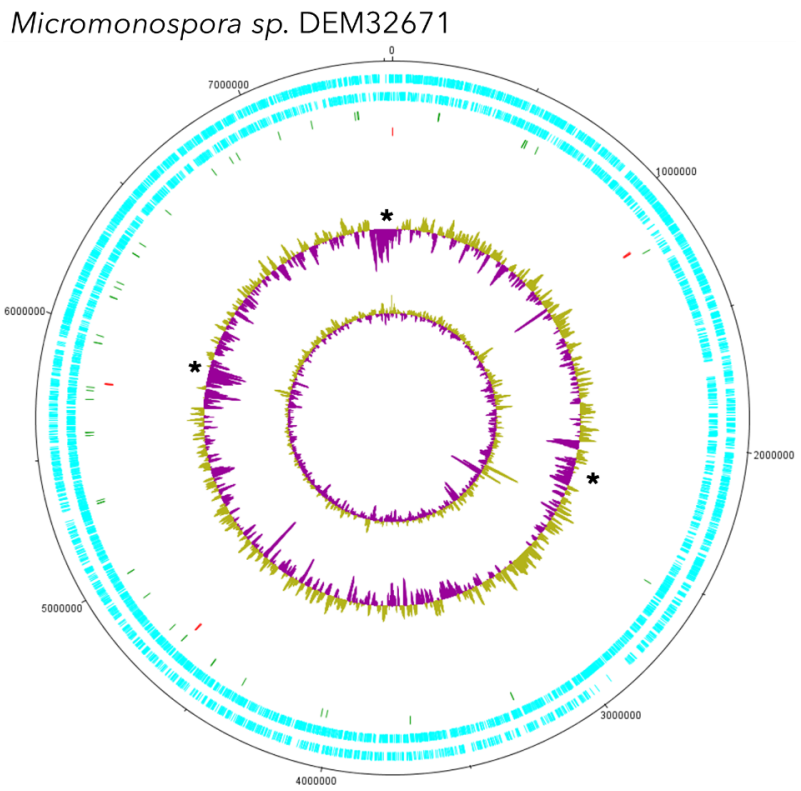


Figure 2.2: Map of Micromonospora sp. DEM32671 genome. Tracks from outside in: CDS (blue) i) forward strand & ii) reverse strand; iii) tRNA (green); iv) rRNA (red); v) G+C % & vi) G+C Skew (both green above average, purple below average). Asterisks denote large areas of low G+C content.

For first estimation of taxonomic placement, the 16S rRNA sequence from *M.* sp. DEM32671 was used as BLAST query sequence to determine closest relatives. The most similar sequence identified was the 16S rRNA sequence from *Micromonospora echinospora* DSM 43816 (99.66% similarity), the strain of the genus perhaps most well-known due to being used for industrial production of gentamicin. This species became a comparison point throughout the analyses due to high similarity, aiming to determine whether *M.* sp. DEM32671 is sufficiently different to be classed as a novel species. Other aminoglycoside producers with 16S rRNA sequences in the top 100 hits include type strains of *Micromonospora sagamiensis* (99.39%; sagamicin producer), *Micromonospora pallida* (99.04%, gentamicin producer) and *Micromonospora inyonensis* (98.84%, sisomicin producer; also known as *M. inyoensis;* **Table 2.1**).

Table 2.1: Comparison of percentage identity of 16S rRNA sequences from *M.* sp. DEM32671 to aminoglycoside-producing *Micromonospora* type strains.

Aminoglycoside-producing *Micromonospora* (type strain)	Aminoglycoside produced	% identity to 16S rRNA sequence from *M.* sp. DEM32671
M. echinospora DSM 43816	Gentamicin	99.66
M. sagamiensis JCM 3310	Sagamicin	99.39
M. pallida DSM 43816	Gentamicin	99.04
M. inyonensis DSM 46123	Sisomicin	98.84
M. olivasterospora DSM 43868	Fortimicin	97.96

These results place *M.* sp. DEM32671 firmly in the genus *Micromonospora*. This genus is well known for the 16S rRNA sequences being highly similar: the top 100 hits in the BLAST repository to the 16S sequence of *M.* sp. DEM32671 (upon analysis 11/08/21) yielded 98 *Micromonospora* 16S rRNA sequences with percentage identity scores between 99.66% and 98.84% **(Table S2.1)**. As the 16S rRNA sequences in this group do not diverge enough for accurate determination of relatedness of each species, it has been recommended that use of the sequence of DNA gyrase subunit B (*gyrB)* is a better measure for creation of a more informative phylogenetic tree. In addition, *gyrB* sequence similarity has been shown to agree with DNA-DNA hybridization studies.[44] The percentage identity of *M.* sp. DEM32671 full-length *gyrB* to *M. echinospora* DSM 43816 was 97.01%, *M. inyonensis* 96.10% and *M. pallida* 95.69%. To confirm the closest relatives of *M.* sp. DEM32671 in the genus *Micromonospora*, the partial *gyrB* and 16S rRNA sequences from 50 *Micromonospora* type strains were extracted from NCBI and aligned using the Molecular Evolutionary Genetics Analysis tool (MEGA). The alignments were truncated at 1200 bp and 1465 bp respectively to ensure that difference in sequence length would not impact the grouping of the strains. Maximum Likelihood phylogenetic trees were constructed from these alignments also using MEGA and visualisations drawn using iTOL with some manual annotation **(Figure 2.3)**. As expected, when looking at *gyrB* sequence *M.* sp. DEM32671 was found to be in a clade with producers of gentamicin and gentamicin-derived aminoglycosides.

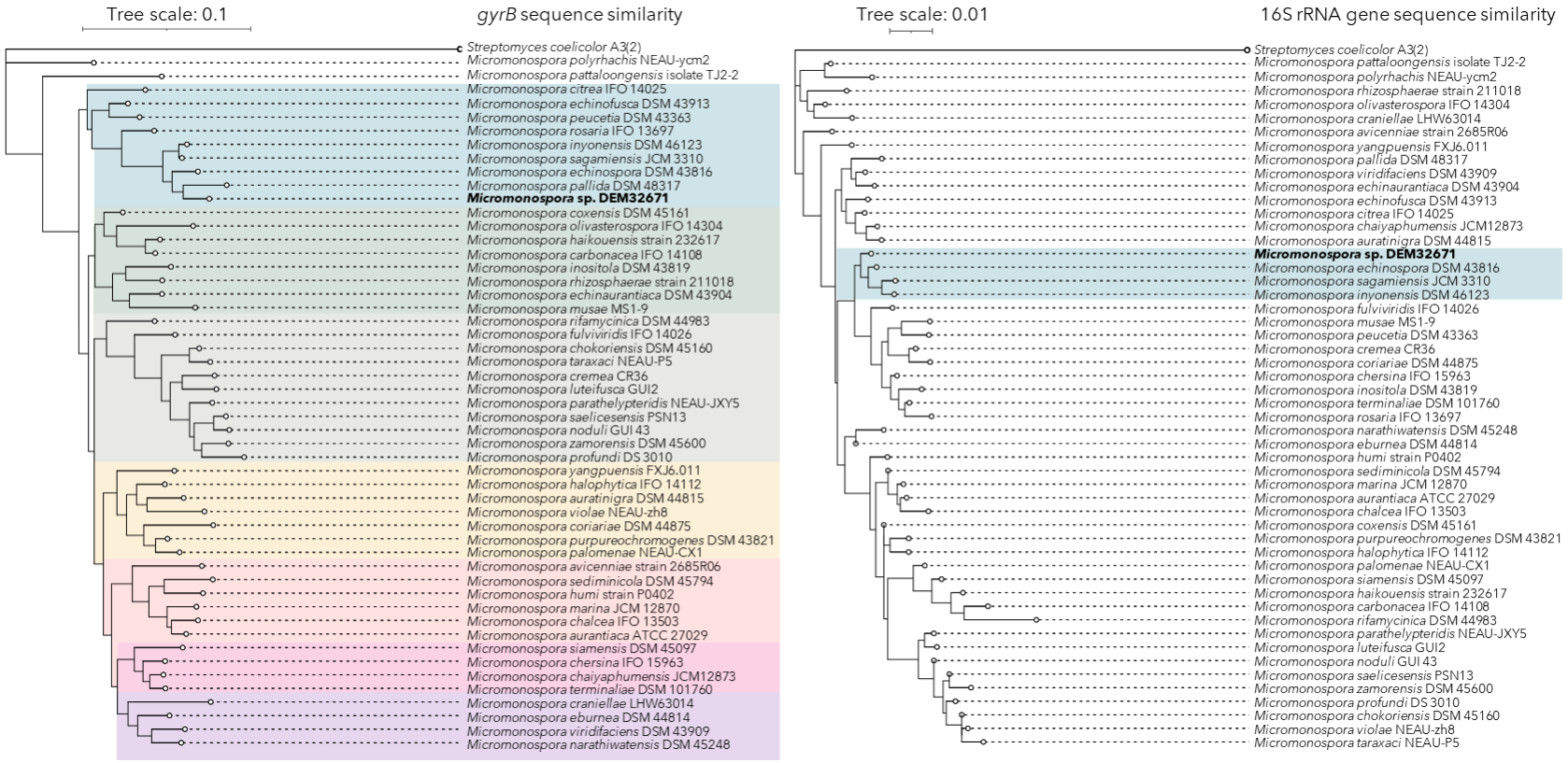


Figure 2.3: Maximum-likelihood phylogenetic trees constructed to determine strain lineage of *M.* sp. DEM32671 based on partial *gyrB* (left; truncated at 1200 bp) and 16S rRNA (right; truncated at 1465 bp) gene sequences. Scale bar is representative of the mean number of nucleotide substitutions per site. Clades for *gyrB* are highlighted in colour; clade for 16S rRNA corresponding to group of M. sp. DEM32671 is highlighted in blue.

2.3.1.1 Comparing *Micromonospora* sp. DEM32671 with type strain *Micromonospora echinospora* DSM 43816

Throughout the genomic analysis of *M.* sp. DEM32671, it became clear that it had a high similarity to nearest neighbour *Micromonospora echinospora* DSM 43816, and so determining whether it is novel enough to be labelled as its own species became challenging. Phenotypically, grown on Medium 554 (N-Z-amine medium; Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ)) the strains are very similar **(Figure 2.4)**, with purplish-black circular convex colonies at low density and wrinkles on areas of the solid medium where cell density is high. *M.* sp. DEM32671 does not show purple diffusible pigment production and orange 'fringing', shown on the areas of high growth in the photograph of *M. echinospora* DSM 43816. It is important to note that a sample of type strain *M. echinospora* DSM 43816 was not able to be obtained for proper phenotypic comparison with equivalent growth conditions; rather, *M.* sp. DEM32671 was grown in the conditions described by the DSMZ.

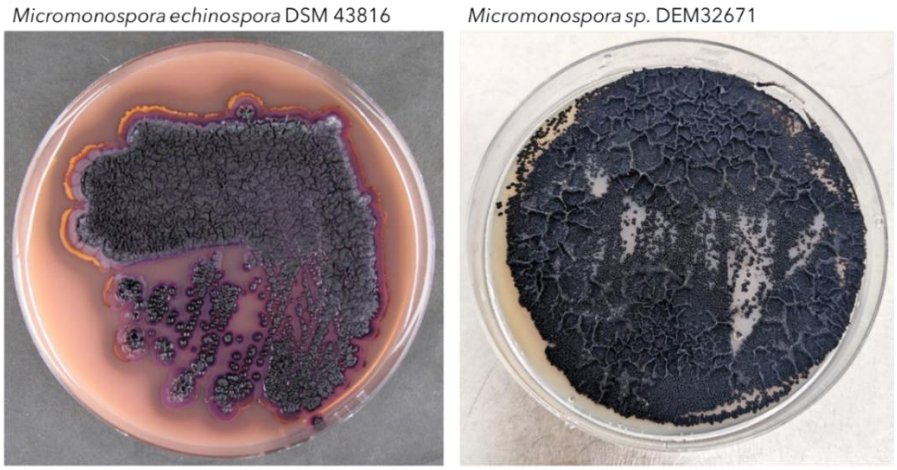


Figure 2.4: Phenotypic comparison of *M. echinospora* DSM 43816 and *M.* sp. DEM32671 grown on Medium 554 (DSMZ). *M.* sp. DEM32671 was grown in liquid culture for 3 weeks before spreading a lawn plate and growing for an additional 2 weeks. Photograph of *M. echinospora* DSM 43816 was retrieved from DSMZ (© Leibniz-Institut DSMZ) under an Attribution-Non Commercial 4.0 International License.[45]

As already stated, the 16S rRNA sequences of the two strains were nearly identical, thus requiring further investigation to determine whether *M.* sp. DEM32671 can be classed as a novel species.[46] The genome size and G+C content was also highly similar (7.78 Mbp vs 7.52 Mbp; 72.3% vs 72.6%). G+C content within species does not usually vary more than 1%.[47]

The Type (Strain) Genome Server (TYGS) was used as an *in silico* mimic of classical DNA-DNA hybridisation (digital DNA-DNA hybridisation; dDDH) to predict how related two bacteria are and therefore measure likelihood that *M.* sp. DEM32671 is a novel species.[48] After aligning the genome with all other *Micromonospora* type strains, 16 were deemed to be similar. *S. coelicolor* A3(2) was included as an outgroup for creation of phylogenetic trees with this set of genomes. The scores are presented in **Table 2.2**. *M.* sp. DEM32671 showed whole genome similarity strongest to the two *M. echinospora* strains available (DSM 43816 and ATCC 15837), in agreement with previous results. A dDDH value of >70% indicates that the two genome sequences likely belong to the same species group, with the d4 measurement usually being taken as it performs best of the three calculations with a dataset where genome sizes differ. The d4 measurement shows a dDDH % of 62.1/62% between *M.* sp. DEM32671 and both *M. echinospora* species respectively, with confidence intervals lower than the 70% cut-off point. However, as both genome sequences being compared are intact, the more relevant statistic may be the d6 dDDH value (72.1/72%) which indicates that the two strains may be the same species.

To further investigate the place of *M.* sp. DEM32671 amongst this group of *Micromonospora* and to attempt to clarify whether the strain can be classed as a new species, a whole-genome OrthoANI (Orthologous Average Nucleotide Identity) analysis was carried out on the ten strains identified by TYGS using OAT **(Figure 2.5)**.[49] OrthoANI fragments the genomes being compared into equally sized sections and then compares those sequences that are orthologous, providing a percentage similarity for the entire genome. The

authors of this work recommend that species demarcation be at values of approximately 95-96% similarity. *M.* sp. DEM32671 shares highest OrthoANI scores with *M. echinospora* DSM 43816 (95.42%), aligning with previous results. Interestingly, *M. inyonensis* DSM 46123 and *M. sagamiensis* JCM 3310 show a higher similarity score of 96.51% but have been defined as separate species.

Table 2.2: digital DNA-DNA hybridisation values for ten most similar strains to *M. sp.* DEM32671 identified by TYGS. *S. coelicolor* was included manually as an outgroup.

Subject strain in comparison to *M.* sp. DEM32671	dDDH (d0; %)	C.I. (d0; %)	dDDH (d4; %)	C.I. (d4; %)	dDDH (d6; %)	C.I. (d6; %)
M. echinospora DSM 43816	71.6	[67.7 - 75.3]	62.1	[59.2 - 64.9]	72.1	[68.6 - 75.3]
M. echinospora ATCC 15837	71.6	[67.7 - 75.3]	62	[59.1 - 64.8]	72	[68.6 - 75.2]
M. sagamiensis JCM 3310	59	[55.4 - 62.5]	42.3	[39.8 - 44.8]	55.8	[52.6 - 58.9]
M. sagamiensis DSM 43912	58.9	[55.3 - 62.4]	42.3	[39.8 - 44.8]	55.7	[52.5 - 58.8]
M. inyonensis DSM 46123	40.8	[37.5 - 44.3]	41.9	[39.4 - 44.4]	40.3	[37.3 - 43.3]
M. pallida DSM 43817	48.6	[45.2 - 52.1]	37.8	[35.4 - 40.4]	45.9	[42.9 - 49.0]
M. citrea DSM 43903	27.6	[24.2 - 31.2]	26.4	[24.0 - 28.9]	26.2	[23.3 - 29.3]
M. rubida NEAU-HG-1	27.9	[24.5 - 31.5]	25.9	[23.6 - 28.4]	26.3	[23.4 - 29.4]
M. fulviviridis JCM 3259	26.6	[23.2 - 30.2]	25.6	[23.3 - 28.1]	25.2	[22.3 - 28.3]
M. acroterricola 5R2A7	25.9	[22.5 - 29.5]	25.6	[23.3 - 28.1]	24.7	[21.8 - 27.8]
M. chersina DSM 44151	26.5	[23.2 - 30.2]	25.5	[23.1 - 27.9]	25.2	[22.3 - 28.3]
M. chaiyaphumensis DSM 45246	27	[23.6 - 30.6]	25.5	[23.2 - 28.0]	25.5	[22.6 - 28.6]
M. haikouensis DSM 45626	28.3	[25.0 - 32.0]	25.5	[23.2 - 28.0]	26.6	[23.7 - 29.7]
M. endolithica JCM 12677	26.1	[22.8 - 29.7]	25.2	[22.9 - 27.7]	24.8	[21.9 - 27.9]
S. coelicolor DSM 40233	13.3	[10.6 - 16.6]	18.1	[16.0 - 20.5]	13.6	[11.2 - 16.4]

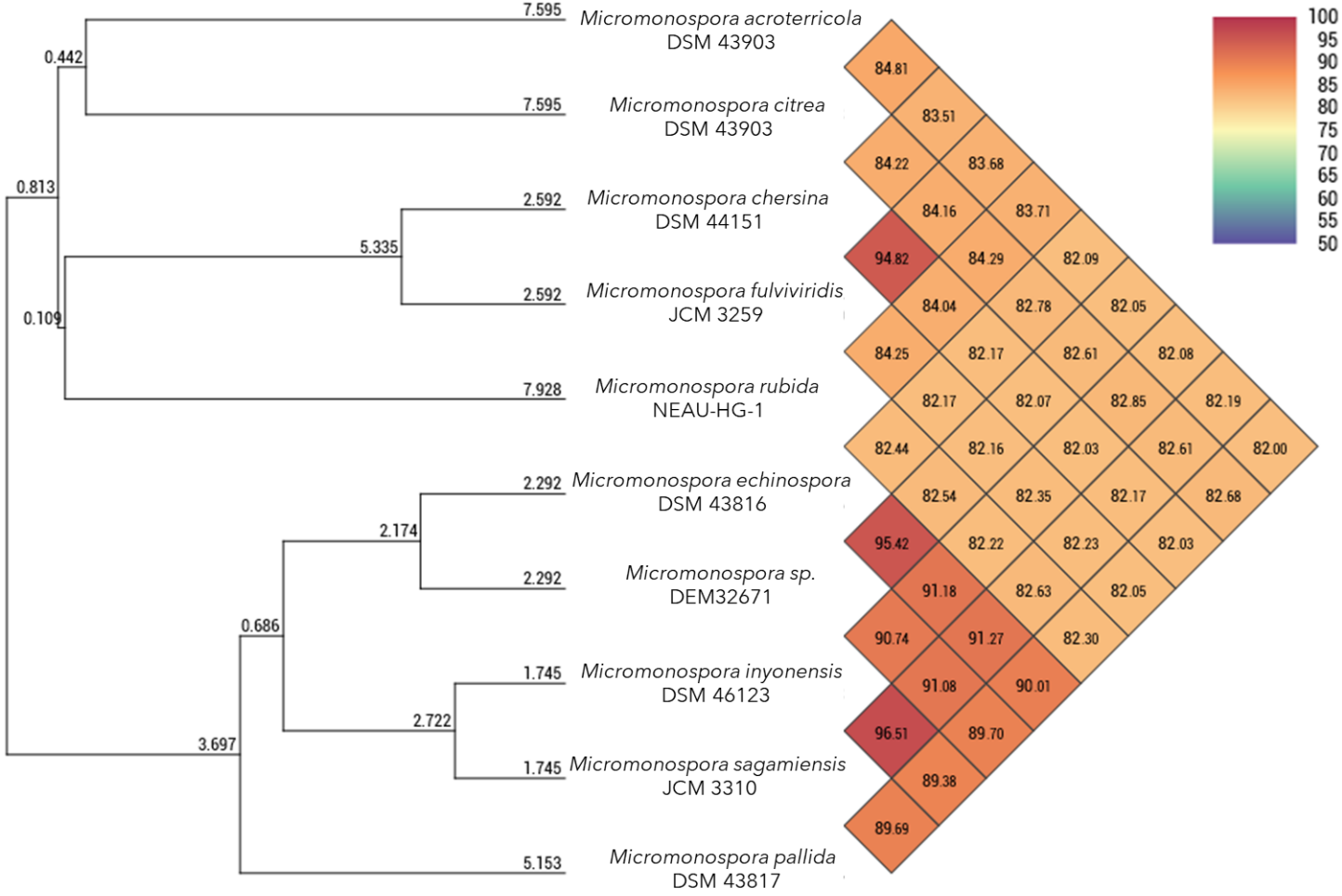


Figure 2.5: OrthoANI analysis of 10 nearest neighbour strains to *M.* sp. DEM32671 identified by TYGS.[48] Orthologous average nucleotide identity as a percentage between two strains is presented as a heatmap. Phylogenetic tree is calculated automatically based on orthologous average nucleotide identity.

Recent work has shown the value of prospecting new strains for natural products, especially those in marine and extreme environments.[11,50–52] It is currently estimated that there is a huge untapped microbial population which has previously been inaccessible to us, which may yield new antibiotic scaffolds once culture conditions are elucidated.[53] Whilst the prospective BGCs from *M.* sp. DEM32671 are of value, further characterisation should be carried out to define whether *M.* sp. DEM32671 is sufficiently different from *M. echinospora* for it to be classified as a novel species. Investigation of the fatty acid composition or carbon source utilization would add to the evidence to suggest novelty but were outside the scope of this work.

2.3.2. *Micromonospora* sp. DEM32671 is predicted to have aminoglycoside production capabilities

Analysis of the draft genome of *M.* sp. DEM32671 with antiSMASH 6.0 using relaxed cluster settings yielded 31 predicted BGCs **(Table S2.2)**.[54] The majority of the natural product gene clusters identified with antiSMASH were non-ribosomal peptide synthase- or polyketide synthase-like clusters, or hybrids of these two. Other peptide gene clusters such as those for ribosomally synthesized and post-translationally modified peptides or lanthipeptides were also present, as well as three terpene gene clusters. Clusters predicted to produce TLN-05220 and isorenieratene are also present, likely contributing to the pigmentation of the strain; carotenoids and other coloured compounds are often produced by *Micromonospora* and cause a wide range of colouration amongst the family.[55–57]

A comparison of the gene clusters of *M.* sp. DEM32671 with nearest phylogenetic neighbours is presented as **Figure 2.6**. *M.* sp. DEM32671 is predicted to be one of the most biosynthetically-talented of the strains with 31 gene clusters predicted, though the other strains have similar numbers of clusters predicted (29 for all others except *M. sagamiensis* JCM 3310, which also

has 31). Most importantly, each of these strains is predicted to produce an aminoglycoside and for all but *M.* sp. DEM32671, the specific aminoglycoside has been confirmed with fermentation: *M. echinospora* and *M. pallida* producing gentamicin, *M. sagamiensis* producing sagamicin and *M. inyonensis* producing sisomicin.[19,44,58,59]

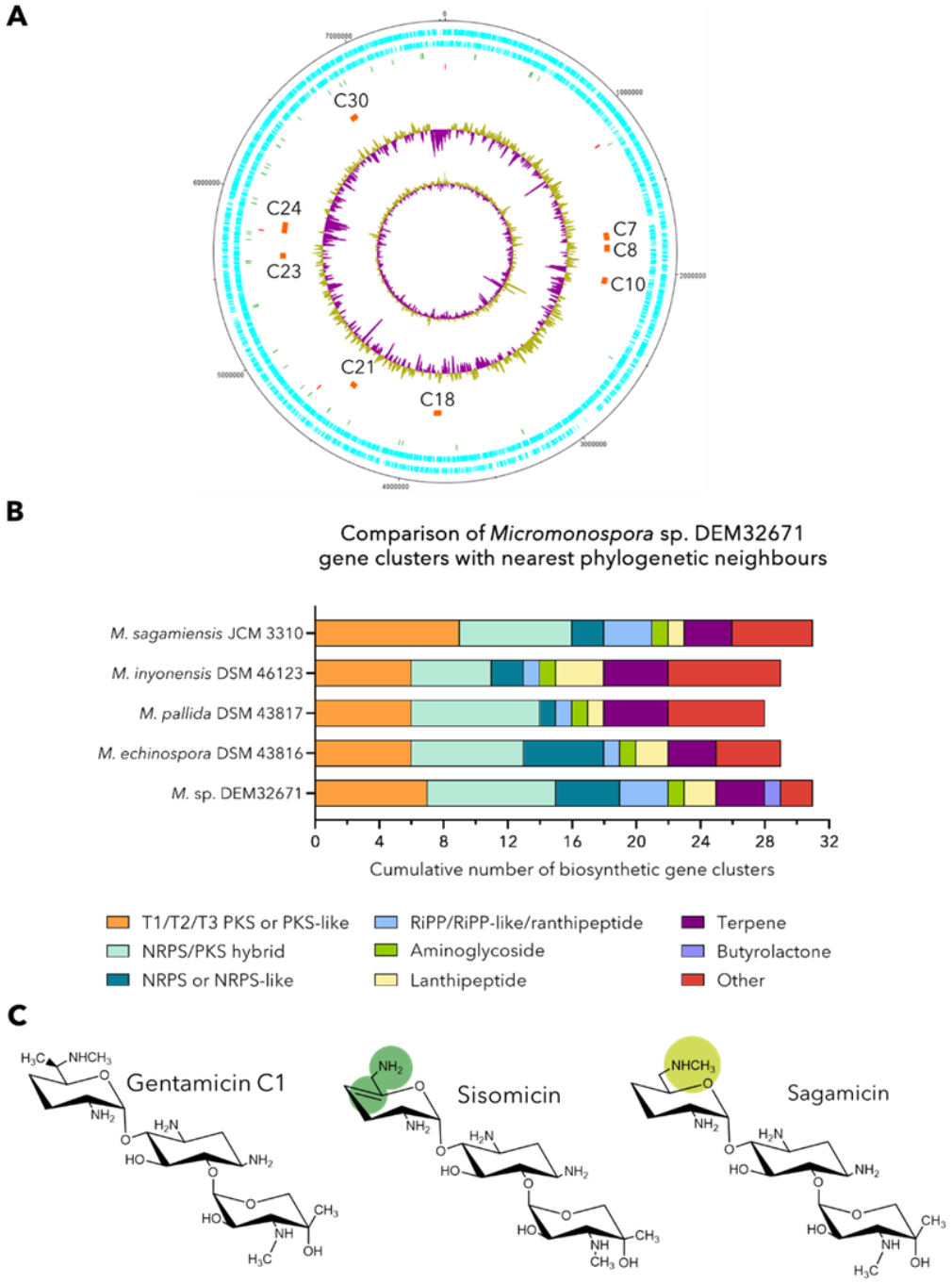


Figure 2.6: Biosynthetic gene cluster diversity in *M.* sp. DEM32671 in comparison to four nearest neighbour strains. A – Location of key BGCs in the genome of *M.* sp. DEM32671. Full list of gene clusters is provided as **Table S2.2. B –** AntiSMASH 6.0 was used to analyse the number of gene clusters of various natural product classes in the genome under relaxed conditions (detecting better-defined full and partial clusters).[54] **C –** aminoglycosides produced by related *Micromonospora* species: gentamicin (*M. echinospora*, *M. pallida*), sisomicin (*M. inyonensis*), sagamicin (*M. sagamiensis*).

2.3.3. Cluster 24 differs from the canonical gentamicin cluster

Cluster 24 was of interest for heterologous expression after early analysis of *M.* sp. DEM32671 culture extracts yielded masses indicative of aminoglycoside production (data not shown). Interestingly, Cluster 24 is located on one of three lower G+C 'islands' in the genome of *M.* sp. DEM32671, suggesting a possible origination from another organism via horizontal gene transfer. The 10 gene clusters detected by AntiSMASH 6.0 as similar to the truncated Cluster 24 detected are presented in **Figure S2.1**.[54] These were from either *Micromonospora*, *Streptomyces*, or *Streptoalloteichus* species, and experimentally characterised gene clusters in the list yielded products such as gentamicin (*M. echinospora* strain DSM 43816, *M. pallida* DSM 43817), sisomicin (*M. inyonensis* DSM 46123), tobramycin (*Streptoalloteichus hindustanus* DSM 44523) or kanamycin (*Streptomyces kanamyceticus* ATCC 23853). The 5 of the 10 gene clusters which showed lowest similarity to Cluster 24 had no experimentally determined aminoglycoside product as of the writing of this work.

Due to Cluster 24 showing highest similarity to the gentamicin cluster (88% of genes in this truncated cluster showing similarity to the *M. echinospora* DSM 43816 cluster; 26% to the next closest match), the gentamicin cluster deposited in the MiBIG database (BGC0000696) was used to probe for the true boundaries of Cluster 24.[60] Due to the cluster architecture including a large gap between genes of biosynthetic relevance, antiSMASH truncated the gene cluster to a single core region spanning the 14858 bp from *genB1* to *genB3*. The boundaries of intact aminoglycoside clusters from four most-related phylogenetic neighbours were also determined using the MiBIG gentamicin gene cluster. Clinker was used to visualise the similarity between these clusters by aligning protein-coding sequences and determining similarity scores. A threshold of either 70% or 95% similarity was required for drawing of 'links' between BGCs **(Figure 2.7)**.[61] From this, even with the full length of the cluster present, highest similarity remains to the *M. echinospora* DSM 43816 gentamicin cluster. The majority of the *M. echinospora* DSM 43816 gentamicin gene cluster shows >96% identity to Cluster 24 **(Table S2.3)**. Early biosynthesis genes

appear to be conserved at this level across all five strains, as each of the aminoglycosides produced by the chosen clusters has a 2-deoxystreptamine precursor.

A closer comparison of *M.* sp. DEM32671 Cluster 24 to the labelled gentamicin cluster from *M. echinospora* DSM 43816 is presented as **Figure 2.8**. The *M. echinospora* DSM 43816 gentamicin cluster contains a single region of difference, consisting of two open reading frames (ORFs), which was found to be present in a region of Cluster 24 outside of the core aminoglycoside biosynthesis region. After using BLASTp to attribute putative function, ORF 1 (according to Figure 2.8; orf_5096) was also annotated as a hypothetical protein and ORF 2 (according to Figure 2.8; orf_5097) was annotated with 95.68% identity to both an ATP-binding protein and putative transcriptional regulator from *M. echinospora* elsewhere in the genome (accessions WP_145834876.1, CAI05926.1 respectively). The highest percentage identity to a named ORF is to a homologue of global regulator AfsR from *Micromonospora* sp. MW-13 (73.12% identity), suggesting that this ORF is likely to be implicated in regulation of transcription but perhaps not specifically regulating gene expression in this cluster. However, analysis using the NCBI Conserved Domain Database suggests that there is similarity to the YlqF-related GTPase superfamily.[62] It remains unclear whether these ORFs being proximally located to the region homologous to the core gentamicin biosynthetic region may have an impact on biosynthesis, whether it be product yield or compound structure.

Another biosynthesis gene, *genL*, is markedly not present in either of these gene clusters. This has previously been described as being located 2.54 Mbp away from the core gene cluster in *M. echinospora* despite being necessary for the conversion of gentamicin C2 to gentamicin C1 and gentamicin C1a to C2b; gentamicin C1 is considered to be one of the most effective congeners antibiotically (generally being measured against a *Bacillus* indicator strain).[63] GenL was identified by Li and colleagues in 2018, after determining that generation of gentamicin C1 was possible after feeding of gentamicin C2 to a

M. echinospora strain missing the core gentamicin gene cluster.[63] After concluding that the gene responsible for this biosynthetic step must lie outside the boundaries of the core gentamicin cluster, the genome of the strain was sequenced, and methyltransferase GenN was used as probe to identify 19 putative candidates with similar protein sequence. The 19 candidates were separately cloned into an *E. coli* expression vector, and after testing cell free lysates for activity, only one (named by the authors as GenL) was able to carry out the conversion of gentamicin C2 to gentamicin C1 (as well as conversion of gentamicin C1a to gentamicin C2b). Homologues of *genL* are present within the genomes of other aminoglycoside producers, similarly spaced from the core cluster region (*M. pallida*: 89.41% identity, 2.64 Mbp from gentamicin cluster; *M. inyonensis*: 92.05% identity, distance from sisomicin cluster unable to be accurately determined (no single contig assembly available at the time of writing); *M. sagamiensis*: 94.09% identity, 2.57 Mbp away from the sagamicin cluster. A homologue of *genL* with 89.54% identity was also found located 2.67 Mbp away from Cluster 24 in the *M.* sp. DEM32671 genome.

A 6 kb island is present at the beginning of Cluster 24 which does not contain any coding sequences, shared in all aminoglycoside producers investigated in this work. In every strain investigated in Figure 2.7, this region contains one rRNA operon (16S and 23S rRNAs annotated). Three rRNA operons were identified in the genome of *M.* sp. DEM23671, but the location of one directly upstream of the aminoglycoside cluster is of interest; as aminoglycosides generally bind to the 16S rRNA, it may be that this operon has evolved with the aminoglycoside cluster to provide a protective effect to the strain as a self-resistance mechanism. The 16S rRNA sequence of this operon did not differ at all from the operon located at 5781059 - 5782250 and differed by a single base pair change (A → G, position 888) from the operon located at 1149791 - 1151309. To the best of our knowledge, this base pair change has not been linked to any known gentamicin resistance mechanisms as of the writing of this work.

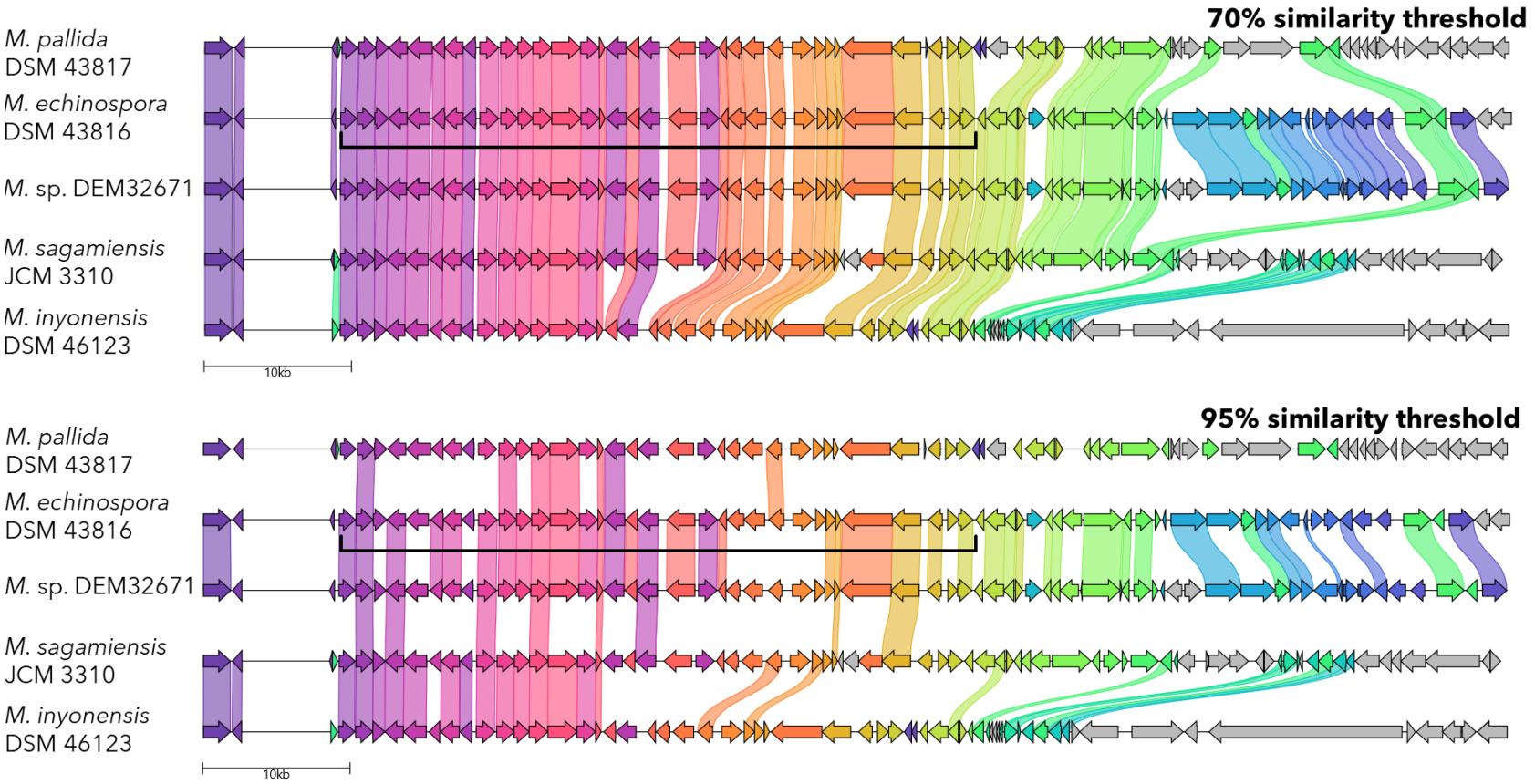


Figure 2.7: clinker diagram of sequence similarity between *M.* sp. DEM32671 Cluster 24 and nearest phylogenetic neighbour aminoglycoside gene cluster.[61] Best link for each gene is visualised. Annotation on *M. echinospora* DSM 43816 marks core region of gentamicin BGC from *genO* to *genN*.

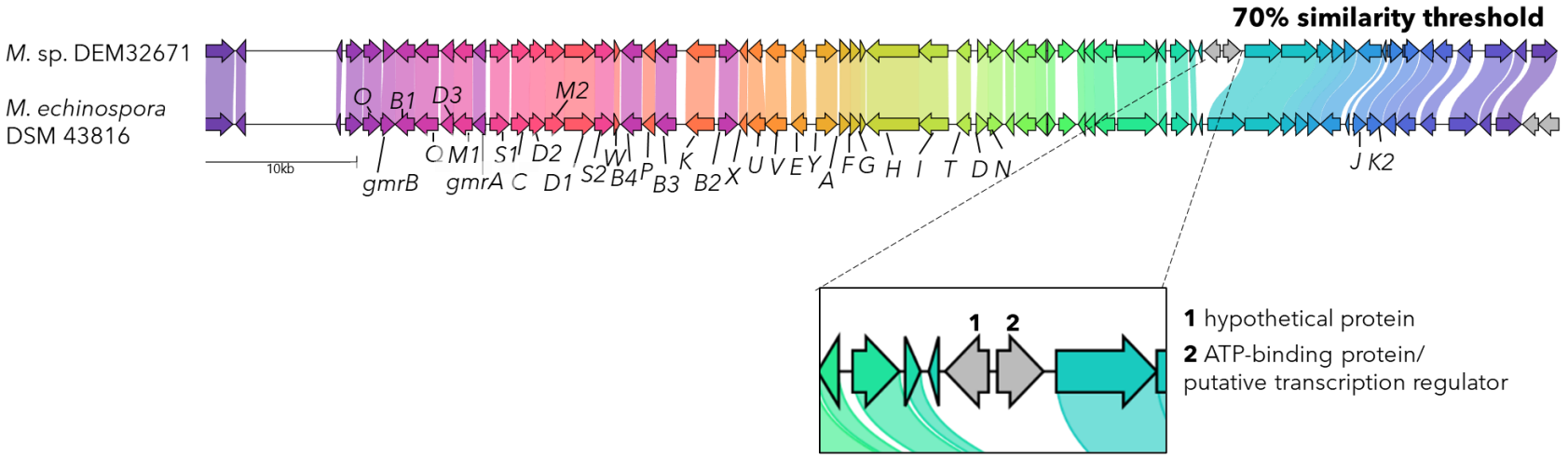


Figure 2.8: clinker diagram of sequence similarity between *M.* sp. DEM32671 Cluster 24 and gentamicin cluster from *M. echinospora* DSM 43816.[61] Best link for each gene is visualised. Gentamicin biosynthesis genes are labelled with the 'gen' labelling scheme favoured by Leadlay and Piepersberg groups, without '*gen*' prefix.[64,65] Open reading frames as labelled: 1 – *M.* sp. DEM32671 orf_5096; 2 – *M.* sp. DEM32671 orf_5097.

2.3.4. Introducing Cluster 24 into *Streptomyces* hosts yields low bioactivity

M. sp. DEM32671 was determined to be challenging to work with, with long culture times (21 days for accumulation of sufficient biomass; Nick Allenby, personal communication) contributing to the need for a heterologous host for characterisation of the output of Cluster 24. While there once was a lack of tools for genetic modification of *Micromonospora* spp., since this work was begun, CRISPR-Cas9 tools have been developed for use in strain *M. echinospora* for the selective production of gentamicin B.[66] Inserting the gene cluster into a heterologous host which does not produce aminoglycosides or other antibiotics should provide a clean background for production screening. Model actinomycetes *Streptomyces coelicolor* M1146 and M1152, and *Streptomyces lividans* TK23 were chosen based on similarity in G+C content (72.2% median G+C of *S. coelicolor* against 72.6% G+C) and their lack of endogenous antibiotic production. *S. coelicolor* M1146 and M1152 have had four antibiotic gene clusters deleted (*Δact Δred Δcpk Δcda*),[43] and while *S. lividans* TK23 can produce antibiotics (such as actinorhodin) which could impact screening, it rarely does so under the culture conditions used in the experiments described here.[67]

An *E. coli* library of phage artificial chromosomes (PACs) containing fragments of *M.* sp. DEM32671 genomic DNA was generated by BioS&T. BioS&T also identified intact Cluster 24 was identified through PCR screening of plates to ensure that the entire gene cluster was present; once obtained, the vector containing Cluster 24 was conjugated into *S. coelicolor* M1146 and M1152, and *S. lividans* TK23 for integration at the φC31 site (data not shown). An extra *S. lividans* strain was generated with extra copies of *genV* (exporter) and *gmrA* (16S methyltransferase conferring resistance) on a plasmid integrated at the φBT1 site (pRes; pR), in the hopes that this may increase antibiotic export into the growth media.

For first trials, liquid cultures of *S. coelicolor* M1152 C24 and *S. lividans* TK23 C24 (+ pRes) were grown in GYM and R2YE medium to test for production of aminoglycosides via bioassay **(Figure 2.9)**. Antimicrobial bioassays were carried out against *Bacillus subtilis* 168, as it has greater susceptibility to gentamicin than other indicator species available in the laboratory (such as *Escherichia coli*).[68,69] After issues with sourcing original strain *M.* sp. DEM32671 or established gentamicin producer *M. echinospora* DSM 43816 for these experiments as a comparison for antibiotic production, purified gentamicin sulfate was used as an indicator of *B. subtilis* growth inhibition. After growing strains for 240 hours, bioactivity was first detected via zones of inhibition at Day 5 (results not shown) which peaked at Day 7. However, as samples were concentrated 50X, and halos remained smaller than the positive control (5 μg gentamicin sulfate), this yield was deemed insufficient for further characterisation based on the detection limits of standards via LC-MS.

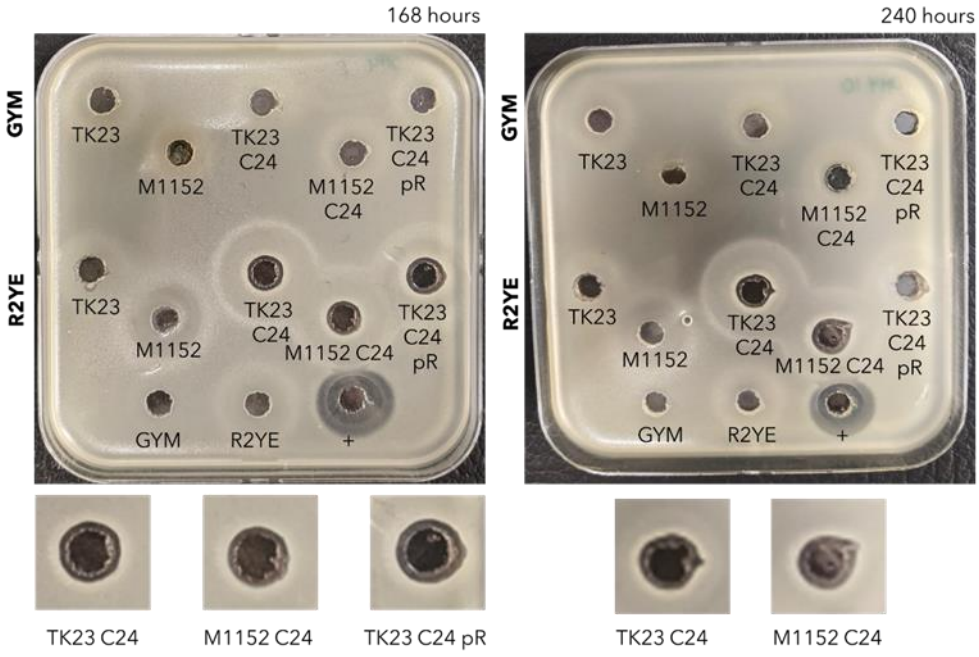


Figure 2.9: Antimicrobial bioassays of *S. coelicolor* M1152 and *S. lividans* TK23 carrying either C24 or a combination of C24 and pRes against *B. subtilis*. Cultures in both GYM and R2YE media were grown for a total of 240 hours (168 hours representing a timepoint of the same set of cultures). Antibiotic activity can be visualised by the inhibition of *B. subtilis* growth. + = gentamicin sulfate, 5 μg.

The growth of *S. coelicolor* M1152 has been previously been reported to be poor, with delayed production of spore pigment being a hallmark of this strain.[43] Due to this, it was posited that further engineering of the strain for yield production may be difficult (taking into account the known problems with the toxicity of expressing Cas9 with existing CRISPR systems in *S. coelicolor*).[70] *S. lividans* TK23 was also not suited to the CRISPR-Cas9 systems available, with much higher editing efficiency being seen with *S. coelicolor*;[70] in addition, the lack of a strain with a clean metabolic background made it a less attractive choice than the *S. coelicolor* strains.

Therefore, with the disadvantages of using these two hosts, an attempt to establish optimal production conditions for strain *S. coelicolor* M1146 C24 was carried out. A panel of eight different media were selected based on a variety of carbon and nitrogen sources and were prepared as both solid and liquid media for testing antibiotic production. A negative control of the parental strain without C24 was grown adjacent. The results are presented as **Figure 2.10**. No production halo was observed for strains grown on solid or liquid media even with the presence of C24, suggesting that either the conditions were not appropriate for production or that the yields of antibiotic produced are too low to visualise with this method.

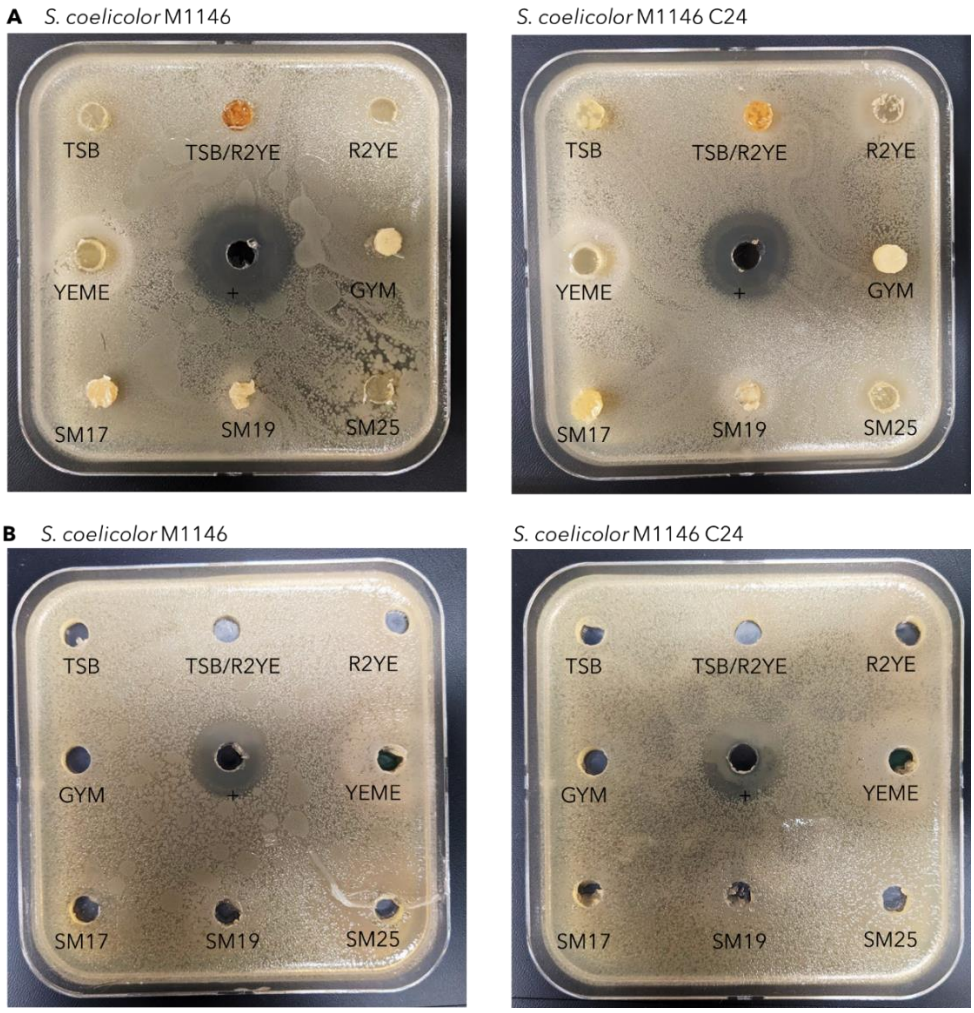


Figure 2.10: Solid and liquid bioassays of *S. coelicolor* M1146 carrying C24. **A** – Bioassays of *S. coelicolor* M1146 C24 on solid medium grown for 14 days before assay against *B. subtilis* 168. + = gentamicin sulfate, 8 μg. **B** – bioassays of *S. coelicolor* M1146 C24 in liquid medium grown for 7 days before assay against *B. subtilis* 168. + = gentamicin sulfate, 5 μg.

Despite challenges, heterologous expression remains an important strategy towards natural product discovery where newly discovered strains are not genetically tractable or culture conditions ill-defined. Testing multiple host strains can also be key to ensure initial production, with two of the three strains tested in this work showing bioactivity under the conditions tested. The *rpoB* mutation introduced to create strain *S. coelicolor* M1152 was originally shown to improve endogenous actinorhodin production only, though from the results presented here, it may be the case that the *rpoB* mutation also functions to improve heterologous antibiotic biosynthesis.[43] By comparison, *S. coelicolor*

M1146 C24 (otherwise identical but lacking the *rpoB* point mutation) does not show antibiotic production in this experiment. Recently, *S. lividans* strains with the gene clusters deleted have been generated by Ahmed et al. (2020); these may provide a better background for heterologous expression than the wild-type strains previously commonly used.[71] Interestingly, the three heterologous hosts tested in this work do not differ in genome sequence by more than 0.72% (when comparing of parental strains of those used in this study, *S. coelicolor* A3(2) against *S. lividans* 1326, by OrthoANI analysis; data not shown), adding to the evidence that both should be tested despite their genetic similarity.

2.3.5. Culturing experiments yield *S. coelicolor* candidates with 5-fold higher resistance to gentamicin

Early heterologous expression experiments with *S. coelicolor* M1146 carrying Cluster 24 did not produce antibiotic at levels able to be detected by antimicrobial bioassay **(Figure 2.10)**. After screening of strains via PCR, it was realised that strain *S. coelicolor* M1146 commonly lost C24 when being passaged on solid medium, despite the vector being integrated to the genome by a φC31 integrase **(Figure 2.11)**.

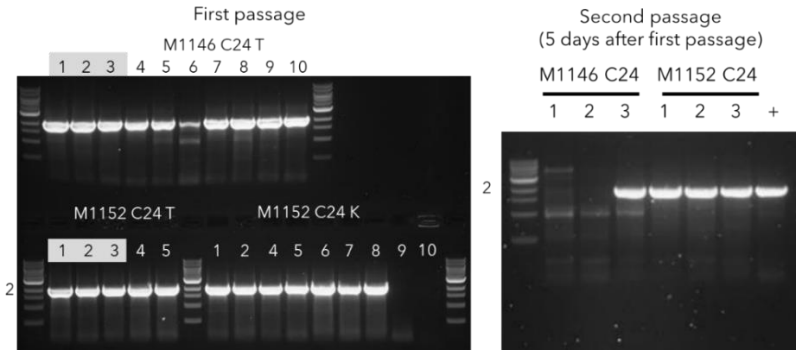


Figure 2.11: Agarose gel of colony PCR of *S. coelicolor* M1146 C24 and *S. coelicolor* M1152 C24 to determine loss of Cluster 24 over time when grown on solid medium. Colony PCR was performed with primers to screen for 1976 bp of the core region of C24 (c44_2kb_gntC_rev, c44_2kb_gntA_for). Lanes representing strains streaked out from passage 1 are marked in grey boxes; colonies grown 5 days later upon repassaging were treated identically upon screening for loss of C24. Molecular weight ladder = 1 kb DNA ladder, NEB.

This cluster loss was hypothesised to be due to production of toxic by-products, which the strain was unable to tolerate. For these experiments, it was only necessary to know whether the predicted biosynthetic region remained intact; as such, the primers used screened for 2 kb of the predicted core biosynthetic region (intragenic regions of *genS1* – *genD2*). It is therefore unclear whether the entire vector was removed from the strain (therefore causing a loss of the thiostrepton/kanamycin resistance cassettes, which would represent a very strong selection against maintenance of Cluster 24) or whether only this core biosynthetic region was deleted. To the best of our knowledge, this strong selection for cluster loss or inactivation after stable integration into the chromosome, without expression of a cognate recombination directionality factor, has not been reported elsewhere.

It was posited that the gene cluster may be producing something toxic to the cells which put pressure on the strain to lose it; therefore, a more resistant strain was required for further testing. It was also hoped that generation of a more resistant strain would increase the yield of antibiotic being produced, allowing for detection and characterisation of the cluster output. While we were able to show bioactivity from *S. coelicolor* M1152 and *S. lividans* TK23 carrying C24, *S. coelicolor* M1146 was chosen to ensure the strain used did not have any issues with poor growth prior to further engineering, and that existing CRISPR-Cas9 modification systems functioned well (respective issues with *S. coelicolor* M1152 and *S. lividans* TK23).[43,70]

A suspension of *S. coelicolor* M1146 carrying C24 was first plated on agar containing 500 μg/mL gentamicin sulfate, before further passaging on a gradient plate with increased gentamicin sulfate concentration of 1000 μg/mL **(Figure 2.12)**. 10 resistant strains were collected from single colonies, with three being prioritised for testing. *S. coelicolor* M1146 C24 2R showed the healthiest growth upon collection of spores **(Figure S2.2)**, while *S. coelicolor* M1146 C24 4R and 9R showed highest resistance to gentamicin on gradient plate (approximated at 800 μg/mL based on location of colonies, dependent on

the assumption that gentamicin diffusion occurred equally across the entire plate; **(Figure 2.12)**.

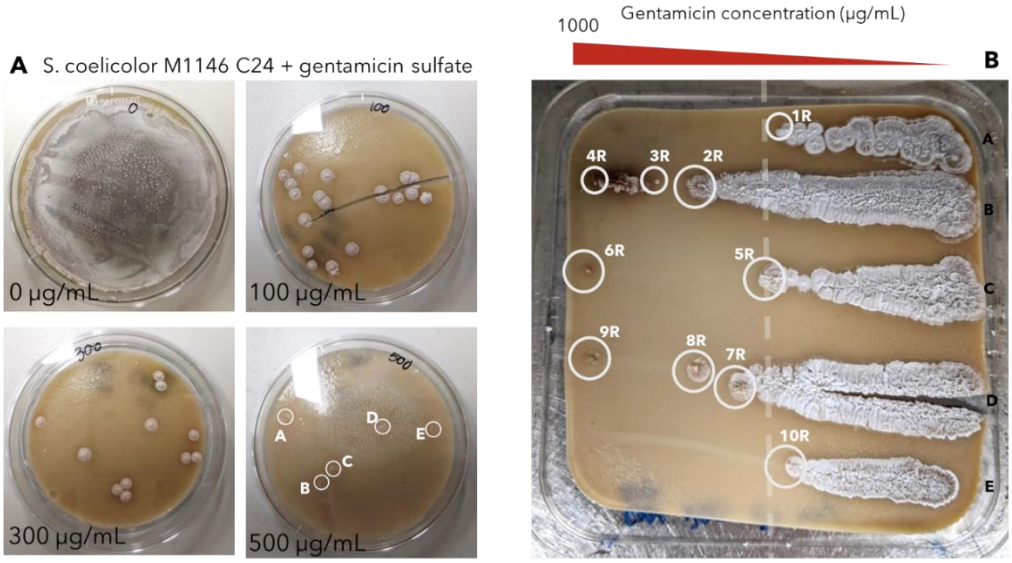
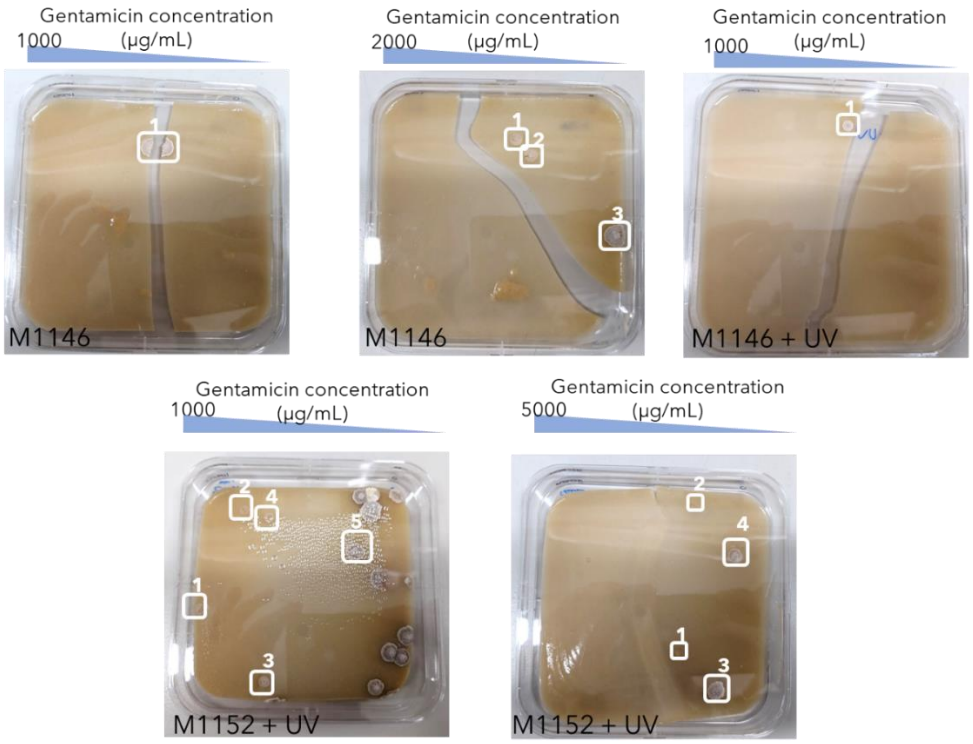


Figure 2.12: Pipeline for generation of resistant *S. coelicolor* M1146 C24 strains. A – ~2x10^7 CFU were plated out on SFM agar with either 0, 100, 300 or 500 µg/mL gentamicin sulfate and 50 µg/mL thiostrepton for maintenance of the biosynthetic gene cluster. **B** – The five colonies from the previous round of culture were streaked out on a gradient plate with maximum gentamicin concentration of 1000 µg/mL (and additional 50 µg/mL thiostrepton) and those circled colonies picked for further testing. Midpoint of the plate where concentration is assumed to be 500 µg/mL is marked with a dashed line.

Similar experiments carried out on *S. coelicolor* M1146 without the cluster present yielded resistant colonies at lower frequency **(Figure 2.13)**. As the only difference between this strain and *S. coelicolor* M1146 C24 was the integration of Cluster 24, it was hypothesised that mutation may be occurring within the gene cluster to allow for growth at such high concentrations of gentamicin. More specifically, it was questioned whether resistance genes or transport genes may be mutated to allow for improved methylation of key nucleotides in the 16S rRNA, or an increased efflux of gentamicin from the cell allowing tolerance at higher concentrations.



Strain	Number of colonies on gradient plates with stated concentration of gentamicin					
	0 - 1000 µg/mL		0 - 2000 µg/mL		0 - 5000 µg/mL	
	-	+ UV (10s)	-	+ UV (10s)	-	+ UV (10s)
S. coelicolor M1146	1	1	3	0	0	0
S. coelicolor M1152	0	10	0	0	0	4

Figure 2.13: Test generation of 'empty' *S. coelicolor* M1146 and *S. coelicolor* M1152 strains resistant to gentamicin sulfate. Strains were plated out on gradient plate of varying concentrations from spore stock and photographed when colonies were present on the plate after 7 days of growth at 30°C

2.3.5.1. Testing resistant strain growth limits in solid and liquid media

As the gradient plate method does not allow for definite quantification of resistance levels, the tolerance of the strains to gentamicin was determined through growth on both solid and liquid medium supplemented with pure gentamicin sulfate. **Figure 2.14A** shows the growth of the strains on agar with increasing concentrations of gentamicin sulfate up to 5000 µg/mL; **Figure 2.14B** shows the same but with liquid medium. The parental strain used as a

control throughout this work was from spore suspension used for the first plating on SFM agar with gentamicin sulfate.

In both cases, all strains displayed healthy growth typical of *S. coelicolor* when grown without antibiotics. On solid medium, when gentamicin was added, all strains (except *S. coelicolor* M1146 C24 4R at 200 μg/mL gentamicin) did not show proper production of spore pigment at day 5 of growth. By 8 days of growth, the phenotype of *S. coelicolor* M1146 C24 4R appeared unaffected by the presence of gentamicin, suggesting a delayed spore pigment production. All other strains appeared atypical in regard to expected growth phenotypes even by day 8 **(Figure S2.3)**. *S. coelicolor* M1146 C24 2R showed the highest resistance of the four strains, with breakthrough growth at 5000 μg/mL: this represents a 5-fold increase in resistance to gentamicin in comparison to the parental strain. On solid medium, *S. coelicolor* M1146 C24 4R showed resistance up until 2000 μg/mL gentamicin, with healthy growth at 1000 μg/mL of gentamicin. *S. coelicolor* M1146 C24 9R performed the weakest of the three, showing breakthrough resistance up to 1000 μg/mL gentamicin but with strong growth on solid medium only occurring up to 200 μg/mL gentamicin.

The strains performed similarly in liquid cultures, with clear differences between control strain *S. coelicolor* M1146 C24 and resistant strains being detectable from 1000 μg/mL gentamicin onwards. Again, *S. coelicolor* M1146 C24 2R performed the best of the strains, with growth occurring even at 5000 μg/mL gentamicin in liquid medium. As culture inoculums were normalised by CFU and so should be growing to similar OD_{450} values, it appeared that *S. coelicolor* M1146 C24 4R and 9R showed either a reduced or a slower growth; however, the growth kinetics of these strains were not verified during this work. Phenotypically, all strains grew dispersed and as typical for *S. coelicolor*.

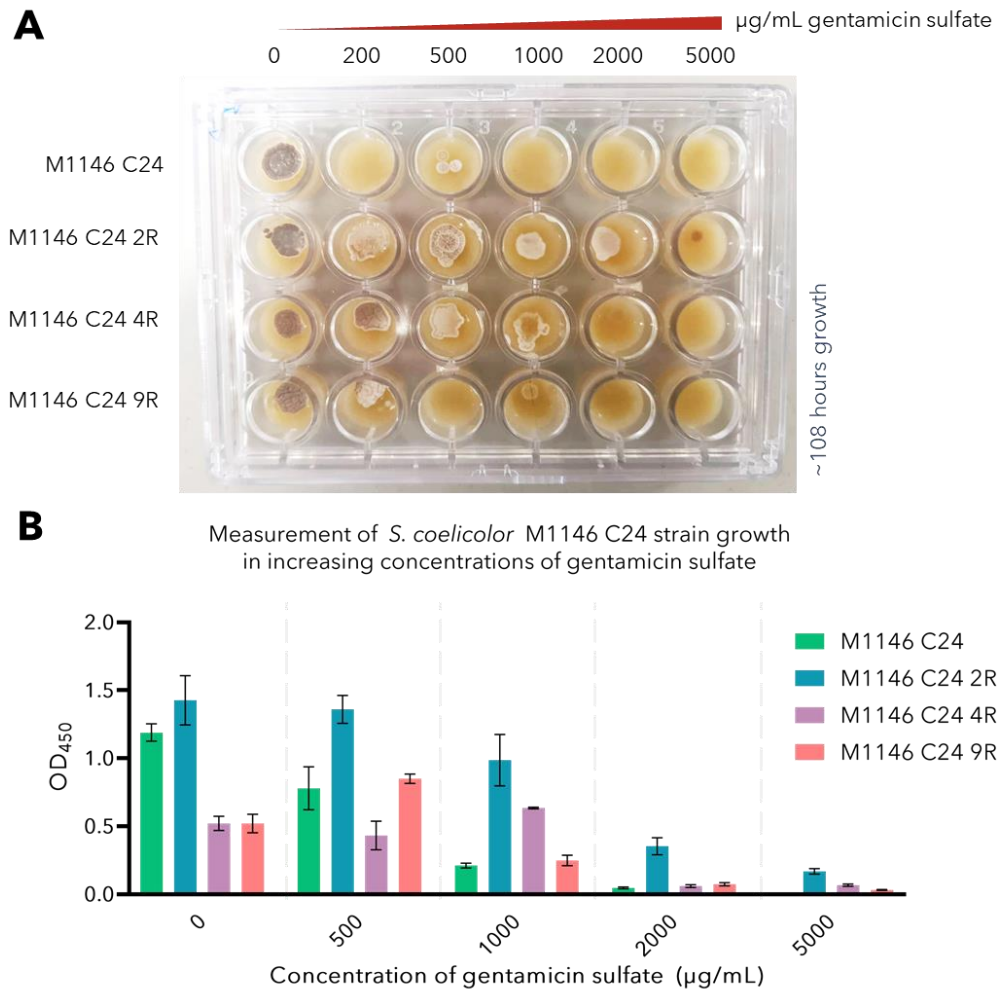


Figure 2.14: Measurement of *S. coelicolor* M1146 C24 and mutant derivative strains' resistance to gentamicin. A – test of strain growth in increasing concentrations of gentamicin sulfate in SFM solid agar medium, photographed after ~108 hours. **B** – test of strain growth in increasing concentrations of gentamicin sulfate in TSB liquid medium. Strains were grown in triplicate biological replicates and growth measured by OD_{450} readings.

2.3.6. *S. coelicolor* M1146 C24 2R and *S. coelicolor* M1146 C24 4R shows differential resistance to aminoglycoside antibiotics

Having determined that the three strains were able to maintain resistance to high levels of gentamicin, it was next decided to test their susceptibility to other aminoglycosides. The aminoglycoside family of antibiotics all bind to the bacterial ribosome to interrupt normal protein biosynthesis.[72] Different concentrations of each antibiotic were used for this experiment to avoid

overwhelming the strain and therefore being unable to detect mutations which may convey very low levels of resistance. For all antibiotics with *S. coelicolor* MICs which were either 1-2 μg/mL or undescribed, 5 μg/mL antibiotic was used;[73,74] for kanamycin and chloramphenicol which have higher MICs for *S. coelicolor*, 1.25X the stated MIC was used.[73]

We expected resistance to have evolved due to modification of components of the 30S subunit of the ribosome, as this is the target of gentamicin.[42,75] Therefore, non-aminoglycoside chloramphenicol was selected as a control antibiotic, as its antibiotic activity is caused by its binding to the 50S subunit of the prokaryotic ribosome.[76]

Table 2.3: Antibiotics and concentrations used for testing cross-resistance of gentamicin-resistant strains.

Antibiotic	MIC in *S. coelicolor*	Concentration used in this work
Kanamycin	> 400 μg/mL with *gmrA* resistance gene; 400 μg/mL with neo phosphotransferase[73]	500 μg/mL
Streptomycin	2 μg/mL[74]	5 μg/mL
Spectinomycin	Not described	5 μg/mL
Hygromycin	Not described	5 μg/mL
Apramycin	< 1 μg/mL without resistance gene[73]	5 μg/mL
Chloramphenicol	80 μg/mL[77]	100 μg/mL
Gentamicin	> 200 μg/mL with *gmrA* resistance gene[73]	2000 μg/mL

The two best performing strains from the previous trial were selected and grown as described in Table 2.3 **(Figure 2.15)**. As expected, the strain growth without antibiotics was as previously shown in **Figure 2.14**, with *S. coelicolor* M1146 C24 2R showing the highest level of growth (with error bars suggesting growth being comparable to *S. coelicolor* M1146 C24) and *S. coelicolor* M1146 C24 4R showing a reduced level of growth despite inoculation of the same estimated CFU. Strain growth in 2000 ug/mL gentamicin also showed similar results to those previously shown, as *S. coelicolor* M1146 C24 2R showed approximately the same levels of growth and again *S. coelicolor* M1146 C24 4R showed a reduced level of growth (but still break-through growth of the

antibiotic). The resistance to kanamycin exhibited by both mutant strains was to be expected; the resistance genes for gentamicin and kanamycin often exhibit cross-resistance to each other due to the similarity in structure of these antibiotics. Some basal resistance to kanamycin can be seen by *S.* coelicolor M1146 C24, likely due to presence of the resistance genes within Cluster 24. However, the mutations causing increased resistance to gentamicin in *S. coelicolor* M1146 C24 2R and *S. coelicolor* M1146 C24 4R also appear to be causing some resistance to kanamycin, though the concentration used here is lower than that of gentamicin (500 ug/mL compared to 2000 ug/mL; still representing a 10-fold increase in usual working concentration of kanamycin with *Streptomyces*).[73] The strains do not appear resistant at all to apramycin. This could be due to apramycin having a slightly different structure (being a non-2-DOS-based aminoglycoside) or the fact that it has a slightly different mode of action on the bacterial ribosome.[40] No difference in resistance was seen to spectinomycin or hygromycin, suggesting that the parental strain maintains a basal level of resistance to these concentrations of these antibiotics. Additionally, the reduced growth of *S. coelicolor* M1146 C24 4R in comparison to the other two strains was seen upon exposure to spectinomycin or hygromycin. Interestingly, strain *S. coelicolor* M1146 C24 4R showed a higher resistance to streptomycin than both *S. coelicolor* M1146 C24 2R and *S. coelicolor* M1146 C24, though the reason for this could not be immediately identified. Lastly, no growth was seen from strains grown in the presence of chloramphenicol; as this antibiotic binds to an entirely different ribosomal target than the others, this was to be expected.[76]

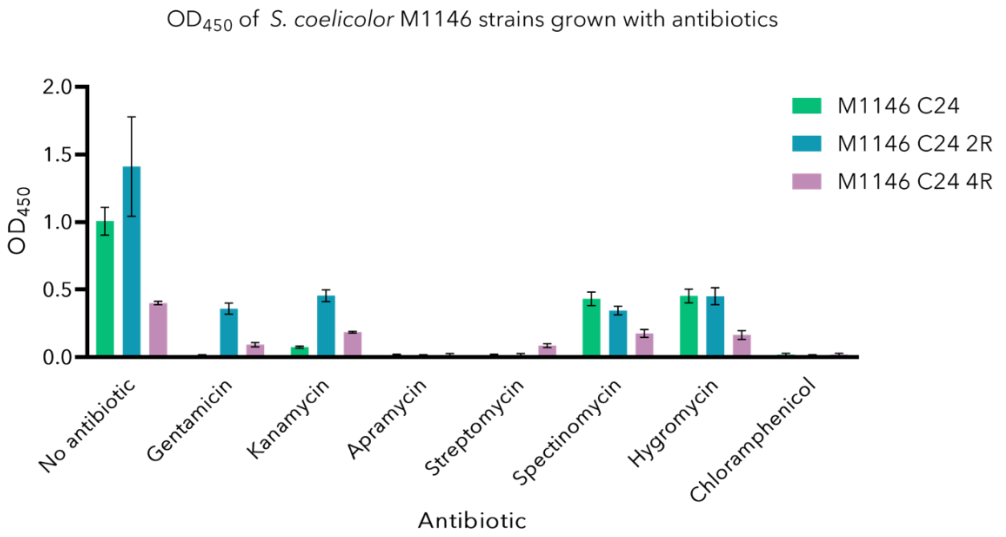


Figure 2.15: Measurement of *S. coelicolor* M1146 C24 and mutant derivative strains' resistance to various antibiotics. Strains were grown in TSB liquid medium with either: 2000 μg/mL, 500 μg/mL kanamycin, 5 μg/mL apramycin, 5 μg/mL streptomycin, 5 μg/mL spectinomycin, 5 μg/mL hygromycin, 100 μg/mL chloramphenicol. Strains were grown in triplicate biological replicates and growth measured by OD_{450} readings.

2.3.7. Characterisation of gentamicin-resistant strains via whole-genome sequencing

Understanding the mechanisms conveying the high levels of gentamicin resistance of strain *S. coelicolor* M1146 C24 2R and the cross-resistance to streptomycin of *S. coelicolor* M1146 C24 4R was of interest, in the event that this could be used for other yield-improvement experiments. Illumina sequencing was performed by MicrobesNG and was used to investigate effects on the entire genome (mean coverage *S. coelicolor* M1146 C24 = 48.63x; *S. coelicolor* M1146 C24 2R = 49.03x; *S. coelicolor* M1146 C24 2R = 70.25x). Assembly statistics are provided as **Table S2.4**.

Across both resistant strains, 148 shared regions of difference (including base substitutions and larger insertions and deletions in both intergenic regions and coding sequences) were present upon comparison with the parental strain. The visualisation of the distribution of mutations across these strains is presented

as **Figure 2.16**; of the 187 total mutations, 24 were unique to *S. coelicolor* M1146 C24 2R and 15 were unique to *S. coelicolor* M1146 C24 4R. These differences between the two resistant strains are most likely to contribute to the different growth phenotypes. Of the mutations to *S. coelicolor* M1146 C24 2R, seven of the 24 areas of difference were present within coding regions, whilst in *S. coelicolor* M1146 C24 4R eight of the 15 regions of difference were present within coding regions **(Table 2.4)**. These follow the general trend of mutations predicted to be genes involved in key cellular processes, including an acetyl-CoA/pyruvate carboxylase,[78,79] c-type cytochrome biogenesis protein,[80] DEAD box helicase,[81] phosphofructokinase,[82] and putative transcriptional regulators.[83,84] The mutations to these basic cellular processes may have caused the reduced growth fitness phenotype of *S. coelicolor* M1146 C24 4R. However, none of these mutations can be easily explained as being a clear cause for the differential resistance phenotypes seen in strains *S. coelicolor* M1146 C24 2R and 4R. Similarly, mutations specific to each strain which occur in intergenic regions do not have a clear link towards impacting the expression of any clear resistance-linked genes **(Table 2.5)**.

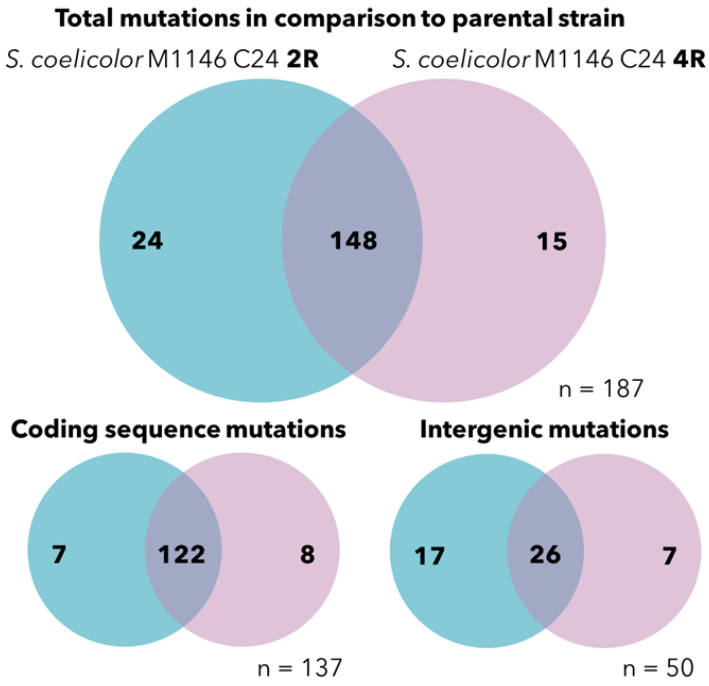


Figure 2.16: Summary information on mutations in *S. coelicolor* M1146 C24 2R and *S. coelicolor* M1146 C24 4R in comparison to parental strain. Numbers in blue circle represent mutations in *S. coelicolor* M1146 C24 2R; numbers in purple circle represent mutations in *S. coelicolor* M1146 C24 4R.

Mutation to ribosomal components is a common cause towards increase of antibiotic resistance for aminoglycosides, as highlighted earlier in this work.[28–31] For example, in the case of gentamicin, modification by methylation of a single nucleotide (G_{1405}) of the 16S rRNA by a 16S rRNA (guanine$_{1405}$-N7)-methyltransferase has been shown to convey resistance.[85] Other aminoglycosides also show this resistance mechanism by either methylation of G1405 or A1408.[86] The 16S rRNA gene was first checked for mutations, as the binding of gentamicin and other similar aminoglycosides to the region of the ribosome encoded by this gene is well described.[42,75,87] Alignment of the *S. coelicolor* 16S rRNA gene (NCBI accession X60514.1, length 1551 bp) to matching sequences in *S. coelicolor* M1146 C24, *S. coelicolor* M1146 C24 2R, and *S. coelicolor* M1146 C24 4R genome data showed that the same five mutations were present across all three strains (G742C, C743G, C903G, A1172C, G1174T).

Table 2.4: Mutations in coding sequences unique to *S. coelicolor* M1146 C24 2R and 4R. Predicted functions were attributed using BLASTp and the NCBI Conserved Domains Database.[88]

Unique mutations in *S. coelicolor* M1146 C24 2R coding regions				
DNA mutation	DNA mutation site	AA mutation	Predicted function	Notes
C -> T	204587 bp; contig 7	R35Q	TetR/AcR family transcriptional regulator	Mutation within predicted pfam00440 TetR; no annotated active site
A -> G	26684 bp; contig 11	T570A	DNA polymerase III subunit	Mutation within region of non-specific hit PRK07764; no annotated active site
390 bp deletion	26707 - 27096 bp; contig 11	Deletion of 130 AAs between G577 and V708		Mutation outside of predicted domain boundaries
165 bp insertion	31449 - 31450 bp; contig 17	Extra 37 AAs between P416 and D417	ADP-ribosylglycohydrolase	Mutation outside of predicted domain boundaries
T -> A	115 bp; contig 20	T299S	1-phosphofructokinase	Mutation outside of predicted substrate- and ATP-binding sites
CA -> GC	37228 - 37229 bp; contig 20	Silent G49G, A50P	FMN-dependent L-lactate dehydrogenase	Mutation outside of predicted domains, mutations outside of predicted active sites, substrate- and FMN-binding sites
T -> A	60 bp, contig 62	X406C	N-acetylglucosamine repressor/sugar kinase	Mutation outside of predicted domains, mutations outside of predicted DNA- and Zn^{2+} binding sites
Unique mutations in *S. coelicolor* M1146 C24 4R coding regions				
DNA mutation	DNA mutation site	AA mutation	Predicted function	Notes
360 bp deletion	278679 - 279038 bp; contig 2	Deletion between E391 and I510	Tetratricopeptide repeat protein	Removes single predicted protein binding surface
21 bp insertion	31491 bp; contig 17	Insertion of 'DRLPPAP' between P430 and D431	ADP-ribosylglycohydrolase	Mutation outside of predicted domain boundaries
A -> G	240 bp; contig 24	T54A	Biotin-dependent enzyme, potentially acetyl-CoA or pyruvate carboxylase	Mutation outside of predicted biotinylation sites/carboxyltransferase interaction sites/biotinyl domain)
G -> T	22536 bp; contig 27	P47H	PAP2 superfamily enzyme	Mutation outside of predicted active site and PAP2 domain

A -> G	43659 bp; contig 28	S366P	Cytochrome C biogenesis protein	Mutation within predicted pfam01578 (cytochrome C assembly)
G -> C	95421 bp; contig 30	Silent G216G	TetR/AcR family transcriptional regulator	Mutation outside of predicted conserved domains
195 bp insertion	51905 bp; contig 40	Extra 65 AAs between G611 and H612	ATP-dependent DEAD box helicase	Mutation outside of predicted conserved domain; mutation outside of ATP binding sites/DEAD box helicase motifs
C -> G	28500 bp; contig 49	V343L	No putative conserved domains detected	No putative conserved domains detected; function of protein unclear

Table 2.5: Mutations in intergenic sequences unique to resistant strains *S. coelicolor* M1146 C24 2R and 4R. Predicted functions were attributed using BLASTp and the NCBI Conserved Domains Database.[88]

Unique mutations in *S. coelicolor* M1146 C24 2R intergenic regions			
DNA mutation	DNA mutation site	Genes flanking	Notes
Insertion of 'CC'	273964 - 273966 bp; contig 2	ATP-binding protein and DUF5926 family protein	Possible disruption of terminator between two proteins; possible disruption of promoter upstream of DUF5926 family protein
G -> A	149642 bp; contig 2	isoprenyl transferase and DUF5324 family protein	Possible disruption of terminator between two genes
ggcgtccgacggacgccggcccgccc to ----------cg- cggcccgc**gg**	43876 - 43941 bp; contig 2	*genB4* and *genP*	Possible effect on translation of *genB4* (C24 from *M.* sp. DEM32671)
T -> G	556916 bp; contig 5	DNA starvation/stationary phase protection protein and orfA (PRC-barrel domain-containing protein)	Possible disruption of terminator between two genes
A -> G	39029 bp; contig 7	hypothetical protein (no conserved domains) and PLP-dependent aminotransferase family protein	Should not affect surrounding genes
204 bp insertion	72348 - 72551 bp; contig 9	LPXTG cell wall anchor domain-containing protein and threonine-phosphate decarboxylase	Should not affect surrounding genes

cggtggcttgagcgcgggggcgggt to ----------gggggcggga	6990 - 7014; contig 11	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase and cysteine--tRNA ligase	No promoter predicted, but could disrupt if there is one
47 bp deletion	15949 - 15995 bp; contig 13	Acyl-homoserine lactone acylase QuiP and 5-formyltetrahydrofolate cyclo-ligase	Should not affect surrounding genes
4 bp insertion followed by 33 bp deletion	24296 - 24328 bp; contig 15	trkA potassium uptake protein and DUF3159 domain-containing protein	Possible disruption of terminator between two genes
G -> C, T -> C in region of mutation	T -> C at 84927, G -> C at 84944 bp; contig 15	serine protease and hypothetical protein (no conserved domains)	Possible disruption of terminator between two genes
56 bp insertion	151868 - 151923 bp; contig 20	beta-ketoadipate pathway transcriptional regulator and 2-keto-3-deoxy-D-arabino-heptulosonate-7-phosphate synthase	Should not affect surrounding genes
75 bp deletion	4961 - 5035 bp; contig 24	phosphomannomutase and purine-nucleoside phosphorylase	mutation upstream of putative phosphomannomutase, no promoter predicted
agga to tgtc	109474 - 109477 bp; contig 24	DUF3017 domain-containing protein and Xenobiotic Response Element (XRE) family transcriptional regulator	Possible disruption of promoter upstream of XRE family regulator but none predicted
Insertion of 8 bp	86588 bp; contig 25	MerR family transcriptional regulator and GNAT family N-acetyltransferase (RimJ/RimL family protein acetyltransferases)	No likely effect (upstream of *speG*, spermidine N(1)-acetyltransferase
cgggggccga to -ggggggcgc	87034 - 87043 bp; contig 29	hypothetical protein (no conserved domains) and helix-turn-helix transcriptional regulator	Possible effect on terminator downstream of hypothetical protein
gtac- to -tacc	89672 - 89675 bp; contig 31	and DUF4191 domain-containing protein	Possible disruption of terminator between two genes
34 bp insertion	55425 bp; contig 32	phosphatase PAP2 family protein and heavy metal translocating P-type ATPase	No promoter predicted, but could disrupt if there is one upstream of PAP2 family protein
Intergenic region mutations specific to *S. coelicolor* M1146 C24 4R			
DNA mutation	DNA mutation site	Predicted function of flanking genes	Notes

42 bp deletion	368165 - 368206 bp; contig 1	Rne/Rng family ribonuclease and glycerol-1-phosphate dehydrogenase	Possible disruption of expression of glycerol-1-phosphate dehydrogenase (no promoter predicted)
1 bp deletion	277896 bp; contig 4	No predicted function for either flanking gene	Should have no effect on surrounding genes
82 bp insertion	7036 bp; contig 11	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase and cysteine--tRNA ligase	No promoter predicted, but could disrupt if there is one
T -> G	5546 bp; contig 19	ADP-forming succinate--CoA ligase subunit beta and hypothetical protein (no conserved domains)	No promoter predicted, but could disrupt promoter upstream of ADP-forming succinate--CoA ligase subunit beta
364 bp insertion	95197 bp; contig 30	MFS transporter and TetR/AcrR family transcriptional regulator	May disrupt terminator if present
'acccccag' to 'tccgccccc'	59345 - 59352 bp; contig 40	trimeric intracellular cation channel family protein and pyrimidine reductase (riboflavin biosynthesis)	Trimeric cation channel protein is predicted to be in an operon with a M20 family metallopeptidase and MerR family transcriptional regulator
48 bp insertion	14004 bp; contig 51	DNA phosphorothioation-associated putative methyltransferase and hypothetical protein (no predicted function)	Hypothetical protein is predicted to be in an operon with two other hypothetical proteins (both without predicted function)

These mutations do not map to the known binding pocket of gentamicin,[42,75,87] but could suggest a burden on the cell from simply carrying Cluster 24. It is important to note that with the Illumina 2x250 bp paired-end reads, repeat sequences longer than 1 kb are not able to be assembled: this includes rRNA operons, of which *S. coelicolor* has 6.[89] Therefore, there could be mutations in the other operons which were unable to be detected using this sequencing technology. One way to establish these mutations would be to combine the Illumina reads with a long-read technology such as PacBio or Oxford Nanopore. Alternatively, PCR of the rRNA operon components separately, cloning into a vector, transformation to *E. coli* and screening of multiple clones with Sanger sequencing would establish a better picture of the complete rRNA operon sequences. This would also avoid the repeat finding of consensus rRNA sequences, which would occur without separating out the operons prior to sequencing.

No other sequence difference to either coding sequences or intergenic regions for ribosomal or RNA polymerase components could be identified. Of particular interest was whether any mutation to *rpsL* was present in cross-resistant strain *S. coelicolor* M1146 C24 4R, as the product (S12 protein) is one of the antibiotic targets for streptomycin.[90] Another candidate which may have conveyed selective streptomycin resistance was the *mscL* gene. The product of *mscL* functions to release molecules upon osmotic shock, aiding cell survival by reducing pressure. Streptomycin's ability to act on a cell has been linked to expression of the MscL channel in *E. coli* (a component of the RpoS regulon),[91] and so it was hypothesised that an altered expression of MscL may convey the resistant phenotype seen for strain *S. coelicolor* M1146 C24 4R. However, no mutation to these genes was seen.

It is possible (and likely) that the resistant phenotypes of both strains occurred with multiple cumulative mutations. As such, the mutations shared between the two mutant strains may also be of interest if these results were to be attempted to be replicated. These mutations were not fully investigated as part

of this work but are inserted as **Table S2.5, Table S2.6**. Even in this list of shared mutations, there does not appear to be any logical single candidate which could cause this level of resistance; however, two point mutations have been made to a predicted ABC-type multidrug transporter, a 3 amino acid insertion to a predicted ABC-transporter and a single point mutation to a predicted ABC transporter permease subunit. As efflux can be a key self-resistance mechanism for antibiotic producers,[92] it may be that these are the most interesting targets established from this work. However, further experiments would be required to confirm whether these predicted transporters can efflux aminoglycosides.

Within the boundaries of C24, three regions of mutation were present in comparison to the genome sequence of *M.* sp. DEM32671. Of these, there was a single point mutation within a coding region: A601H in the homologue for *genH*. *genH* was predicted to function as an efflux MFS transporter when the sequence was analysed with BLASTp and the NCBI Conserved Domains Database,[88] but to the best of our knowledge the function has never been experimentally characterised. Whilst mutation in this candidate aligns with the phenotypes of the resistant strains, this point mutation was also present in the parental strain and so can be discarded as a cause of the difference in phenotype. The other regions of mutation were located in intergenic regions up- and downstream of the homologue of *genB4*, recently characterised as being involved in late-stage gentamicin C biosynthesis.[93] The promoter for this operon was predicted by BPROM to be further upstream (in front of *genB3*), and analysis with the Salis Lab RBS calculator suggested that the ribosome binding site for *genB4* should not have been disrupted by the region of mutation **(Figure 2.17)**.[94,95] The region downstream of *genB4* contained several potential point mutations differing slightly between the three strains sequenced. Considering cluster architecture, it was predicted that the transcription and translation of surrounding genes should not be impacted. However, this could be further explored by using RT-qPCR to analyse the expression of *genB4* in these culture conditions.

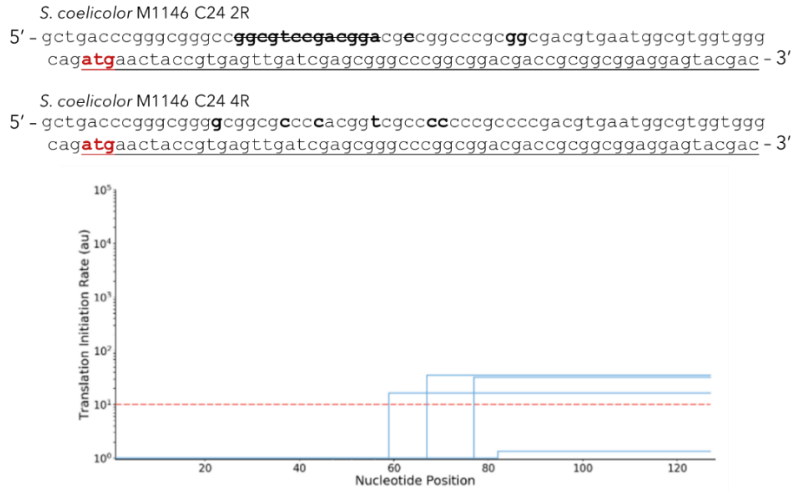


Figure 2.17: Analysis of mutations upstream of genB4 to determine whether an impact is likely on cluster gene expression. First nucleotide 'G' represents first nucleotide after the stop codon for *genP*, the gene upstream in the operon. Deleted nucleotides are striked through substituted nucleotides are in bold, start codon for *genB4* is in red, coding region is underlined. Graph is an estimation of the translation initiation rate (au) of different nucleotide positions upstream of *genB4* to determine whether mutation was likely to disrupt ribosome binding site. Figure was generated by the Salis Lab RBS calculator.[94] Regions of mutation were present from position 15 to position 43, dependent on strain. Position 67 represents the annotated start codon for the homologue for *genB4.*

Multiple studies in recent years have used antibiotics as a mutagenic pressure on actinomycetes for increase of production of endogenous secondary metabolites. This approach has so far not been well-utilised for heterologous production approaches.[29–31] Generally in these previous works, mutation to ribosomal sequences yields the large increase in secondary metabolite production. The mutations identified in this work were unexpected, as none of them were easily linked to previously studied mechanisms of resistance. The mean coverage of the Illumina sequencing used for the three strains was as follows: *S. coelicolor* M1146 C24 = 48.63x; *S. coelicolor* M1146 C24 2R = 49.03x; *S. coelicolor* M1146 C24 2R = 70.25x, and as expected with using Illumina sequencing alone, a single contig was not able to be assembled. Importantly, long repeat sequences such as rRNA operons – of high importance to this work – are unable to be resolved using the sequencing platform used. To further

confirm these mutations and generate a closed genome sequence, the short-read Illumina sequencing could be combined with a secondary sequencing technology such as Oxford Nanopore or PacBio,[96–99] which provide long-reads to bridge gaps. The work described here is solely predictive, and so to truly characterise the mutations causing this resistant phenotype complementation of the mutations for the resistant strains would be the next step in confirming relevance. This could be achieved through further CRISPR-Cas9 engineering. While our main objective for these experiments was to increase strain fitness for production, a high-throughput screen to establish which of many independent colonies may overproduce antibiotics would have been of additional value. Despite advances over the past decade, [100–103] high-throughput culturing can remain challenging with actinomycetes – in some cases, these strains only produce antibiotics in very limited culture conditions incompatible with small-scale growth. If this can be improved upon in the coming years, it will likely impact massively on the genome-mining/natural products production pipeline**.**

2.3.8. Resistant strain *S. coelicolor* M1146 C24 2R shows interesting production activity

Antibiotic bioassay of the three resistant strains of interest (*S. coelicolor* M1146 C24 2R, 4R and 9R) showed varying results. **Figure 2.18** shows antibiotic bioassay of culture extracts from *S. coelicolor* M1146 C24 2R, *S. coelicolor* M1146 C24 4R and *S. coelicolor* M1146 C24 9R. Upon first test with a similar panel of media as used previously **(Figure S2.4)**, production was only able to be achieved in one of nine media (TSB/R2YE). This was then used for all further experiments with any strains carrying Cluster 24. *S. coelicolor* M1146 C24 2R showed production of antibiotic reproducible across biological and media replicates, but *S. coelicolor* M1146 C24 4R and 9R showed diffused halos which were not as easily visualised. From the bioassay, a titre of approximately 1 μg/mL gentamicin in 50 mL shake flasks was roughly estimated for *S. coelicolor*

M1146 C24 2R by comparing halo sizes with dilutions of gentamicin standards against *B. subtilis* **(Figure 2.18)**.

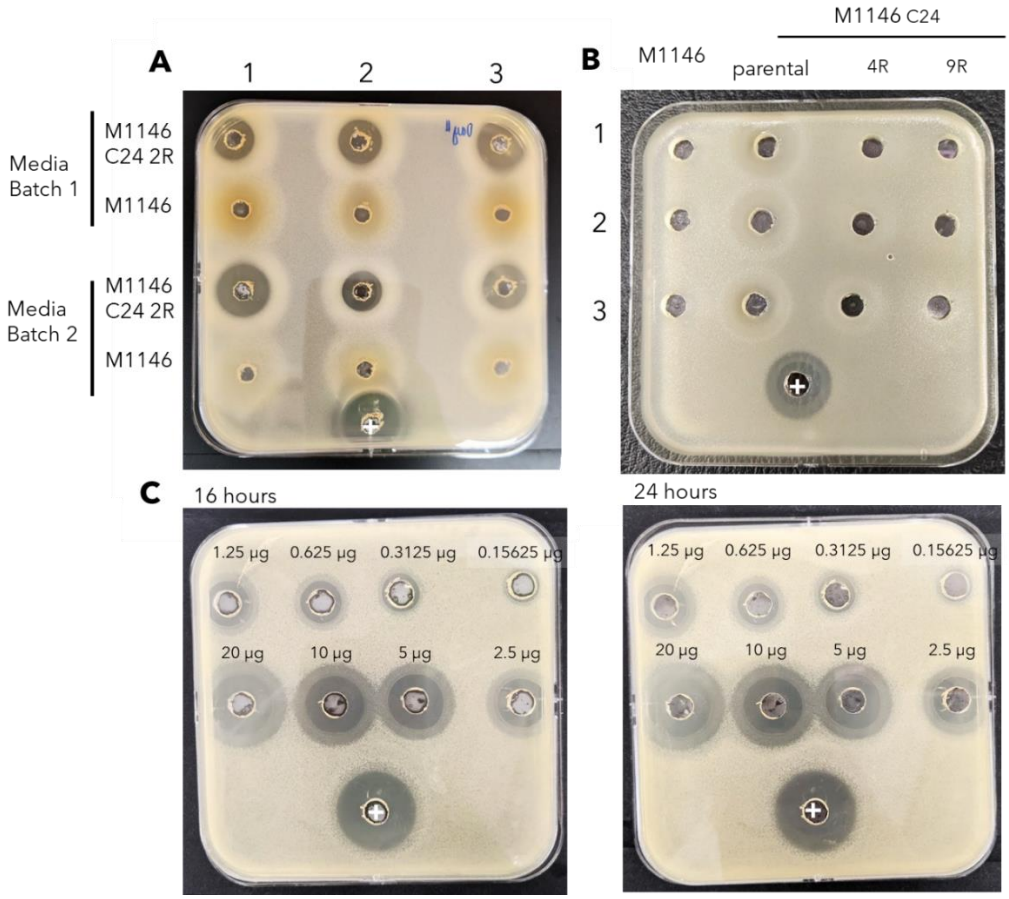


Figure 2.18: Bioassay of *S. coelicolor* M1146 C24 2R, *S. coelicolor* M1146 C24 4R and *S. coelicolor* M1146 C24 9R against *B. subtilis* 168. A – Bioassay of *S. coelicolor* M1146 C24 2R. Samples from day 11 of culture were concentrated 100X by freeze drying and 100 µL added to each well. + = gentamicin sulfate, 5 µg. **B** – Bioassay of *S. coelicolor* M1146 C24 4R and *S. coelicolor* M1146 C24 9R against *B. subtilis*. Samples from day 11 of culture were concentrated 100X by freeze drying and 100 µL added to each well. + = gentamicin sulfate, 5 µg. **C** – Serial dilutions of gentamicin sulfate in sterile ddH_2O, bioassayed against *B. subtilis* in LB agar and photographed twice at 16 and 24 hours of growth at 37°C. + = 10 µg of chloramphenicol.

Samples were grouped into 4 experimental groups of independently cultured samples for analysis of LC-MS samples. Group 1 consisted of a single *S. coelicolor* M1146 C24 2R culture extract from first media test with TSB/R2YE medium; Group 2 and 3 consisted of extracts from biological replicates of *S. coelicolor* M1146 C24 2R shown in Figure 2.18 grown in media batch 1 and

media batch 2 respectively; Group 4 consisted of extracts from *S. coelicolor* M1146 C24 4R and 9R grouped together due to their similar performance on bioassay. For each of these, a negative control group was included (*S. coelicolor* M1146 samples numbered 1 through 3 respective to the experimental group), with negative control group 2 also serving as a control for experimental group 4. Upon analysis of extracts from these strains with LC-MS, peaks matching gentamicin intermediates 2-deoxy-scyllo-inososе (2-DOI), 2-deoxy-scyllo-inosamine (2-DOIA), 2-deoxystreptamine (2-DOS) and paromamine were able to be detected in experimental samples **(Figure 2.19 & Figure 2.20)**. For peak selection in Figure 2.19 and Figure 2.20, peaks with masses with a ppm of >5 were discounted, as well as peaks with intensity below 1000. Peaks matching paromamine were detected multiple times within the chromatograms; all other intermediates appeared once as an M+H adduct. Confirmation of the identity of these peaks with standards or with further fragmentation was not able to be carried out, but this would increase confidence that these species are present. 2-DOI represents the second metabolite in the gentamicin pathway, 2-DOIA the third, 2-DOS the fifth, and paromamine the seventh.[104]

Of the gentamicin congeners, C2a was the only one able to be assigned to a peak, albeit in the negative controls only **(Figure 2.21)**. This is therefore likely to be invalid assignation.

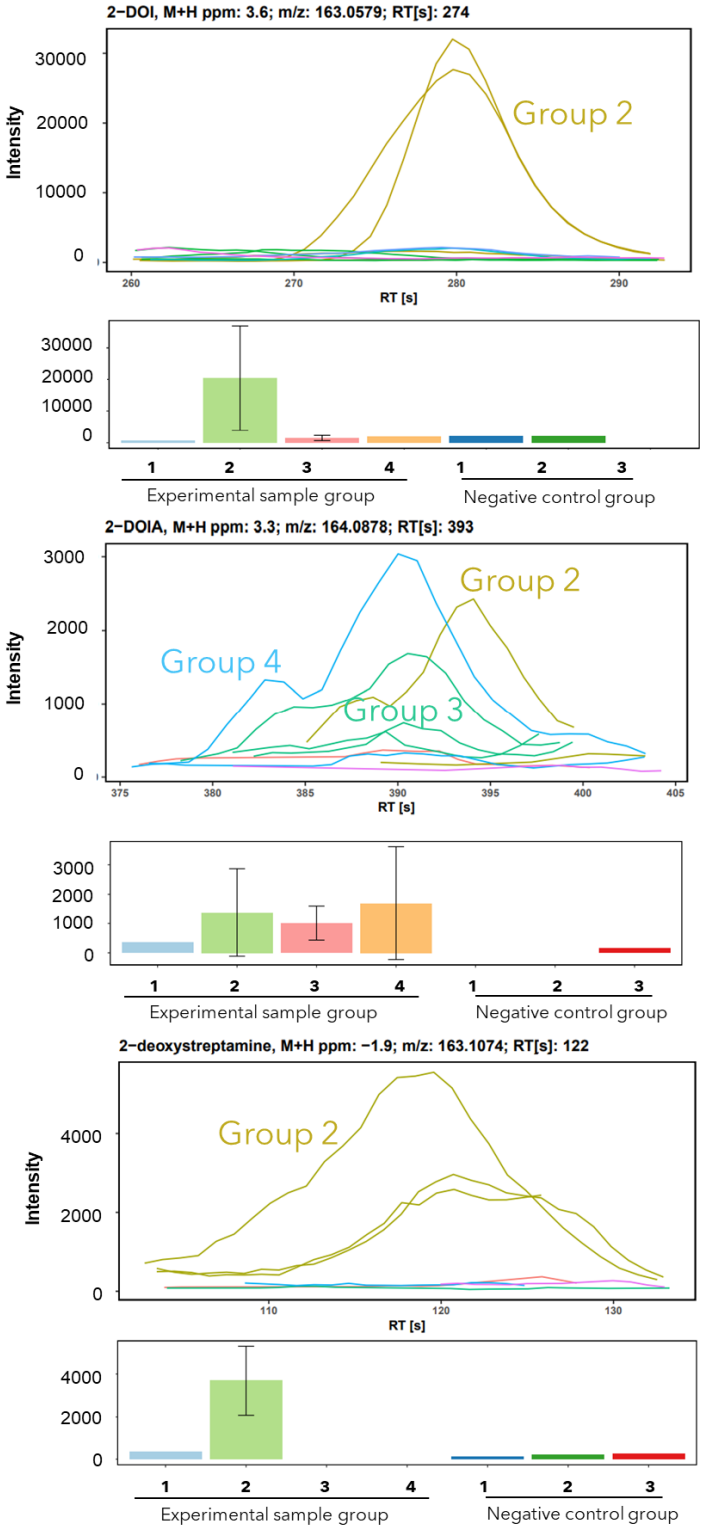


Figure 2.19: LC-MS peaks assigned to 4,6-disubstituted 2-DOS aminoglycoside precursors 2-DOI, 2-DOIA and 2-DOS across experimental groups. All detected M+H adducts are included. Groups 1-3 represent *S. coelicolor* M1146 C24 2R cultured at independent times, Group 4 represents *S. coelicolor* M1146 C24 4R and 9R. Negative control groups (*S. coelicolor* M1146) were cultured at the same time as the experimental sample groups; negative control for experimental group 4 is control group 2.

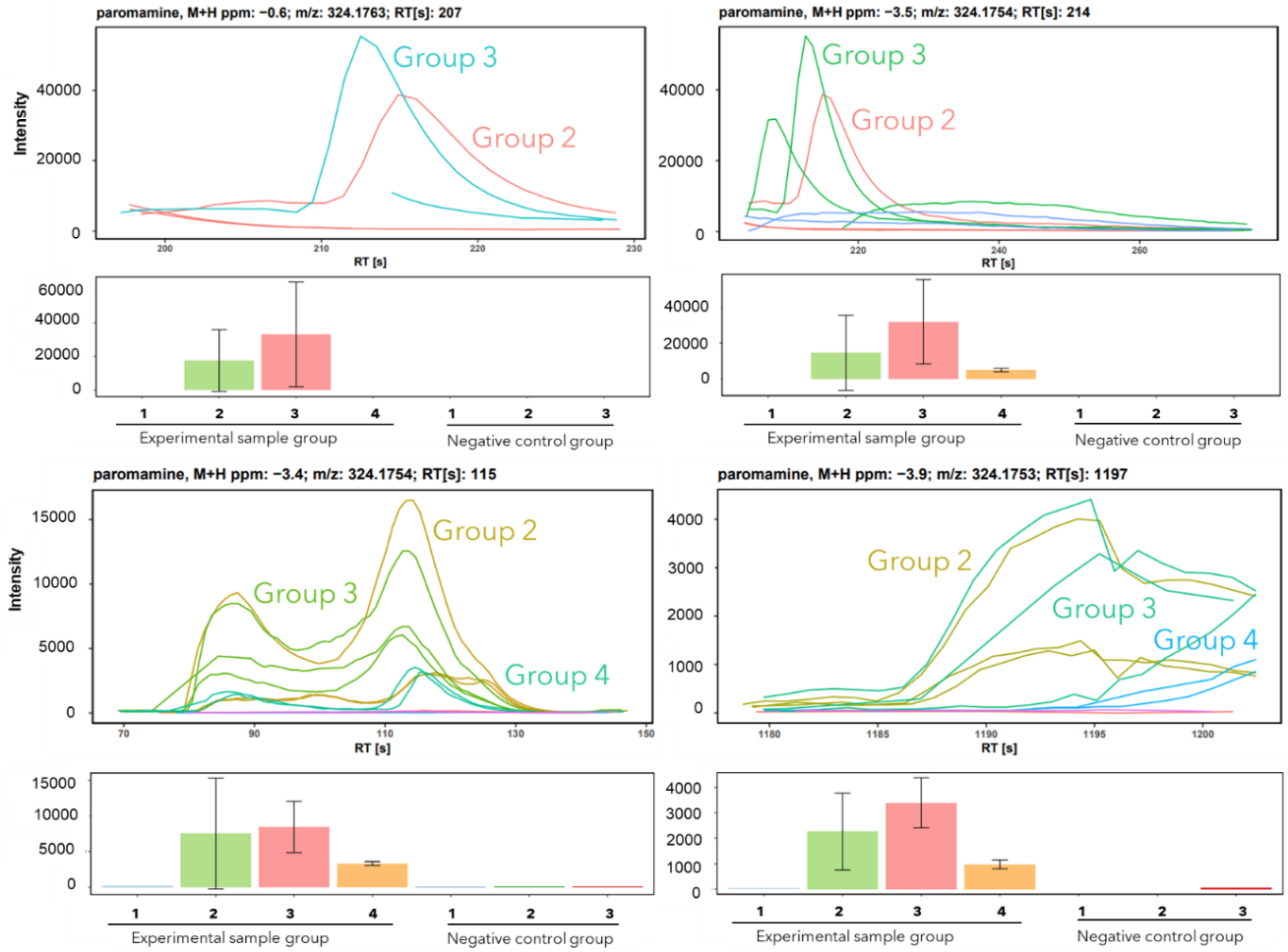


Figure 2.20: LC-MS peaks assigned to 4,6-disubstituted 2-DOS aminoglycoside precursor paromamine across experimental groups. All detected M+H adducts are included. Groups 1-3 represent *S. coelicolor* M1146 C24 2R cultured at independent times, Group 4 represents *S. coelicolor* M1146 C24 4R and 9R. Negative control groups (*S. coelicolor* M1146) were cultured at the same time as the experimental sample groups; negative control for experimental group 4 is control group 2.

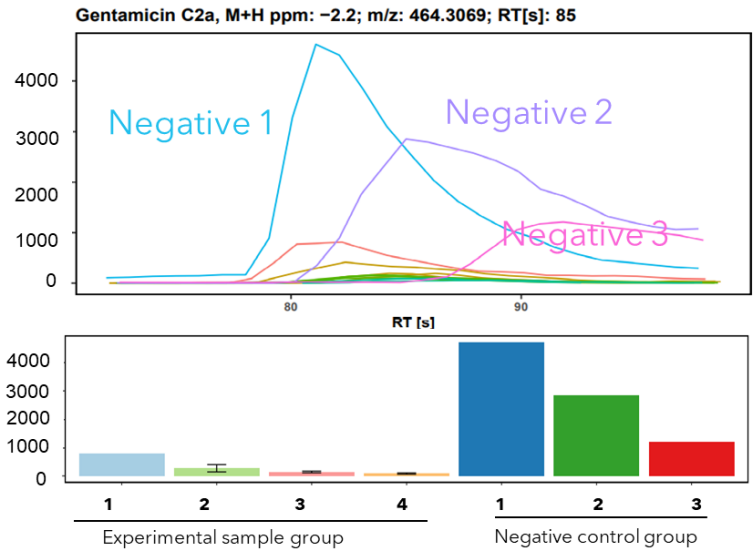


Figure 2.21: LC-MS peaks assigned to gentamicin C2a. Presence of C2a congener was only detected in the negative controls which have no gentamicin production capability (*S. coelicolor* M1146).

The only other ion matching the aminoglycoside class of compounds was of m/z 502.2246, which upon fragmentation showed a pattern similar to the *in silico* fragmentation of hybrimycin-related antibiotic LL-BM 434β.[105] From this, the chemical formula $C_{18}H_{35}N_3O_{13}$ was suggested (M+H m/z = 502.2248). This formula is compatible with three known aminoglycosides: 2-hydroxy-6'-deaminokanamycin, LL-BM 27β (produced by *Streptomyces* sp., including native kanamycin producer *Streptomyces kanamyceticus*),[105] and chitotriose (three linked 2-amino-2-deoxy-D-glucopyranose units). The manual fragmentation of each of these is presented as **Figure S2.6**. All putative compounds were of interest, as while the kanamycin and gentamicin pathways share precursors and homologous enzymes, no specialised kanamycin biosynthesis past paromamine should be possible in theory.[104,106] While the peak was of low intensity, it was present only in groups of samples from *S. coelicolor* M1146 C24 2R which showed clear bioactivity on bioassay; samples from *S. coelicolor* M1146 C24 4R and 9R (grouped with PeakML to Group 4 in **Figure 2.22**) did not show this ion.

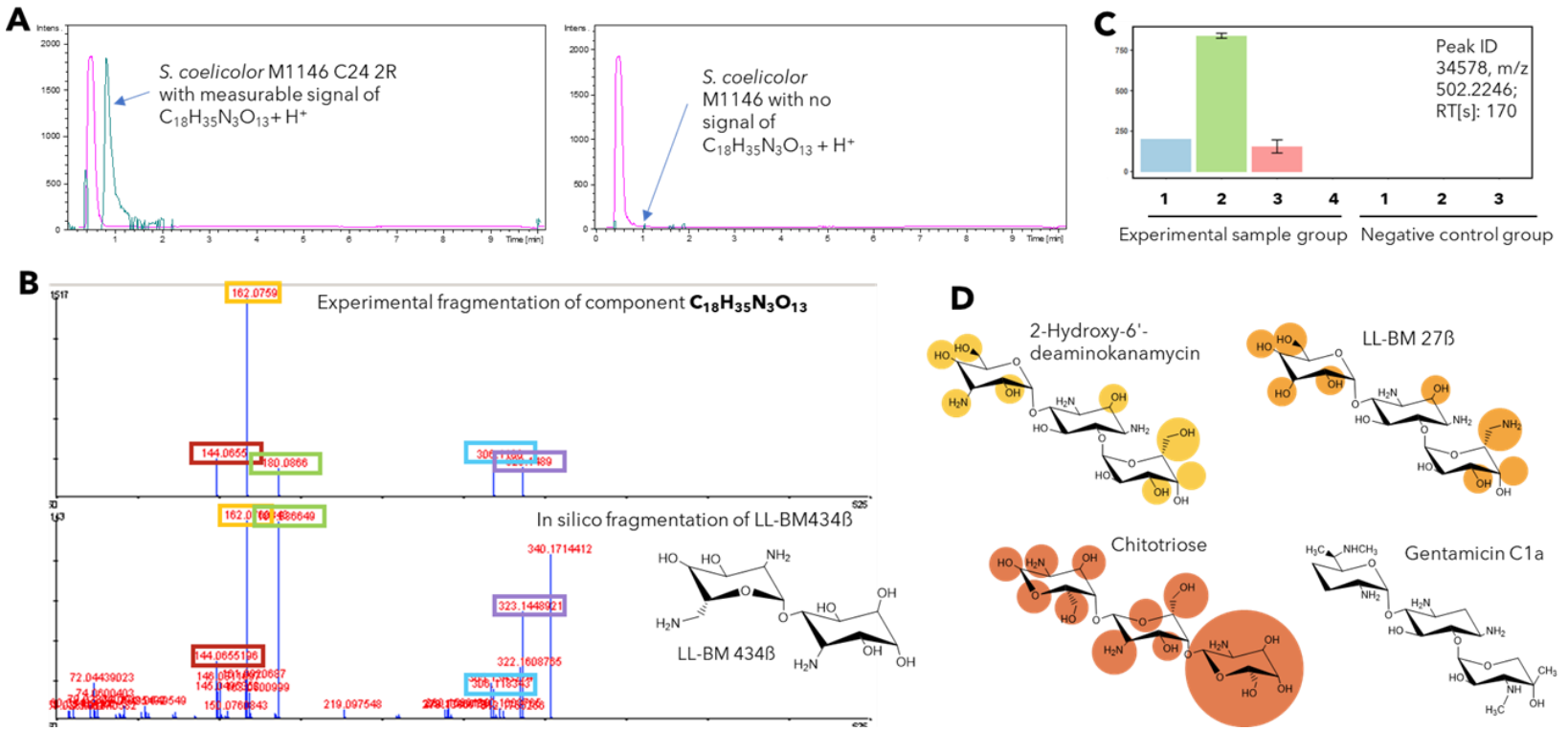


Figure 2.22: LC-MS results showing presence of aminoglycoside ion with m/z 502.2246. A – extracted ion chromatograms of single experimental group 2 sample (left) and single negative control group 2 sample (right), selecting for m/z 502.2240 ± 0.003. Dark blue trace = 502.2240 ±0.003; Purple trace = UV_{210nm}. **B** – Fragmentation MS of m/z 502.2246 in comparison to *in silico* fragmentation of aminoglycoside LL-BM 434ß; peaks of similar mass are highlighted in same colours. **C** – PeakML-generated estimation of peak intensity for m/z 502.2246 across experimental groups. Groups 1-3 represent *S. coelicolor* M1146 C24 2R cultured at independent times, Group 4 represents M1146 C24 4R and 9R. Negative control groups (*S. coelicolor* M1146) were cultured at the same time as the experimental sample groups; negative control for experimental group 4 is control group 2. **D** – Structures of putative compound of m/z 502.2246 and chemical formula $C_{18}H_{35}N_3O_{13}$ compared to expected product gentamicin C1a. Coloured circles represent areas of difference.

It is important to note that while gentamicin C congeners in standards were able to be detected in water **(Figure S2.5)**, detection of gentamicin congeners or intermediates in standards dissolved in the same TSB/R2YE media used for bacterial cultures was not possible (in concentrations from 62.5 μg/mL to 0.9765625 μg/mL, reducing by half each time). When considering the results from bioassay in comparison to the peak intensity of m/z 502.2246, it can be realised that while the zones of inhibition between media batches do not show marked differences in size, the peak intensity of species with m/z 502.2246 drops by approximately 75% when comparing samples from batch 2 to batch 1. We therefore hypothesised that gentamicin or gentamicin intermediates were present in these samples, but that the high intensity peaks of the media components, through ion suppression, impacted on visibility of gentamicin congeners in culture samples. Aminoglycosides are notoriously challenging to purify; their hydrophilic nature ensures that organic solvent-based extraction procedures are not suitable.[107] In many complex matrices, use of a solid phase extraction (SPE) column is one of the few extraction options. These allow for the retention of non-polar species and the flow-through of aminoglycosides.[107] However, a C18 SPE column was utilised as such in this work and was insufficient to clean up samples in the TSB/R2YE culture medium enough to avoid a possible ion suppression effect. An additional challenge with detection is that aminoglycosides lack a UV chromophore which would enable them to be detected via spectrophotometric methods, such as HPLC coupled to a UV detector.[108] Previous studies have used derivatization to improve gentamicin detection, either by chromatographic methods[109–111] or in the case of Omar et al. (2013), measurement of fluorescent product by coupling to biological stain safranin.[112] However, in this case, further yield improvement was the only path forward to aid in characterisation.

While yields of antibiotic were not able to be quantified via LC-MS, we estimate production in 250 mL shake flasks to be on the micrograms per liter scale. Previous reports have suggested that with tailoring of media conditions, a yield of gentamicin on the grams per liter scale could be expected when

culturing native producer *M. echinospora* (in the cited work, named *Micromonospora purpurea*).[113] There therefore remains much room for improvement, but this study highlights the feasibility of using model actinomycete *S. coelicolor* M1146 for production of aminoglycosides, giving further options for increasing product yields.

The results from our analysis of culture extracts from strains carrying Cluster 24 suggest the possible production of an interesting metabolite not previously linked to gentamicin biosynthesis, though further LC-MS analysis on further culture replicates must be conducted to definitively confirm this. We showed earlier in this work that Cluster 24 from *M.* sp. DEM32671 differs from the canonical gentamicin gene cluster in *M. echinospora* DSM 43816 by the addition of two ORFs: one predicted to be a transcriptional regulator, and the other with no predicted function **(Figure 2.8)**. Homologues to these genes are present in *M. echinospora* DSM 43816 but located elsewhere in the genome. Whilst the function of these ORFs in the context of aminoglycoside biosynthesis have not been characterised, it remains possible that these genes can generate a scaffold with slight molecular differences to the expected gentamicin C congeners. Further experiments investigating whether the products of these genes show *in vitro* enzyme activity on gentamicin C congeners could establish this much needed characterisation. A construct could be generated with each gene linked to an epitope tag. After purifying the enzyme using appropriate column chromatography, various gentamicin congeners could be used as feedstocks to determine modification via LC-MS and tandem MS.

While the sizes of zones of inhibition in bioassay of samples between media batches were not shown to have great differences, a large difference in peak intensity for putative product $C_{18}H_{35}N_3O_{13}$ was seen. It is possible that this could indicate that antibiotic activity is being caused by another molecular species which was unable to be determined under the current chromatography conditions. Without carrying out quantitative LC-MS (with an internal standard of known concentration), this however remains unclear. The media

components remain a likely source of the proposed ion suppression effect – when gentamicin sulfate standards in another growth medium (GYM medium, containing glucose, yeast extract and malt extract) were initially tested via LC-MS, the correct congeners were able to be visualized, albeit with lower peak intensity than samples in pure water. However, we were unable to establish antibiotic production by *S. coelicolor* M1146 C24 in GYM medium when this was tested **(Figure 2.23)**.

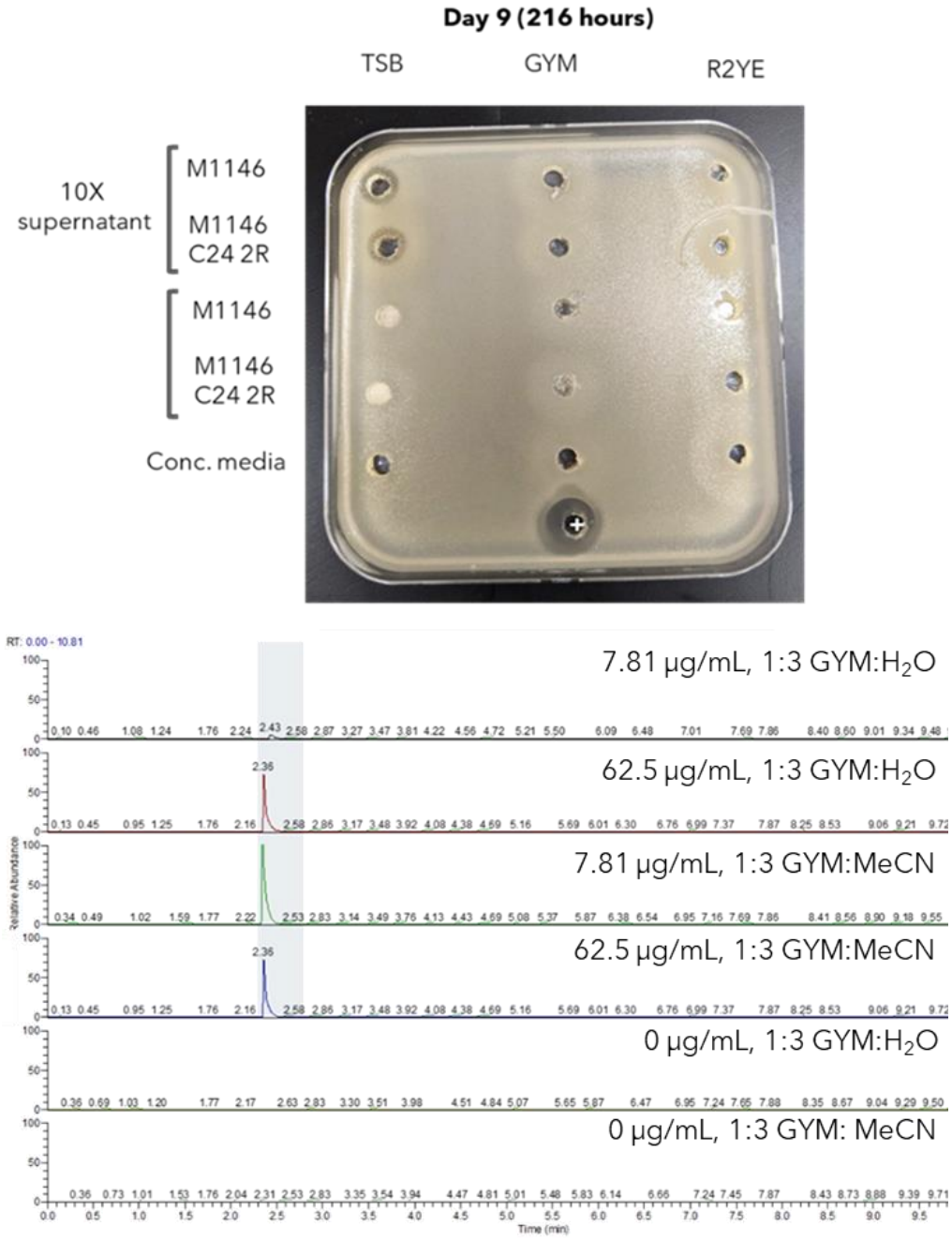


Figure 2.23: Bioassay of extracts from *S. coelicolor* M1146 C24 2R in three different media (TSB, GYM, R2YE) and extracted ion chromatograms of 62.5 µg/mL gentamicin sulfate and 7.81 µg/mL gentamicin sulfate standards in mixtures of GYM and water/acetonitrile (MeCN). EIC = 478.30 – 478.35 for gentamicin C1 (M+H m/z =478.32). Antibiotic production was not established in any of the media conditions tested via bioassay, but gentamicin C1 was able to be visualised in diluted GYM media via LC-MS.

In this study, we report the identification of *M.* sp. DEM32671 as a talented producer of natural products. Based on the evidence gathered, we suggest that *M.* sp. DEM32671 can be classed either as a new species, or new subspecies of *M. echinospora*.

After genome sequencing the strain, we identified 31 gene clusters predicted to produce a variety of secondary metabolites, including echinosporamicin-type antibiotic TLN-5520, and aminoglycoside gentamicin.[55] The cluster was successfully integrated into the genomes of three test *Streptomyces* hosts. We improved the suitability of our candidate strain by increasing resistance to the predicted cluster output, gentamicin. With the resultant increase in antibiotic yield visualised via antimicrobial bioassay, we were able to begin to characterise the output of Cluster 24 and identified interesting metabolite of mass 501.2246 which could be linked to presence of this gene cluster. To our knowledge, this is the first report of an intact full-sized aminoglycoside cluster being introduced to *Streptomyces coelicolor*.

Using heterologous expression as a tool for production of new secondary metabolites is likely to continue to have impacts on drug development, particularly of development of antibiotics. By genome sequencing strains it is possible to predict the full portfolio of natural products available *in silico* and identify those most useful to us for production in a more amenable host. Nevertheless, low starter yields remain an issue with the technique and remains a bottleneck for full characterisation of the compound(s) of interest. This work identifies significant challenges in heterologous expression – the reliance upon very specific conditions for production of compound is well-illustrated in the approach towards media choice in this work. While production of antimicrobial from strains carrying Cluster 24 was a major breakthrough, it is important to note that this was only able to be achieved with very specific media conditions. Success in heterologous expression without further cluster or strain editing remains highly challenging, with

many published attempts yielding no compound or less compound than the original host strain.[114] Coupling heterologous expression strategies with strain or cluster engineering may yield a path forward in this area,[115,116] though strategies are likely to be required to be highly tailored to the specific gene cluster of interest.

This study takes further steps towards generation of a platform strain for aminoglycoside production, but yield could be further improved through engineering known rate-limiting enzymes,[117] or optimizing media conditions as done for the native producer of gentamicin.[113] Utilising a more genetically amenable strain gives many more options to further increase yields and may soon provide a challenge for currently-used industrial producers of valuable secondary metabolites.

2.5. Materials and Methods

2.5.1. Strains and plasmids

All bacterial strains and plasmids used and generated in this work are listed in **Table 2.6.**

Table 2.6: Strains and plasmids used and generated in this work.

Bacterial strain	Genotype	Use	Reference
Micromonospora sp. DEM32671	Proprietary	Phenotypic comparison to type strains	This work
Escherichia coli NEB5α	*fhuA2 (argF-lacZ) U169 phoA glnV44 80 (lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	Cloning	[118]
E. coli ET12567 pUZ8002	*Fdam13::Tn9 dcm6 hsdM hsdR zjj202::Tn10 recF143 galK2 galT22 ara14 lacY1 xyl5 leuB6 thi1 tonA31 rpsL136 hisG4 tsx78 mtl1 glnV44 pUZ8002*	Conjugation of plasmids into Streptomyces	[119]
Bacillus subtilis 168	*trpC2*	Indicator strain for bioassays	[120,121]
Streptomyces lividans TK23	*spc-1 SLP2- SLP3-*	Expression of Cluster 24-derived plasmids	[73,122]
Streptomyces coelicolor M1152	*S. coelicolor* M145 *Δact Δred Δcda Δcpk rpoB*[C1298T])	Expression and modification of Cluster 24-derived plasmids	[43]
Streptomyces coelicolor M1146	*S. coelicolor* M145 *Δact Δred Δcda Δcpk*	Expression of Cluster 24-derived plasmids	[43]
Streptomyces coelicolor M1146 C24 2R	*S. coelicolor* M145 *Δact Δred Δcda Δcpk* C24 (172 mutations, detailed in Table 2.4, Table 2.5, Table S2.5, Table S2.6)	Improvement of antibiotic yield	This work
Streptomyces coelicolor M1146 C24 4R	*S. coelicolor* M145 Δact Δred Δcda Δcpk C24 (163 mutations, detailed in Table 2.4,	Improvement of antibiotic yield	This work

Streptomyces coelicolor M1146 C24 9R	Table 2.5, Table S2.5, Table S2.6) *S. coelicolor* M145 Δact Δred Δcda Δcpk C24 (Unknown mutations)	Improvement of antibiotic yield	This work

Plasmid	**Backbone**	**Insert**	**Reference**
C24	pESAC-13	Cluster 24 (170 kb region homologous to region of gentamicin cluster from *M. echinospora* DSM 43816)	This work
pRes	pT801	*ermEp1-gmrA-genV*	[123]

2.5.2. Media and cultivation conditions

All actinomycete liquid cultures were grown in 50 mL of media in siliconized 250 mL Erlenmeyer flasks with 10 mM diameter coil springs. *S. coelicolor* seed cultures (~10^7 CFU from spore suspension) were cultivated in tryptone soya broth (TSB; 1.7% pancreatic digest of casein, 0.3% enzymatic digest of soya bean, 0.5% sodium chloride, 0.25% K_2HPO_4, 0.5% glucose; obtained as pre-mixed powder from Oxoid cat. no. CM0129) for 48 hours (30°C, 180 rpm) before inoculation of 1 mL of seed culture to production medium. *S. coelicolor* production cultures were carried out in multiple media: GYM (0.4% glucose, 0.4% yeast extract, 1% malt extract);[73] R2YE (10.3% sucrose, 1% glucose, 1.12% $MgCl_2.6H_2O$, 0.025% K_2SO_4, 0.01% Difco casamino acids, 0.5% Difco yeast extract, 0.05% KH_2PO_4, 0.57% TES Buffer, 0.3% $CaCl_2.2H_2O$, 0.3% L-proline, 0.02% NaOH, 0.2% trace elements solution.);[73] INA5 (3% glycerol, 1.5% soybean meal, 0.2% NaCl, 0.5% $CaCO_3$. pH adjusted to 7.2 before autoclaving);[124] YEME (0.3% yeast extract, 0.3% malt extract, 0.5% peptone, 1% glucose, 34% sucrose, 1% $MgCl_2.6H_2O$);[73] SM17 (0.2% glucose, 4% glycerol, 0.2% soluble starch, 0.5% soy flour, 0.5% peptone, 0.5% yeast extract, 0.5% NaCl, 0.2% $CaCO_3$);[125] SM19 (4% tomato paste, 1.5% oat flour, 0.2% glucose);[125] SM25 (1% peptone, 2.1% malt extract, 4% glycerol).[126] TSB/R2YE medium was

prepared with a 1:1 ratio of unautoclaved TSB and previously autoclaved R2YE Media A (10.3% sucrose, 1% glucose, 1.12% $MgCl_2.6H_2O$, 0.025% K_2SO_4, 0.01% Difco casamino acids, 0.5% Difco yeast extract). Production cultures were grown for up to 10 days at 30°C and 180 rpm before final collection of samples. For solid medium, 1.5% agar was added to all the above liquid media with no further modification aside from GYM, which required addition of 0.2% $CaCO_3$. *S. coelicolor* was also grown on solid soy flour mannitol (SFM) agar (2% mannitol, 2% soy flour, 2% agar). All cultures on solid medium were grown at 30°C. Where appropriate, thiostrepton (working concentration 50 μg/mL) was used for maintenance of the gene cluster in *Streptomyces*; this was not used where it may interfere with bioassay results.

M. sp. DEM32671 was inoculated from single colonies on agar plate and cultured in a seed culture medium modified from Ni et al. 2014 (1% soluble starch, 3.5% soy flour, 0.1% glucose, 0.3% $CaCO_3$)[127] for 14 days at 34°C and 220 rpm before collection. Medium 554 agar was used as a solid medium (1% glucose, 2% soluble starch, 0.5% yeast extract, 0.5% N-Z-Amine, 0.1% $CaCO_3$, 1.5% agar, pH 7.2).[128]

E. coli NEB5α, *E. coli* ET12567 pUZ8002 and *B. subtilis* strain 168 were cultured in Lysogeny Broth Miller (LB; Formedium; 1% NaCl, 1% tryptone, 0.5% yeast extract) at 37°C and 180 rpm, or on Lysogeny Broth Miller agar (Formedium; as above with 1.5% agar). When maintenance of pRes was required, Lysogeny Broth Lennox (0.5% NaCl) or Lysogeny Broth Lennox agar was used to ensure proper action of hygromycin. Where appropriate, *E. coli* was cultured with 50 μg/mL kanamycin or 25 μg/mL hygromycin.

2.5.3. Genome assembly and comparison of *M.* sp. DEM32671 to similar strains

Whole genome sequencing was performed by TGAC (Norwich, UK) using PacBio SMRT. The genome was assembled into a single contig for further

analysis. AntiSMASH 6.0 ("Antibiotic and Secondary Metabolites Analysis Shell") was used to predict the gene clusters present in the assembled genome with relaxed settings and all extra features on.[129] The Orthologous Average Nucleotide Identity Tool (OAT) was used to run OrthoANI to measure overall similarity between candidate genome sequences.[49] The Type Strain Genome Server (TYGS) was used to carry out dDDH analysis of *M.* sp. DEM32671 against other type strains in the NCBI database.[48] After using BLASTn to obtain 16S rDNA and *gyrB* gene sequences from other *Micromonospora* spp., MEGA X (Molecular Evolutionary Genetics Analysis) was used to construct multiple sequence alignments (ClustalW, gap opening penalty of 15.00 and gap extension penalty of 6.66 for both pairwise and multiple alignment) and phylogenetic trees (Maximum Likelihood).[130] iTOL was used to draw phylogenetic trees.[131]

2.5.4. Comparison of Cluster 24 to clusters from known aminoglycoside producers

AntiSMASH 6.0 analysis of *M.* sp. DEM32671 was used for first detection of Cluster 24 in the genome.[129] Manual comparison to gentamicin gene cluster BGC0000696 in the MiBIG (Minimum Information about a Biosynthetic Gene cluster) database was carried out to determine the full-length sequence of the gene cluster.[60] Clinker was first used to select cluster boundaries for similar strains by inclusion of region homologous to BGC000696 + 30 ORFs on either side and manual trimming, then secondly used to annotate similarity in cluster ORFs at thresholds of 70% and 95%.[61]

2.5.5. Insertion of Cluster 24 and pRes to *Streptomyces* spp.

BioS&T (Québec, Canada) carried out creation of high molecular weight PAC library from the *M.* sp. DEM32671 genome (average insert size of >100 kb; pESAC-13 backbone) and identified clones containing Cluster 24 by PCR

screening of ~800 bp-long regions at start, middle and end of cluster. Conjugation from *E. coli* ET12567 pUZ8002 was carried out as described by Kieser et al. in Practical *Streptomyces* Genetics.[73]

2.5.6. Determination of cluster presence in *Streptomyces coelicolor*

Screening for the presence of Cluster 24 was done using PCR amplification using Terra polymerase (Takara; cat. no. 639270). Primers used covered 1976 bp of the predicted core biosynthetic region of the gene cluster and were located in the intragenic regions of *genS1* and *genD2* homologues (c44_2kb_gntC_rev 5′-GAGAGCACGAGCGGCTTCTC-3′ and c44_2kb_gntA_for 5′-CCACCGAGCGCACCTACTAC-3′). The reaction was run on an Applied Biosystems ProFlex PCR System with the following conditions: initial denaturation 98°C for 2 min; amplification (35X cycles) 98°C for 10s, 68°C for 2 min. Agarose gel electrophoresis was used to determine that the correct product was amplified.

2.5.7. Amplification of 16S rRNA

16S rRNA gene sequence was obtained by PCR amplification using Terra polymerase (Takara; cat. no. 639270). A single colony of *M.* sp. DEM32671 grown on Medium 554 agar for 10 days was lysed in NaOH 50 mM at 98°C for 10 min before neutralising with Tris-HCl 1M pH 8; 5 μL of this was used in the final reaction (25 μL). Primers used were universal primers 27F and 1492R (27F 5′-AGAGTTTGATCCTGGCTCAG-3′; 1492R 5′-GGTTACCTTGTTACGACTT-3′). The reaction was run on an Applied Biosystems ProFlex PCR System with the following conditions: initial denaturation 98°C for 2 min; amplification (35X cycles) 98°C for 10s, 68°C for 1:30 min. Agarose gel electrophoresis was used to determine that the product amplified was clean and purification was carried out using a QIAquick PCR cleanup kit (Qiagen; cat. no. 28104). Sanger sequencing to confirm sequence validity was carried out by Eurofins

Genomics using 27F and 1492R primers. Snapgene was used to align sequence to *M.* sp. DEM32671 genome sequence to determine errors.

2.5.8. Preparation of liquid samples for bioassay

Samples for bioassay were obtained through concentration of culture supernatant via freeze drying. Cultures frozen at -20°C were thawed before centrifugation at 7000 x *g* for 5 minutes. Afterwards, an aliquot of culture supernatant was taken into a 50 mL Corning tube and snap-frozen in liquid nitrogen before freezing at -80°C for an hour. Samples were freeze dried for 22 hours before resuspension in sterile ddH_2O to 50X the original concentration.

2.5.9. Antibiotic bioassay

Indicator strain *B. subtilis* strain 168 was grown in LB-Miller broth at 37°C for 16 hours (180 rpm) before subculturing in the same medium to an OD_{600} = 0.6. At OD_{600} = 0.6 *B. subtilis* was inoculated to 50 mL LB-Miller agar in an 1/1000 dilution, before pouring to 120 mm square bioassay plates. For analysis of antibiotic production by *S. coelicolor* on solid medium, 13 mm diameter disks were cut from agar plates with lawns of strains of interest and laid on top of the indicator strain plate. For analysis of antibiotic production in liquid, a 13 mM diameter hole was cut into the indicator plate and 50 μL of concentrated culture extract was added to each. The plate was incubated for 16 hours at 37°C before photographing.

2.5.10. Generation of resistant strains

~2x10^7 CFU of *S. coelicolor* M1146 carrying Cluster 24 was inoculated into 600 μL TSB medium before plating out 100 μL to SFM agar (2% soy flour, 2% mannitol, 2% agar) at the following concentrations of gentamicin sulfate: 0, 100, 300, 500 μg/mL. These were grown for 7 days before streaking out of single

colonies on a gradient plate made as in the protocol by Weinberg, with top concentration of gentamicin being 1000 μg/mL.[132] These were grown for a further seven days before selection of 10 single colonies.

2.5.11. Testing susceptibility of resistant strains to antibiotics

For testing susceptibility of resistant strains to gentamicin in solid, 3 mL of SFM agar (+0, 200, 500, 1000, 2000, 5000 μg/mL gentamicin sulfate) was added to each well in a Costar® 24-well plate (Corning, cat. no. 3524) before inoculating with ~10^5 CFU of each spore suspension and growing at 30°C for 5 days.

For testing susceptibility of resistant strains to antibiotics in liquid, ~10^5 CFU of each spore suspension was inoculated as a seed culture to TSB (with the stated antibiotics at stated concentrations) and growth was calculated by measurement of OD_{450} of 1 mL of culture after 48 hours in a spectrophotometer. 1 mL was taken in triplicates and measurements carried out three times for each (with mixing in between) to mitigate effects of *Streptomyces* sinking in liquid culture.

2.5.12. Preparation of genomic DNA for whole-genome sequencing

Streptomyces liquid cultures were grown by inoculating ~10^7 CFU to 50 mL TSB and culturing for 48 hours at 30°C and 180 rpm before harvesting. After pelleting 5 mL from each culture (5000 x *g* for 10 minutes), cells were lysed by adding 0.5 vol lysis solution (2 mg/mL lysozyme and 50 μg/ml RNase A in 0.3 M sucrose, 25 mM Tris (pH 8.8), 25 mM EDTA (pH 8)) and incubating at 37°C for 30 minutes to allow for action of lysozyme, before freeze-thawing four times in liquid nitrogen and a 60°C water bath. An equal volume of phenol-chloroform isoamyl alcohol (25:24:1) was added and the aqueous layer removed after centrifugation (1000 x *g*, 10 minutes; 2X addition of phenol-

chloroform isoamyl alcohol and centrifugation), before DNA was precipitated with 0.6 vol isopropanol and washed with 70% ethanol.

2.5.13. Whole-genome sequencing of resistant *Streptomyces* strains and analysis of mutations

Genome sequencing (Illumina) and subsequent bioinformatic analyses (identification of closest reference genome with Kraken,[133] read-mapping with BWA-MEM for data quality assessment,[134] *de novo* contig assembly with SPAdes and BWA-MEM)[134,135] was provided by MicrobesNG (http://www.microbesng.com).

Further analysis was carried out by using BWA-MEM to align each set of reads to the *S. coelicolor* A3(2) genome with Pilon for the determination of region of difference between parental strains and resistant strains.[134,136] Regions of difference were manually investigated by aligning sequences to the parental strain with Snapgene to look at genomic context.

2.5.14. Characterisation of cluster output

Samples for LC-MS were prepared by filtering culture supernatant through a C18 SPE cartridge (Thermo Scientific 60108-701, 2000 mg bed weight, 15 mL column capacity) for partial clean-up and snap-freezing in liquid nitrogen before freeze drying. LC-MS was carried out using a Waters Atlantis T3 column (4.6 x 100 mm, 5 mm particle size) on an Agilent 1200 Series Rapid Resolution LC coupled to a Bruker maXis HR-qTOF mass spectrometer. Mobile phase: solvent A water:MeCN 90:10, solvent B water:MeCN 10:90, both with 13mM ammonium formate and 0.01% trifluoroacetic acid (TFA). The gradient composition was: 100% A for 12 minutes, 90% A 10% B for 6 seconds, 100% B for 4 minutes 54 seconds, 100% A for 4 minutes. Ionization was generated using the standard maxis ESI source adjusted to a drying gas flow

of 11 L/min at 200°C and a nebulizer pressure of 40 psig. The capillary voltage was set to 4000 V. Mass spectra were collected from 50 to 2000 m/z in positive mode. Samples were further cleaned up by using a Waters Atlantis T3 SemiPrep with a water:MeCN gradient without additives. The gradient was run from 0 to 50% MeCN over 30 min. For manual inspection of MS data, mzMatch/PeakML were used for the selection of i) peaks present only in samples from strains carrying C24; ii) selection of peaks of mass similar to adducts of gentamicin or gentamicin intermediates across all samples.[137]

1. J. O'Neill, "The Review on Antimicrobial Resistance. Antimicrobial Resistance: Tackling a Crisis for the Health and Wealth of Nations. UK Government Report," 2014.

2. M. Kumar *et al.*, "Futuristic Non-antibiotic Therapies to Combat Antibiotic Resistance: A Review," *Frontiers in Microbiology*, vol. 12, no. 609459, 2021, doi: 10.3389/FMICB.2021.609459.

3. J. Ørskov, *Investigations into the morphology of the Ray fungi*. Copenhagen: Levin & Munksgaard, 1923.

4. A. C. Parte, "LPSN—list of prokaryotic names with standing in nomenclature," *Nucleic Acids Research*, vol. 42, no. D1, pp. D613–D616, 2014, doi: 10.1093/NAR/GKT1111.

5. L. Carro *et al.*, "Uncovering the potential of novel *Micromonosporae* isolated from an extreme hyper-arid Atacama Desert soil," *Scientific Reports*, vol. 9, no. 1, pp. 1–16, 2019, doi: 10.1038/s41598-019-38789-z.

6. G. Fu *et al.*, "*Micromonospora zhangzhouensis* sp. nov., a Novel Actinobacterium Isolated from Mangrove Soil, Exerts a Cytotoxic Activity in vitro," *Scientific Reports*, vol. 10, no. 1, pp. 1–12, 2020, doi: 10.1038/s41598-020-60677-0.

7. C. Thawai, S. Tanasupawat, and T. Kudo, "*Micromonospora caldifontis* sp. nov., isolated from hot spring soil," *International Journal of Systematic and Evolutionary Microbiology*, vol. 69, no. 5, pp. 1336–1342, 2019, doi: 10.1099/IJSEM.0.003321.

8. C. Kittiwongwattana, D. Thanaboripat, C. Laosinwattana, P. Koohakan, N. Parinthawong, and C. Thawai, "*Micromonospora oryzae* sp. nov., isolated from roots of upland rice," *International Journal of Systematic and Evolutionary Microbiology*, vol. 65, no. Pt_11, pp. 3818–3823, 2015, doi: 10.1099/IJSEM.0.000500.

9. M. Ortúzar, M. E. Trujillo, B. Román-Ponce, and L. Carro, "*Micromonospora* metallophores: A plant growth promotion trait useful for bacterial-assisted phytoremediation?," *Science of The Total Environment*, vol. 739, p. 139850, 2020, doi: 10.1016/J.SCITOTENV.2020.139850.

10. M. E. Trujillo *et al.*, "*Micromonospora mirobrigensis* sp. nov.," *International Journal of Systematic and Evolutionary Microbiology*, vol. 55, no. 2, pp. 877–880, 2005, doi: 10.1099/IJS.0.63361-0.

11. C. R. Back *et al.*, "A New *Micromonospora* Strain with Antibiotic Activity Isolated from the Microbiome of a Mid-Atlantic Deep-Sea Sponge," *Marine Drugs*, vol. 19, no. 2, p. 105, 2021, doi: 10.3390/MD19020105.

12. B. Intra *et al.*, "*Micromonospora pelagivivens* sp. nov., a new species of the genus *Micromonospora* isolated from deep-sea sediment in Japan," *International Journal of Systematic and Evolutionary Microbiology*, vol. 70, no. 5, pp. 3069–3075, 2020, doi: 10.1099/IJSEM.0.004136.

13. K. Phillips, F. Zaidan, O. R. Elizondo, and K. L. Lowe, "Phenotypic characterization and 16S rDNA identification of culturable non-obligate halophilic bacterial communities from a hypersaline lake, La Sal del Rey, in extreme South Texas (USA)," *Aquatic Biosystems*, vol. 8, no. 1, pp. 1–2012, doi: 10.1186/2046-9063-8-5.

14. Y. B. Lin *et al.*, "*Micromonospora nickelidurans* sp. nov., isolated from soil from a nickel-mining site," *International Journal of Systematic and Evolutionary Microbiology*, vol. 65, pp. 4615–4620, 2015, doi: 10.1099/IJSEM.0.000623.

15. S. Nimaichand *et al.*, "*Micromonospora kangleipakensis* sp. nov., isolated from a sample of limestone quarry," *International Journal of Systematic and Evolutionary Microbiology*, vol. 63, no. Pt_12, pp. 4546–4551, 2013, doi: 10.1099/IJS.0.052746-0.

16. P. Hirsch, U. Mevs, R. M. Kroppenstedt, P. Schumann, and E. Stackebrandt, "Cryptoendolithic Actinomycetes from Antarctic Sandstone Rock Samples: *Micromonospora endolithica* sp. nov. and two Isolates Related to *Micromonospora coerulea* Jensen 1932," *Systematic and Applied Microbiology*, vol. 27, no. 2, pp. 166–174, 2004, doi: 10.1078/072320204322881781.

17. M. E. Trujillo, K. Hong, and O. Genilloud, "The Family *Micromonosporaceae*," *The Prokaryotes: Actinobacteria*, pp. 499–569, 2014, doi: 10.1007/978-3-642-30138-4_196.

18. J. Bérdy, "Bioactive Microbial Metabolites," *The Journal of Antibiotics*, vol. 58, no. 1, pp. 1–26, 2005, doi: 10.1038/ja.2005.1.

19. M. J. Weinstein *et al.*, "Gentamicin, a New Antibiotic Complex from *Micromonospora*," *Journal of Medicinal Chemistry*, vol. 6, no. 4. pp. 463–464, 1963. doi: 10.1021/jm00340a034.

20. G. H. Wagman, R. D. Gannon, and M. J. Weinstein, " Production of Vitamin B 12 by *Micromonospora* ," *Applied Microbiology*, vol. 17, no. 4, pp. 648–649, 1969, doi: 10.1128/AM.17.4.648-649.1969.

21. J. W. Park *et al.*, "Genetic dissection of the biosynthetic route to gentamicin A2 by heterologous expression of its minimal gene set.," *PNAS*, vol. 105, no. 24, pp. 8399–404, 2008, doi: 10.1073/pnas.0803164105.

22. F. Lombó *et al.*, "Deciphering the biosynthesis pathway of the antitumor thiocoraline from a marine actinomycete and its expression in two *streptomyces* species," *ChemBioChem*, vol. 7, no. 2, pp. 366–376, 2006, doi: 10.1002/CBIC.200500325.

23. Y. Chunfang, H. Chunshuai, Z. Wenjun, Z. Yiguang, and Z. Changsheng, "Heterologous Expression of Fluostatin Gene Cluster Leads to a Bioactive Heterodimer," *Organic Letters*, vol. 17, no. 21, pp. 5324–5327, 2015, doi: 10.1021/ACS.ORGLETT.5B02683.

24. K. K. Nepal and G. Wang, "Streptomycetes: Surrogate Hosts for the Genetic Manipulation of Biosynthetic Gene Clusters and Production of Natural Products," *Biotechnol Adv*, vol. 37, no. 1, p. 1, 2019, doi: 10.1016/J.BIOTECHADV.2018.10.003.

25. K. V. Baker, E. Takano, and R. Breitling, "The 'Three Cs' of Novel Antibiotic Discovery and Production through Synthetic Biology: Biosynthetic Gene Clusters, Heterologous Chassis, and Synthetic Microbial Consortia," *Advanced Biosystems*, vol. 2, no. 11, p. 1800064, 2018, doi: 10.1002/ADBI.201800064.

26. M. Xu, W. Wang, N. Waglechner, E. J. Culp, A. K. Guitor, and G. D. Wright, "GPAHex-A synthetic biology platform for Type IV-V glycopeptide antibiotic production and discovery", *Nature Communications*, vol. 11, no. 5232, 2020, doi: 10.1038/s41467-020-19138-5.

27. K. D. Bauman *et al.*, "Refactoring the Cryptic Streptophenazine Biosynthetic Gene Cluster Unites Phenazine, Polyketide, and Nonribosomal Peptide Biochemistry," *Cell Chemical Biology*, vol. 26, no. 5, pp. 724-736.e7, 2019, doi: 10.1016/j.chembiol.2019.02.004.

28. M. R. Estévez *et al.*, "Novel Fredericamycin Variant Overproduced by a Streptomycin-Resistant *Streptomyces albus* subsp. *chlorinus* Strain," *Marine Drugs*, vol. 18, no. 6, 2020, doi: 10.3390/MD18060284.

29. X. P. Shentu, Z. Y. Cao, Y. Xiao, G. Tang, K. Ochi, and X. P. Yu, "Substantial improvement of toyocamycin production in *Streptomyces diastatochromogenes* by cumulative drug-resistance mutations," *PLoS ONE*, vol. 13, no. 8, 2018, doi: 10.1371/JOURNAL.PONE.0203006.

30. J. X. Fan, Y. Song, G. Tang, K. Ochi, X. P. Shentu, and X. P. Yu, "Substantial improvement of tetraene macrolide production in *Streptomyces*

diastatochromogenes by cumulative drug resistance mutations," *PLoS ONE*, vol. 15, no. 5, 2020, doi: 10.1371/JOURNAL.PONE.0232927.

31. L. Wang *et al.*, "Efficiently activated ε-poly-L-lysine production by multiple antibiotic-resistance mutations and acidic pH shock optimization in *Streptomyces albulus*," *Microbiologyopen*, vol. 8, no. 5, 2019, doi: 10.1002/MBO3.728.

32. D. Fourmy, M. I. Recht, and J. D. Puglisi, "Binding of neomycin-class aminoglycoside antibiotics to the A-site of 16 s rRNA," *Journal of Molecular Biology*, vol. 277, no. 2, pp. 347–362, 1998, doi: 10.1006/JMBI.1997.1552.

33. M. Misumi and N. Tanaka, "Mechanism of inhibition of translocation by kanamycin and viomycin: A comparative study with fusidic acid," *Biochemical and Biophysical Research Communications*, vol. 92, no. 2, pp. 647–654, 1980, doi: 10.1016/0006-291X(80)90382-4.

34. H. Demirci, F. Murphy, E. Murphy, S. T. Gregory, A. E. Dahlberg, and G. Jogl, "A structural basis for streptomycin-induced misreading of the genetic code," *Nature Communications*, vol. 4, 2013, doi: 10.1038/NCOMMS2346.

35. K. B. Gromadski and M. v. Rodnina, "Streptomycin interferes with conformational coupling between codon recognition and GTPase activation on the ribosome," *Nat Struct Mol Biol*, vol. 11, no. 4, pp. 316–322, 2004, doi: 10.1038/NSMB742.

36. P. C. Tai, B. J. Wallace, and B. D. Davis, "Streptomycin causes misreading of natural messenger by interacting with ribosomes after initiation," *PNAS*, vol. 75, no. 1, pp. 275–279, 1978, doi: 10.1073/PNAS.75.1.275.

37. A. P. Carter, W. M. Clemons, D. E. Brodersen, R. J. Morgan-Warren, B. T. Wimberly, and V. Ramakrishnan, "Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics," *Nature*, vol. 407, pp. 340–348, 2000, doi: 10.1038/35030019.

38. N. Bilgin, A. A. Richter, M. Ehrenberg, A. E. Dahlberg, and C. G. Kurland, "Ribosomal RNA and protein mutants resistant to spectinomycin.," *The EMBO Journal*, vol. 9, no. 3, pp. 735–739, 1990, doi: 10.1002/J.1460-2075.1990.TB08167.X.

39. J. H. D. Borovinskaya, M. A., Shoji, S., Holton, J. M., Fredrick, K. & Cate, "Steric block translation caused by spectinomycin aminoglycoside," *ACS Chem. Biol.*, 2007.

40. S. Perzynski, M. Cannon, E. Cundliffe, S. B. Chahwala, and J. Davies, "Effects of Apramycin, a Novel Aminoglycoside Antibiotic on Bacterial Protein

Synthesis," *European Journal of Biochemistry*, vol. 99, no. 3, pp. 623–628, Sep. 1979, doi: 10.1111/J.1432-1033.1979.TB13295.X.

41. C. M. T. Spahn and C. D. Prescott, "Throwing a spanner in the works: antibiotics and the translation apparatus," *Journal of Molecular Medicine*, vol. 74, no. 8, pp. 423–439, Aug. 1996, doi: 10.1007/BF00217518.

42. M. A. Borovinskaya *et al.*, "Structural basis for aminoglycoside inhibition of bacterial ribosome recycling," *ACS Chemical Biology*, vol. 2, no. 8, pp. 545–552, 2007, doi: 10.1038/nsmb1271.

43. J. P. Gomez-Escribano and M. J. Bibb, "Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters," *Microbial Biotechnology*, vol. 4, no. 2, pp. 207–215, 2011, doi: 10.1111/j.1751-7915.2010.00219.x.

44. H. Kasai, T. Tamura, and S. Harayama, "Intrageneric relationships among *Micromonospora* species deduced from *gyrB*-based phylogeny and DNA relatedness.," *International Journal of Systematic and Evolutionary Microbiology*, vol. 50, no. 1, pp. 127–134, 2000, doi: 10.1099/00207713-50-1-127.

45. Curators of the DSMZ, "Leibniz Institut DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH DSM 43816," *BacDive, DSMZ*, 2021. https://bacdive.dsmz.de/index.php?site=pdf_view&id=7972&doi=doi:10.13145/bacdive7972.20201210.5 (accessed Dec. 28, 2021).

46. E. Stackebrandt and B. M. Goebel, "Taxonomic Note: A Place for DNA-DNA Reassociation and 16S rRNA Sequence Analysis in the Present Species Definition in Bacteriology," *International Journal of Systematic and Evolutionary Microbiology*, vol. 44, no. 4, pp. 846–849, 1994, doi: 10.1099/00207713-44-4-846.

47. J. P. Meier-Kolthoff, H. P. Klenk, and M. Göker, "Taxonomic use of DNA G+C content and DNA-DNA hybridization in the genomic age," *International Journal of Systematic and Evolutionary Microbiology*, vol. 64, no. Pt 2, pp. 352–356, 2014, doi: 10.1099/IJS.0.056994-0.

48. J. P. Meier-Kolthoff and M. Göker, "TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy," *Nature Communications 2019 10:1*, vol. 10, no. 1, pp. 1–10, 2019, doi: 10.1038/s41467-019-10210-3.

49. I. Lee, Y. Ouk Kim, S. C. Park, and J. Chun, "OrthoANI: An improved algorithm and software for calculating average nucleotide identity," *International Journal of Systematic and Evolutionary Microbiology*, vol. 66, no. 2, pp. 1100–1103, 2016, doi: 10.1099/IJSEM.0.000760.

50. W.-Z. Zhu, Y.-M. Ge, H.-M. Gao, J. Dai, X.-L. Zhang, and Q. Yang, "*Gephyromycinifex aptenodytis* gen. nov., sp. nov., isolated from gut of Antarctic emperor penguin *Aptenodytes forsteri*," *Antonie Van Leeuwenhoek*, vol. 114, no. 2, pp. 2003–2017, 2021, doi: 10.1007/S10482-021-01657-W.

51. S. Králová *et al.*, "*Streptomyces tardus* sp. nov.: A Slow-Growing Actinobacterium Producing Candicidin, Isolated From Sediments of the Trondheim Fjord," *Frontiers in Microbiology*, vol. 0, p. 2131, 2021, doi: 10.3389/FMICB.2021.714233.

52. L. Carro *et al.*, "Hunting for cultivable *Micromonospora* strains in soils of the Atacama Desert," *Antonie van Leeuwenhoek 2018 111:8*, vol. 111, no. 8, pp. 1375–1387, 2018, doi: 10.1007/S10482-018-1049-1.

53. S. Louca, F. Mazel, M. Doebeli, and L. W. Parfrey, "A census-based estimate of Earth's bacterial and archaeal diversity," *PLOS Biology*, vol. 17, no. 2, p. e3000106, 2019, doi: 10.1371/JOURNAL.PBIO.3000106.

54. K. Blin *et al.*, "antiSMASH 6.0: improving cluster detection and comparison capabilities," *Nucleic Acids Research*, vol. 49, no. W1, p. W29, 2021, doi: 10.1093/NAR/GKAB335.

55. A. H. Banskota *et al.*, "TLN-05220, TLN-05223, new Echinosporamicin-type antibiotics, and proposed revision of the structure of bravomicins," *The Journal of Antibiotics 2009 62:10*, vol. 62, no. 10, pp. 565–570, 2009, doi: 10.1038/ja.2009.77.

56. L. Carro *et al.*, "Genome-based classification of *Micromonosporae* with a focus on their biotechnological and ecological potential," *Scientific Reports*, vol. 8, no. 1, 2018, doi: 10.1038/S41598-017-17392-0.

57. A. Z. Boumehira, H. A. El-Enshasy, H. Hacène, E. A. Elsayed, R. Aziz, and E. Y. Park, "Recent progress on the development of antibiotics from the genus *Micromonospora*," *Biotechnology and Bioprocess Engineering 2016 21:2*, vol. 21, no. 2, pp. 199–223, May 2016, doi: 10.1007/S12257-015-0574-2.

58. H. Kase, Y. Odakura, and K. Nakayama, "Sagamicin and the related aminoglycosides: fermentation and biosynthesis. I. Biosynthetic studies with the blocked mutants of *Micromonospora sagamiensis*," *The Journal of Antibiotics*, vol. 35, no. 1, pp. 1–9, 1982, doi: 10.7164/ANTIBIOTICS.35.1.

59. M. J. Weinstein and G. H. Wagman, "Discovery and Isolation of Sisomicin," *Infection*, vol. 4, pp. 285–288, 1976.

60. S. A. Kautsar *et al.*, "MIBiG 2.0: a repository for biosynthetic gene clusters of known function," *Nucleic Acids Research*, vol. 48, pp. D454–D458, 2020, doi: 10.1093/NAR/GKZ882.

61. C. L. M. Gilchrist and Y.-H. Chooi, "clinker & clustermap.js: automatic generation of gene cluster comparison figures," *Bioinformatics*, vol. 37, no. 16, pp. 2473–2475, 2021, doi: 10.1093/BIOINFORMATICS/BTAB007.

62. NCBI Resource Coordinators, "Database Resources of the National Center for Biotechnology Information.," *Nucleic Acids Research*, vol. 45, no. D1, pp. D12–D17, 2017, doi: 10.1093/nar/gkw1071.

63. S. Li *et al.*, "Methyltransferases of gentamicin biosynthesis," *PNAS*, vol. 115, no. 6, pp. 1340–1345, 2018, doi: 10.1073/pnas.1711603115.

64. M. Khaled and M. A. Aboshanab, "Genetic studies on the biosynthesis of the major aminoglycoside antibiotics: Isolation, analysis and comparison of the biosynthetic gene clusters for 12 aminoglycoside antibiotics," PhD thesis, Bergische Wuppertal University, Wuppertal, 2005.

65. A. Reva, "Enzymology of gentamicin biosynthesis", PhD thesis, University of Cambridge, Cambridge, 2019, doi: 10.17863/CAM.25238.

66. Y. Chang *et al.*, "Overproduction of gentamicin B in industrial strain *Micromonospora echinospora* CCTCC M 2018898 by cloning of the missing genes *genR* and *genS*," *Metabolic Engineering Communications*, 2019, doi: 10.1016/j.mec.2019.e00096.

67. E. S. Kim, H. J. Hong, C. Y. Choi, and S. N. Cohen, "Modulation of actinorhodin biosynthesis in *Streptomyces lividans* by glucose repression of *afsR2* gene transcription," *Journal of Bacteriology*, vol. 183, no. 7, pp. 2198–2203, 2001, doi: 10.1128/JB.183.7.2198-2203.2001.

68. D. B. Adimpong *et al.*, "Antimicrobial Susceptibility of *Bacillus* Strains Isolated from Primary Starters for African Traditional Bread Production and Characterization of the Bacitracin Operon and Bacitracin Biosynthesis," *Applied and Environmental Microbiology*, vol. 78, no. 22, 2012, doi: 10.1128/AEM.00730-12.

69. Y. Tashiro, H. Eida, S. Ishii, H. Futamata, and S. Okabe, "Generation of small colony variants in biofilms by *Escherichia coli* harboring a conjugative F plasmid," *Microbes and Environments*, vol. 32, no. 1, pp. 40–46, 2017, doi: 10.1264/JSME2.ME16121.

70. S. Ye, B. Enghiad, H. Zhao, and E. Takano, "Fine-tuning the regulation of Cas9 expression levels for efficient CRISPR-Cas9 mediated recombination in

Streptomyces," *Journal of Industrial Microbiology & Biotechnology*, vol. 47, no. 4–5, pp. 413–423, 2020, doi: 10.1007/s10295-020-02277-5.

71. Y. Ahmed, Y. Rebets, M. R. Estévez, J. Zapp, M. Myronovskyi, and A. Luzhetskyy, "Engineering of *Streptomyces lividans* for heterologous expression of secondary metabolite gene clusters," *Microbial Cell Factories*, vol. 19, no. 1, pp. 1–16, 2020, doi: 10.1186/S12934-020-1277-8.

72. K. M. Krause, A. W. Serio, T. R. Kane, and L. E. Connolly, "Aminoglycosides: An overview," *Cold Spring Harbor Perspectives in Medicine*, vol. 6, no. 6, 2016, doi: 10.1101/cshperspect.a027029.

73. T. Kieser, M. J. Bibb, M. J. Buttner, K. F. Chater, and D. A. Hopwood, "Practical *Streptomyces* Genetics," *John Innes Foundation*, p. 529, 2000. doi: 10.4016/28481.01.

74. S. Westhoff, T. M. van Leeuwe, O. Qachach, Z. Zhang, G. P. van Wezel, and D. E. Rozen, "The evolution of no-cost resistance at sub-MIC concentrations of streptomycin in Streptomyces coelicolor," *The ISME Journal 2017 11:5*, vol. 11, no. 5, pp. 1168–1178, 2017, doi: 10.1038/ismej.2016.194.

75. S. Yoshizawa, D. Fourmy, and J. D. Puglisi, "Structural origins of gentamicin antibiotic action," *The EMBO Journal*, vol. 17, no. 22, pp. 6437–6448, Nov. 1998, doi: 10.1093/EMBOJ/17.22.6437.

76. K. S. Long and B. T. Porse, "A conserved chloramphenicol binding site at the entrance to the ribosomal peptide exit tunnel," *Nucleic Acids Research*, vol. 31, no. 24, pp. 7208–7215, 2003, doi: 10.1093/NAR/GKG945.

77. J. J. Vecchione, B. Alexander, and J. K. Sello, "Two distinct major facilitator superfamily drug efflux pumps mediate chloramphenicol resistance in Streptomyces coelicolor," *Antimicrobial Agents and Chemotherapy*, vol. 53, no. 11, pp. 4673–4677, 2009, doi: 10.1128/AAC.00853-09

78. J. E. Cronan and G. L. Waldrop, "Multi-subunit acetyl-CoA carboxylases," *Prog Lipid Res*, vol. 41, no. 5, pp. 407–435, 2002, doi: 10.1016/S0163-7827(02)00007-3.

79. S. Jitrapakdee, M. St Maurice, I. Rayment, W. W. Cleland, J. C. Wallace, and P. v. Attwood, "Structure, Mechanism and Regulation of Pyruvate Carboxylase," *Biochem J*, vol. 413, no. 3, p. 369, 2008, doi: 10.1042/BJ20080709.

80. R. G. Kranz, C. Richard-Fogal, J.-S. Taylor, and E. R. Frawley, " Cytochrome c Biogenesis: Mechanisms for Covalent Modifications and Trafficking of Heme and for Heme-Iron Redox Control ," *Microbiology and Molecular Biology Reviews*, vol. 73, no. 3, pp. 510–528, 2009, doi: 10.1128/MMBR.00001-09

81. A. A. Putnam and E. Jankowsky, "DEAD-box helicases as integrators of RNA, nucleotide and protein binding," *Biochim Biophys Acta*, vol. 1829, no. 8, pp. 884–893, 2013, doi: 10.1016/J.BBAGRM.2013.02.002.

82. R. S. Ronimus and H. W. Morgan, "The biochemical properties and phylogenies of phosphofructokinases from extremophiles," *Extremophiles*, vol. 5, no. 6, pp. 357–373, 2001, doi: 10.1007/S007920100215.

83. L. Cuthbertson and J. R. Nodwell, "The TetR Family of Regulators," *Microbiology and Molecular Biology Reviews: MMBR*, vol. 77, no. 3, p. 440, 2013, doi: 10.1128/MMBR.00018-13.

84. B. Floriano and M. Bibb, "*afsR* is a pleiotropic but conditionally required regulatory gene for antibiotic production in *Streptomyces coelicolor* A3(2)," *Molecular Microbiology*, vol. 21, no. 2, pp. 385–396, 1996, doi: 10.1046/J.1365-2958.1996.6491364.X.

85. N. Husain *et al.*, "Structural basis for the methylation of G1405 in 16S rRNA by aminoglycoside resistance methyltransferase Sgm from an antibiotic producer: a diversity of active sites in m7G methyltransferases," *Nucleic Acids Research*, vol. 38, no. 12, pp. 4120–4132, 2010, doi: 10.1093/NAR/GKQ122.

86. A. A. D. Beauclerk and E. Cundliffe, "Sites of action of two ribosomal RNA methylases responsible for resistance to aminoglycosides," *Journal of Molecular Biology*, vol. 193, no. 4, pp. 661–671, 1987, doi: 10.1016/0022-2836(87)90349-4.

87. J. Thompson, P. A. Skeggs, and E. Cundliffe, "Methylation of 16S ribosomal RNA and resistance to the aminoglycoside antibiotics gentamicin and kanamycin determined by DNA from the gentamicin-producer, *Micromonospora purpurea*," *Molecular and General Genetics*, vol. 201, no. 2, pp. 168–173, 1985, doi: 10.1007/BF00425655.

88. M. Yang, M. K. Derbyshire, R. A. Yamashita, and A. Marchler-Bauer, "NCBI's Conserved Domain Database and Tools for Protein Domain Analysis," *Current Protocols in Bioinformatics*, vol. 69, no. 1, p. e90, 2020, doi: 10.1002/CPBI.90.

89. H. L. Kim *et al.*, "Heterogeneous rRNAs are differentially expressed during the morphological development of *Streptomyces coelicolor*," *FEMS Microbiology Letters*, vol. 275, no. 1, pp. 146–152, 2007, doi: 10.1111/J.1574-6968.2007.00872.X.

90. D. Sharma, A. R. Cukras, E. J. Rogers, D. R. Southworth, and R. Green, "Mutational analysis of S12 protein and implications for the accuracy of decoding by the ribosome," *Journal of Molecular Biology*, vol. 374, no. 4, p. 1065, Dec. 2007, doi: 10.1016/J.JMB.2007.10.003.

91. I. Iscla, R. Wray, S. Wei, B. Posner, and P. Blount, "Streptomycin potency is dependent on MscL channel expression," *Nature Communications*, Sep. 2014, doi: 10.1038/ncomms5891.

92. E. Severi and G. H. Thomas, "Antibiotic export: Transporters involved in the final step of natural product production," *Microbiology (United Kingdom)*, vol. 165, no. 8, pp. 805–818, 2019, doi: 10.1099/MIC.0.000794

93. S. Li *et al.*, "Mechanistic Insights into Dideoxygenation in Gentamicin Biosynthesis," *ACS Catalysis*, vol. 11, no. 19, pp. 12274–12283, 2021, doi: doi.org/10.1021/acscatal.1c03508.

94. A. C. Reis and H. M. Salis, "An automated model test system for systematic development and improvement of gene expression models," *ACS Synthetic Biology*, vol. 9, no. 11, pp. 3145–3156, 2020, doi: 10.1021/ACSSYNBIO.0C00394

95. V. Solovyev and A. Salamov, "Automatic annotation of microbial genomes and metagenomic sequences. In metagenomics and its applications in agriculture, biomedicine and environmental studies (Ed. R.W. Li).," *Nova Science Publishers*, 2011.

96. A. Rhoads and K. F. Au, "PacBio Sequencing and Its Applications," *Genomics, Proteomics & Bioinformatics*, vol. 13, no. 5, pp. 278–289, 2015, doi: 10.1016/J.GPB.2015.08.002.

97. M. Eisenstein, "Oxford Nanopore announcement sets sequencing sector abuzz," *Nature Biotechnology*, vol. 30, no. 4, pp. 295–296, 2012, doi: 10.1038/NBT0412-295.

98. H. Lu, F. Giordano, and Z. Ning, "Oxford Nanopore MinION Sequencing and Genome Assembly," *Genomics, Proteomics & Bioinformatics*, vol. 14, no. 5, pp. 265–279, 2016, doi: 10.1016/J.GPB.2016.05.004.

99. A. D. Tyler *et al.*, "Evaluation of Oxford Nanopore's MinION Sequencing Device for Microbial Whole Genome Sequencing Applications," *Scientific Reports 2018 8:1*, vol. 8, no. 1, pp. 1–12, 2018, doi: 10.1038/s41598-018-29334-5.

100. O. Genilloud *et al.*, "Current approaches to exploit actinomycetes as a source of novel natural products", doi: 10.1007/s10295-010-0882-7.

101. R. Tu *et al.*, "Droplet-based microfluidic platform for high-throughput screening of *Streptomyces*," *Communications Biology*, vol. 4, no. 1, pp. 1–9, 2021, doi: 10.1038/s42003-021-02186-y.

102. Q. Tong, Y. Li, S. Wang, and S. Yan, "High-Throughput Screening of *Streptomyces virginiae* Strains Using Microtiter Plates for the High-Titer

Production of Virginiamycin," *Analytical Letters*, vol. 53, no. 8, pp. 1233–1246, 2019, doi: 10.1080/00032719.2019.1700516.

103. C. Zhang, M. R. Seyedsayamdost, "Discovery of a Cryptic Depsipeptide from *Streptomyces ghanaensis* via MALDI-MS-Guided High-Throughput Elicitor Screening," *Angewandte Chemie International Edition*, vol. 59, no. 51, pp. 23005–23009, 2020, doi: 10.1002/ANIE.202009611.

104. J. W. Park *et al.*, "Genetic dissection of the biosynthetic route to gentamicin A2 by heterologous expression of its minimal gene set.," *PNAS*, vol. 105, no. 24, pp. 8399–404, 2008, doi: 10.1073/pnas.0803164105.

105. B. W. Bycroft and D. J. Payne, "Dictionary of antibiotics and related substances". 2nd ed. Boca Raton, CRC Press, 2013.

106. Y. H. Ban *et al.*, "Complete reconstitution of the diverse pathways of gentamicin B biosynthesis," *Nature Chemical Biology*, vol. 15, pp. 295–303, 2019, doi: 10.1038/s41589-018-0203-4.

107. F. Farouk, H. M. E. Azzazy, and W. M. A. Niessen, "Challenges in the determination of aminoglycoside antibiotics, a review," *Anal Chim Acta*, vol. 890, pp. 21–43, 2015, doi: 10.1016/J.ACA.2015.06.038.

108. B. Blanchaert, S. Huang, K. Wach, E. Adams, and A. van Schepdael, "Assay Development for Aminoglycosides by HPLC with Direct UV Detection," *Journal of Chromatographic Science*, vol. 55, no. 3, pp. 197–204, 2017, doi: 10.1093/CHROMSCI/BMW169.

109. S. B. Turnipseed, S. B. Clark, C. M. Karbiwnyk, W. C. Andersen, K. E. Miller, and M. R. Madson, "Analysis of aminoglycoside residues in bovine milk by liquid chromatography electrospray ion trap mass spectrometry after derivatization with phenyl isocyanate," *J Chromatogr B Analyt Technol Biomed Life Sci*, vol. 877, no. 14–15, pp. 1487–1493, 2009, doi: 10.1016/J.JCHROMB.2009.03.025.

110. E. Kaale, C. Govaerts, J. Hoogmartens, and A. van Schepdael, "Mass spectrometric study to characterize thioisoindole derivatives of aminoglycoside antibiotics," *Rapid Commun Mass Spectrom*, vol. 19, no. 20, pp. 2918–2922, 2005, doi: 10.1002/RCM.2152.

111. R. Sardella *et al.*, "Optimized Extraction of Amikacin from Murine Whole Blood," *Molecules*, vol. 26, no. 3, 2021, doi: 10.3390/MOLECULES26030665.

112. M. A. Omar, D. M. Nagy, M. A. Hammad, and A. A. Aly, "Highly Sensitive Spectrofluorimetric Method for Determination of Certain Aminoglycosides in

Pharmaceutical Formulations and Human Plasma," *AAPS PharmSciTech*, vol. 14, no. 2, p. 828, 2013, doi: 10.1208/S12249-013-9969-6.

113. A.-Z. A. Abou-Zeid, A. El-Wahab, I. Eissa, and H. M. Salem, "The fermentative production of gentamicins by *Micromonospora purpurea*," *Journal of Applied Chemistry and Biotechnology*, vol. 24, no. 11, pp. 655–661, 1974, doi: 10.1002/JCTB.2720241107.

114. H.-J. Nah, H.-R. Pyeon, S.-H. Kang, S.-S. Choi, and E.-S. Kim, "Cloning and Heterologous Expression of a Large-sized Natural Product Biosynthetic Gene Cluster in *Streptomyces* Species," *Frontiers in Microbiology*, vol. 8, no. Mar, p. 394, 2017, doi: 10.3389/fmicb.2017.00394.

115. C. Olano *et al.*, "Activation and identification of five clusters for secondary metabolites in *Streptomyces albus* J1074.," *Microbial Biotechnology*, vol. 7, no. 3, pp. 242–56, May 2014, doi: 10.1111/1751-7915.12116.

116. R. Novakova *et al.*, "Increased heterologous production of the antitumoral polyketide mithramycin A by engineered *Streptomyces lividans* TK24 strains," *Applied Microbiology and Biotechnology*, vol. 102, no. 2, pp. 857–869, 2018, doi: 10.1007/s00253-017-8642-5.

117. Z. Wu, W. Gao, S. Zhou, Z. Wen, X. Ni, and H. Xia, "Improving gentamicin B and gentamicin C1a production by engineering the glycosyltransferases that transfer primary metabolites into secondary metabolites biosynthesis," *Microbiological Research*, vol. 203, pp. 40–46, Oct. 2017, doi: 10.1016/J.MICRES.2017.06.006.

118. B. P. Anton and E. A. Raleigh, "Complete Genome Sequence of NEB 5-alpha, a Derivative of *Escherichia coli* K-12 DH5α.," *Genome Announcements*, vol. 4, no. 6, Nov. 2016, doi: 10.1128/genomeA.01245-16.

119. D. J. MacNeil, K. M. Gewain, C. L. Ruby, G. Dezeny, P. H. Gibbons, and T. MacNeil, "Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector," *Gene*, vol. 111, no. 1, pp. 61–68, 1992, doi: 10.1016/0378-1119(92)90603-M.

120. P. R. Burkholder and N. H. Giles, "Induced biochemical mutations in *Bacillus subtilis*.," *Am J Bot*, vol. 34, no. 6, pp. 345–8, 1947.

121. F. Kunst *et al.*, "The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*.," *Nature*, vol. 390, no. 6657, pp. 249–256, 1997, doi: 10.1038/36786.

122. O. Hara and T. Beppu, “Mutants blocked in streptomycin production in *Streptomyces griseus* - the role of A-factor.,” *J Antibiot (Tokyo)*, vol. 35, no. 3, pp. 349–58, 1982.

123. C. Howden, “Generating New Antibiotics by Cloning Antibiotic Biosynthetic Gene Clusters,” Thesis (MSc), University of Newcastle, Newcastle, 2017.

124. S. Donadio, P. Monciardini, and M. Sosio, “Chapter 1 Approaches to Discovering Novel Antibacterial and Antifungal Agents,” *Methods in Enzymology*, vol. 458, no. A, pp. 3–28, 2009, doi: 10.1016/S0076-6879(09)04801-0.

125. M. G. Malmierca *et al.*, “Searching for Glycosylated Natural Products in Actinomycetes and Identification of Novel Macrolactams and Angucyclines,” *Frontiers in Microbiology*, vol. 9, no. 39, 2018, doi: 10.3389/FMICB.2018.00039.

126. I. Voitsekhovskaia *et al.*, “Baikalomycins A-C, New Aquayamycin-Type Angucyclines Isolated from Lake Baikal Derived *Streptomyces* sp. IB201691-2A,” *Microorganisms*, vol. 8, no. 5, p. 680, 2020, doi: 10.3390/MICROORGANISMS8050680.

127. X. Ni, Z. Sun, H. Zhang, H. He, Z. Ji, and H. Xia, “Genetic engineering combined with random mutagenesis to enhance G418 production in *Micromonospora echinospora*,” *Journal of Industrial Microbiology and Biotechnology*, vol. 41, no. 9, pp. 1383–1390, Sep. 2014, doi: 10.1007/s10295-014-1479-3.

128. DSMZ, “Medium 554 | BacMedia,” Sep. 2021. https://bacmedia.dsmz.de/medium/554?bacdive=132377 (accessed Nov. 03, 2021).

129. K. Blin *et al.*, “antiSMASH 6.0: improving cluster detection and comparison capabilities,” *Nucleic Acids Research*, vol. 49, no. W1, pp. W29–W35, 2021, doi: 10.1093/NAR/GKAB335.

130. S. Kumar, G. Stecher, M. Li, C. Knyaz, and K. Tamura, “MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms,” *Molecular Biology and Evolution*, vol. 35, no. 6, pp. 1547–1549, 2018, doi: 10.1093/MOLBEV/MSY096.

131. I. Letunic and P. Bork, “Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation,” *Nucleic Acids Research*, vol. 49, no. W1, pp. W293–W296, 2021, doi: 10.1093/NAR/GKAB301.

132. E. D. Weinberg, “Gradient Agar Plates,” *The American Biology Teacher*, vol. 21, no. 8, pp. 347–350, 1959, doi: 10.2307/4439185.

133. D. E. Wood and S. L. Salzberg, “Kraken: Ultrafast metagenomic sequence classification using exact alignments,” *Genome Biology*, vol. 15, no. 3, pp. 1–12, 2014, doi: 10.1186/GB-2014-15-3-R46/FIGURES/5.

134. H. Li, “Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM,” 2013, Accessed: Nov. 12, 2021. [Online]. Available: http://github.com/lh3/bwa.

135. A. Bankevich *et al.*, “SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing,” *Journal of Computational Biology*, vol. 19, no. 5, p. 455, 2012, doi: 10.1089/CMB.2012.0021.

136. B. J. Walker *et al.*, “Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement,” *PLOS ONE*, vol. 9, no. 11, p. e112963, 2014, doi: 10.1371/JOURNAL.PONE.0112963.

137. R. A. Scheltema, A. Jankevics, R. C. Jansen, M. A. Swertz, and R. Breitling, “PeakML/mzMatch: A File Format, Java Library, R Library, and Tool-Chain for Mass Spectrometry Data Analysis,” *Analytical Chemistry*, vol. 83, no. 7, pp. 2786–2793, 2011, doi: 10.1021/AC2000994.

Table S2.1: Top 100 hits on the NCBI repository when using *M.* sp. DEM32671 16S rRNA as a query sequence

Scientific Name	Query Cover	E value	% identity	Accession
Micromonospora echinospora	100%	0	99.66	LT607413.1
Micromonospora echinospora	100%	0	99.66	AY524043.1
Micromonospora echinospora	100%	0	99.66	AJ628149.4
Micromonospora sp.	99%	0	99.66	MH299476.1
Micromonospora sp. 27021/10ATCC11	100%	0	99.59	JQ836684.1
Micromonospora sp. GUI28	100%	0	99.59	FN658643.1
Micromonospora sp. NEAU-ZJ3-9	100%	0	99.52	KF982697.2
Micromonospora sagamiensis subsp. *flava*	100%	0	99.52	FN552409.1
Micromonospora echinospora	100%	0	99.52	X92597.1
Micromonospora sp. UMM500	98%	0	99.86	AY552775.1
Micromonospora sp. 1A01699	100%	0	99.39	EF059959.1
Micromonospora sagamiensis	100%	0	99.39	AP023438.1
Micromonospora sp. MKTS11	99%	0	99.52	KU382314.1
Micromonospora sp. NEAU-h5	100%	0	99.32	KC439460.1
Micromonospora carbonacea	100%	0	99.32	CP058905.1
Micromonospora sp.	99%	0	99.52	MN625942.1
Micromonospora sp.	99%	0	99.52	MG770739.1
Micromonospora echinospora	99%	0	99.32	NR_044883.1
Micromonospora sp. NEAU-CYQ1	100%	0	99.25	KC344362.1
Micromonospora sp. 26021/10ATCC6	100%	0	99.25	JQ836670.1
Micromonospora chersina	100%	0	99.25	EF552206.1
Micromonospora sp.	99%	0	99.38	MG770833.1
Micromonospora citrea	99%	0	99.18	NR_044886.1
Micromonospora sp. 22MKTSM	99%	0	99.45	KU382260.1
Micromonospora sp. 2803GPT1-10	100%	0	99.11	JQ836674.1
Micromonospora sp.	100%	0	99.11	MN058278.1
Micromonospora sp. B026	100%	0	99.11	KP009553.1
Micromonospora inyonensis	100%	0	99.11	JF431003.1
Micromonospora sp. 2701GPT1-10	100%	0	99.11	JQ836680.1
Micromonospora sp. NEAU-N1	100%	0	99.11	JQ073727.1
Micromonospora chersina	99%	0	99.11	NR_044892.1
Micromonospora sp.	99%	0	99.11	MT527686.1
Micromonospora sp.	99%	0	99.18	LC383890.1
Micromonospora sp.	99%	0	99.31	MG770631.1
Micromonospora sp. 2802OA10	100%	0	99.05	JQ867355.1
Micromonospora sp. 0138	100%	0	99.05	AF367587.1
Micromonospora coerulea	99%	0	99.18	AJ560637.1
Micromonospora coerulea	99%	0	99.18	AJ560636.1
Micromonospora sp.	98%	0	99.45	MT374866.1
Micromonospora fluminis	100%	0	99.05	LR130241.1

Micromonospora sp.	99%	0	99.31	MG770630.1
Micromonospora echinofusca	100%	0	99.05	LT607733.1
Micromonospora auratinigra	100%	0	99.05	LT594323.1
Actinobacterium J93	99%	0	99.18	KP292604.1
Micromonospora sp. NEAU-ycm3	100%	0	99.05	KC161228.1
Micromonospora inyonensis	100%	0	99.05	FJ160413.1
Micromonospora sagamiensis	100%	0	99.05	NR_044890.1
Micromonospora sp.	99%	0	99.25	MN812695.1
Micromonospora pallida	99%	0	99.04	NR_044884.1
Micromonospora terminaliae	100%	0	98.98	CP045309.1
Micromonospora endolithica	99%	0	99.11	NR_027200.1
Micromonospora sp.	100%	0	98.98	MT527692.1
Micromonospora sp.	98%	0	99.58	MT374906.1
Micromonospora sp.	98%	0	99.51	MT374890.1
Micromonospora sp.	98%	0	99.51	MT374884.1
Micromonospora tulbaghiae	100%	0	98.98	CP024087.1
Micromonospora sp. K2-05	100%	0	98.98	KU289076.1
Micromonospora soli	100%	0	98.98	NR_146360.1
Micromonospora sp. 27021/10ATCC9	100%	0	98.98	JQ836683.1
Micromonospora sp. 213425	100%	0	98.98	FJ263420.1
Micromonospora sp. CNS-767_SD06	99%	0	99.11	EU214980.1
Micromonospora sp.	98%	0	99.58	MT374926.1
Micromonospora sp.	98%	0	99.45	MT374886.1
Micromonospora sp.	98%	0	99.58	MT374885.1
Micromonospora sp. S2905	99%	0	99.24	KJ782561.1
Micromonospora sp. EUSN1S8	99%	0	98.98	LC085581.1
Micromonospora chersina	100%	0	98.91	AB648998.1
Micromonospora sp. 6	100%	0	98.91	JF905621.1
Micromonospora auratinigra	100%	0	98.91	NR_028659.1
Micromonospora sp.	98%	0	99.31	LC487778.1
Micromonospora sp. B006	100%	0	98.91	CP030865.1
Micromonospora endolithica	98%	0	99.24	KX354312.1
Micromonospora sp. 213425	100%	0	98.91	FJ263421.1
Micromonospora sp. 213431	100%	0	98.91	FJ263414.1
Micromonospora sp.	98%	0	99.44	MT374888.1
Micromonospora sp.	99%	0	99.11	MG770749.1
Micromonospora sp. EUST2H12	99%	0	99.04	LC085582.1
Micromonospora chersina	99%	0	98.91	LC128359.1
Actinobacterium R48	99%	0	98.97	KP638485.1
Micromonospora sp. A4065	100%	0	98.91	JN989294.1
Micromonospora chersina	99%	0	98.91	FJ756552.1
Micromonospora sp.	99%	0	99.11	MG770804.1
Micromonospora sp.	99%	0	99.11	MG770792.1
Micromonospora inositola	100%	0	98.84	LT607754.1
Micromonospora echinaurantiaca	100%	0	98.84	LT607750.1
Micromonospora sp. HCI39	99%	0	99.11	KT004657.1
Micromonospora sp. NEAU-lc7	100%	0	98.84	KC287243.1
Micromonospora auratinigra	98%	0	99.38	FJ547133.1

Micromonospora sp. CNP-847_SD01	99%	0	99.11	EU214924.1
Micromonospora inyonensis	100%	0	98.84	NR_044893.1
Micromonospora echinospora	100%	0	98.84	MF467904.1
Micromonospora sp. MSM11	99%	0	99.04	KU382315.1
Micromonospora narathiwatensis	100%	0	98.84	LT594324.1
Micromonospora sp. S07	99%	0	98.91	KP903370.1
Micromonospora sp. NEAU-zt13	100%	0	98.84	KC287245.1
Micromonospora sp. 2603GPT1-5	100%	0	98.84	JQ836671.1
Micromonospora sp. 20	100%	0	98.84	FJ205724.1
Micromonospora sp. 210901	100%	0	98.84	FJ261958.1
Micromonospora sp. CNQ-083_SD01	99%	0	98.91	EU214939.1

Table S2.2: Biosynthetic gene clusters identified in the genome of *M.* sp. DEM32671 after analysis with AntiSMASH 6.0. Cluster 24 is highlighted in blue-green as it is the subject of this work.

Region	Type	Cluster size (Kb)	Most similar known cluster	Similarity to most similar known cluster (%)
1	butyrolactone	11.01	-	-
2	RiPP-like	8.11	-	-
3	NRPS	56.17	-	-
4	T1PKS, T2PKS, PKS-like	73.91	TLN-05220	86
5	NRPS-like, NRPS, T1PKS	64.58	bleomycin	15
6	NRPS-like, T1PKS, NRPS, T3PKS	106.52	stigmatellin	40
7	TK3PKS, NRPS	51.18	feglymycin	47
8	NRPS, PKS-like	55.11	kedarcidin	4
9	betalactone, NRPS, T1PKS, RiPP-like	94.35	fosfomycin	13
10	T2PKS, arylpolyene	72.53	WS79089A/hexaricin B/hexaricin C	15
11	NRPS, lanthipeptide class III, lanthipeptide class II, RiPP-like	90.59	gobichelin A/gobichelin B	33
12	NRPS	52.73	thiolutin	12
13	RiPP-like, lanthipeptide class III	29.48	SapB	75
14	NRPS-like, T1PKS, NRPS, nucleoside, lassopeptide	107.40	muraymycin C1	21
15	T1PKS, NRPS-like, NRPS	235.77	rifamycin	38
16	other, T1PKS	49.71	tetronasin	9
17	T1PKS	52.03	butyrolactol A	33
18	NRPS, T1PKS	49.26	-	-
19	RiPP-like	10.86	lymphostin/neolymphostinol B/lymphostinol/neolymphostin B	23
20	terpene	19.33	-	-
21	T3PKS	40.50	herbimycin A	10
22	NRPS-like	43.62	-	-
23	T3PKS	41.05	alkyl-O-dihydrogeranyl-methoxyhydroquinones	57
24	aminoglycoside	21.19	gentamicin	23
25	lanthipeptide class I	25.21	glycopeptidolipid	20
26	terpene	19.35	isorenieratene	25
27	terpene	20.96	phosphonoglycans	3
28	siderophore	11.88	desferrioxamine E	100
29	lanthipeptide class III	22.59	-	-
30	T1PKS	52.42	vazabitide A	4

31	N-acetylglutaminylglutamine amide	14.87	-	-

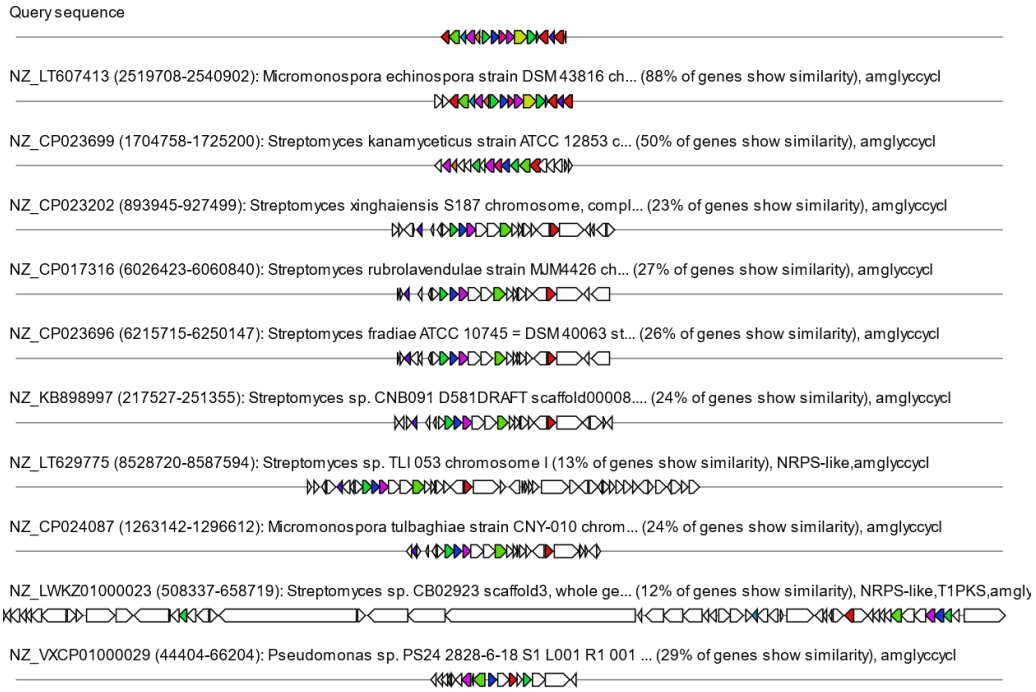


Figure S2.1: Most similar clusters to C24 determined by ClusterBlast in antiSMASH 6.0. Similar genes are labelled in same colour scheme as query sequence (automatically annotated by antiSMASH 6.0).

Table S2.3: BLASTp analysis of Cluster 24 homologues in comparison to gentamicin genes from *M. echinospora* DSM 43816.

Gene in *M.* sp. DEM32671	Gene in *M. echinospora* DSM 43816	Identities %	Positives %	E
genO	*genO*	99 (384/385)	100 (385/385)	0
gmrB	*gmrB*	98 (256/261)	99 (260/261)	1E-174
genB1	*genB1*	96 (346/359)	97 (351/359)	0
genQ	*genQ*	94 (475/508)	96 (490/508)	0
genD3	*genD3*	97 (268/276)	97 (268/276)	2E-94
genM1	*genM1*	99 (397/403)	99 (399/403)	0
gmrA	*gmrA*	96 (258/270)	98 (266/270)	5E-151
genS1	*genS1*	98 (415/423)	99 (421/423)	0
genC	*genC*	99 (394/398)	99 (396/398)	0
genD2	*genD2*	98 (319/324)	99 (321/324)	0
genM2	*genM2*	99 (417/423)	99 (421/423)	0
genD1	*genD1*	96 (635/660)	99 (654/660)	0
genS2	*genS2*	99 (415/419)	99 (418/419)	0
genW	*genW*	100 (372/372)	100 (372/372)	3E-82
genB4	*genB4*	97% (433/446)	98% (440/446)	0
genP	*genP*	98% (265/271)	99% (269/271)	3E-178
genB3	*genB3*	98% (440/451)	99% (449/451)	0
genK	*genK*	98% (625/636)	99% (632/636)	0
genB2	*genB2*	99% (411/415)	99% (413/415)	0
genX	*genX*	98% (167/171)	98% (169/171)	2E-97
genU	*genU*	99% (295/299)	99% (297/299)	0
genV	*genV*	97%(420/433)	98% (426/433)	6E-170
genE	*genE*	98% (301/307)	99%(305/307)	0
genY	*genY*	82% (376/459)	89%(411/459)	0
genA	*genA*	98%(222/226)	99%(225/226)	6E-150
genF	*genF*	98% (211/216)	98%(213/216)	7E-117
genG	*genG*	99% (117/118)	100%(118/118)	2E-80
genH	*genH*	96% (1092/1139)	97%(1110/1139)	0
genI	*genI*	99% (626/634)	99% (632/634)	0
genT	*genT*	96% (291/303)	97% (296/303)	0
genD	*genD*	95% (166/174)	97% (170/174)	1E-108
genN	*genN*	96% (298/310)	97% (303/310)	0
genJ	*genJ*	93% (313/336)	94% (319/336)	4E-177
genK2	*genK2*	96% (341/354)	97% (345/354)	0

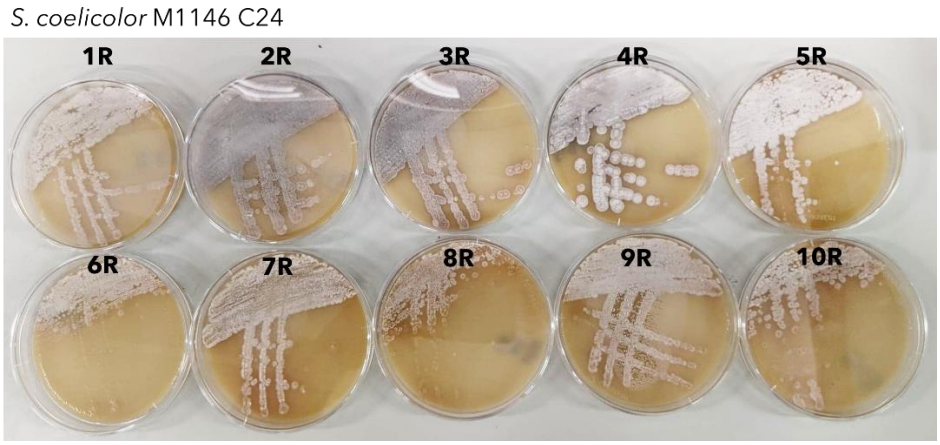


Figure S2.2: Streak plates of *S. coelicolor* M1146 C24 cohort of resistant strains on SFM + 500 µg/mL gentamicin sulfate, photographed after 7 days growth at 30°C.

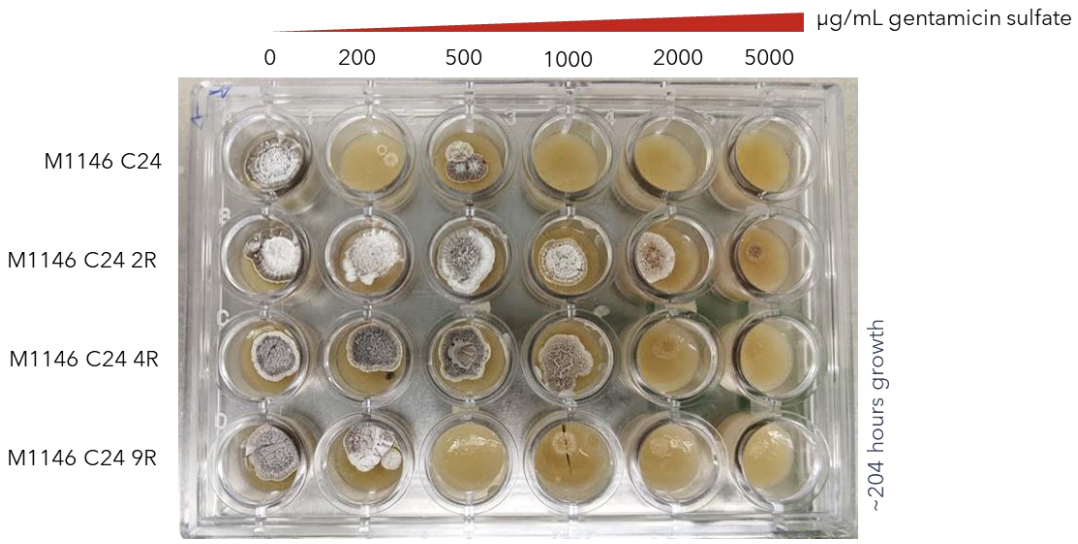


Figure S2.3: Measurement of *S. coelicolor* M1146 C24 and mutant derivative strains' resistance to gentamicin grown on SFM agar. Image of further strain growth (shown in **Figure 2.14**) after a further ~108 hours, ~204 hours total.

Table S2.4: Assembly statistics provided by MicrobesNG for the contig assembly of reads from *S. coelicolor* M1146 C24, *S. coelicolor* M1146 C24 2R, *S. coelicolor* M1146 C24 4R. Where appropriate, data is heatmapped from best (dark blue) through to worst (dark red).

Statistics without reference	***S. coelicolor* M1146 C24**	***S. coelicolor* M1146 C24 2R**	***S. coelicolor* M1146 C24 4R**
# contigs	145	68	72
# contigs (>= 0 bp)	274	100	133
# contigs (>= 1000 bp)	110	63	61
# contigs (>= 5000 bp)	88	52	52
# contigs (>= 10000 bp)	81	48	49
# contigs (>= 25000 bp)	65	40	42
# contigs (>= 50000 bp)	49	35	40
Largest contig	411176	622656	622681
Total length	8516642	8471998	8501097
Total length (>= 0 bp)	8569568	8482497	8520525
Total length (>= 1000 bp)	8494265	8468981	8493650
Total length (>= 5000 bp)	8450327	8443766	8477608
Total length (>= 10000 bp)	8394442	8415140	8450585
Total length (>= 25000 bp)	8154885	8293549	8345937
Total length (>= 50000 bp)	7587243	8135289	8285077
N50	176756	324513	257413
N75	96794	168725	164481
L50	17	11	12
L75	33	19	23
GC (%)	72.17	72.21	72.2
Mismatches			
# N's	0	0	0
# N's per 100 kbp	0	0	0

Table S2.5: Mutations in coding regions shared between *S. coelicolor* M1146 C24 2R and 4R. Predicted functions were attributed using BLASTp and the NCBI Conserved Domains Database.[88]

Shared mutations in *S. coelicolor* M1146 C24 2R and *S. coelicolor* M1146 C24 4R coding regions			
DNA mutation	**DNA mutation site**	**AA mutation**	**Predicted function**
Deletion of GG	282264; contig 2	Frame shift	Hypothetical protein, no putative domains
A -> G	337200; contig 3	D701G	MerR family transcriptional regulator
G -> C	100154; contig 4	Silent S6S	D-inositol-3-phosphate glycosyltransferase
T -> G	347284; contig 4	Silent D22D	L, D-transpeptidase
A -> T	61192; contig 5	C371S	Sensor histidine kinase
G -> C	61310; contig 5	E331D	
A -> T	61320; contig 5	F328Y	
G -> T	61331; contig 5	Silent V324V	
G -> T	61335; contig 5	A323D	
G -> C	61344; contig 5	A320G	
G -> T	61369; contig 5	H312N	
Insertion of AGCACCA	147360; contig 5	L-V-L insertion (position 508)	ABC-transporter
G -> T	153854; contig 5	H191N	SidA/IucD/PvdA family monooxygenase
G -> C	246878; contig 5	E240D	cytochrome p450
G -> T	288480; contig 5	D95E	Hypothetical protein, no putative domains
C -> A	292529; contig 5	R21L	FAD-dependent oxidoreductase
deleted cytosine	303659; contig 5	Frame shift from A555	Hypothetical protein, no putative domains
55 bp region of insertion	151581 - 151634; contig 6	Frame shift from H51	LLM class flavin-dependent oxidoreductase
T -> A	177997; contig 10	H88L	ABC-type multidrug transporter

G -> C	178025; contig 10	P79A	
6bp deletion	149348 - 149353; contig 12	Deletion of T231, G232	ribonuclease III
C -> A	130647; contig 16	G93V	glycerate 2-kinase
165 bp insertion	31432 - 31450; contig 17	Extra amino acids added to hypothetical protein	ADP-ribosylglycohydrolase family protein
C -> A	162805; contig 20	Silent G298G	acetoin utilization protein AcuC
T -> G	43114; contig 21	Silent A367A	SWIM zinc finger family protein
A -> C	43126; contig 21	D363E	
A -> T	100277; contig 21	S221T	ABC transporter permease subunit
C -> A	134615; contig 21	A580S	glycoside hydrolase family C-terminal domain-containing protein
G -> C	59721; contig 24	First mutation frame shifts rest of coding sequence of hypothetical protein	Hypothetical protein, no putative domains
G -> T	59728; contig 24		
G -> C	59738; contig 24		
T -> G	59761; contig 24		
T -> G	59764; contig 24		
17 bp Insertion	59768; contig 24		
1281 bp insertion	24372; contig 26	In frame addition of bases	Hypothetical protein
635 bp insertion	110492; contig 26	Frame shift	Hypothetical protein
T -> A	12968; contig 31	Removes stop codon allowing for translation read though to next CDS along	Both CDS are hypothetical proteins with no putative annotated domains
9 bp region of mutation	11139 - 11147; contig 34	E27C, G28P, A29F	cytochrome p450
A -> T	29944; contig 34	L34H	ADP-binding cassette domain-containing protein
948 bp insertion	69612; contig 35	In frame addition of bases at position 1231	Type I PKS
C -> A	77080; contig 35	P37T	
G -> T	58670; contig 36	P30H	CBS pair and BON domain-containing protein
T -> A	58676; contig 36	H28L	

Insertion of 167 bp	45118; contig 37	Frame shift from V51	Hypothetical protein, no putative domains
CGTGGC -> CGCGGT	51700 - 51706; contig 40	Silent R544R, silent G545G	DEAD box helicase
G -> T	11837; contig 42	P720H	CHAT domain-containing protein
C -> A	11558; contig 42	R663L	
C -> A	36071; contig 45	T379L	serine/threonine protein kinase
65 bp changes across single coding sequence	13 - 1062; contig 46	G74A, silent S75S, silent G78G, A81F, L82N, R85M, H98L, D89A, G91N, silent G96G, V105I, silent R123R, silent A124A, silent A125A, V126T, E130D, silent G131G, N134D, silent S136S, silent V138V, silent P139P, silent V140V, V141K, silent R143R, S144A, silent G145G, silent L146L, V147T, silent L149L, R151H, M155P, silent R156R, E158G, D162E, silent S16S, silent L164L, silent A165A, silent G216G, D218K, silent V237V, silent S238S, A240S, E241D, silent P242P, silent I264I, silent D265D, silent G299G	acetyl-CoA C-acetyltransferase
C -> A	3809; contig 47	L263M	VWA domain-containing protein
C -> A	17539; contig 47	P156T	ROK family protein
C -> A	34448; contig 47	G73V	Zinc-binding dehydrogenase
C -> G	34502; contig 47	Silent G68G	
T -> G	57893; contig 47	Silent R275R	SGNH/GDSL hydrolase family protein
T -> A	55019; contig 48	Q24L	Gln to Leu in hypothetical protein
Region of difference CCGGCGCGGCG to ACGGCGCGGAC	13690 - 13700; contig 53	A604D, A607D	DNA ligase
C -> A	34681; contig 57	R11L	Hypothetical protein, no putative domains
G -> T	13312; contig 65	E184*	Hypothetical protein, no putative domains

Table S2.6: Mutations in intergenic sequences shared between *S. coelicolor* M1146 C24 2R and 4R. Predicted functions were attributed using BLASTp and the NCBI Conserved Domains Database.[88]

Shared mutations in *S. coelicolor* M1146 C24 2R and *S. coelicolor* M1146 C24 4R intergenic regions			
DNA mutation	**DNA mutation site**	**Genes flanking**	**Notes**
Deletion of guanine	10431; contig 1	putative membrane transporter, glycoside hydrolase	Possible disruption of terminator (9 run guanine with mutation)
A -> C	410229; contig 1	amino acid permease, biotin transporter	Possible disruption of terminator
G -> T	410240; contig 1		
A -> C	283335; contig 5	rhodanese-like domain-containing protein, TetR family transcriptional regulator	No likely effect on expression of hypothetical proteins
C -> G	285751; contig 5	MFS transporter, twin-arginine translocation signal domain-containing protein	Possible disruption of terminator
C -> G	69305; contig 6	alpha-lytic protease prodomain-containing protein, serine hydrolase	Possible disruption of terminator
G -> T	187649; contig 6	glycosyl hydrolase-related protein, cellulase family glycosylhydrolase	Possible disruption of promoter of hypothetical protein
G -> T	187663; contig 6		
72 bp region of insertion	50082 - 50153; contig 6	NADP-dependent phosphogluconate dehydrogenase, aspartate/glutamate racemase family protein	Unlikely to cause any effect on surrounding genes
13 bp insertion	128 - 140; contig 7	none annotated upstream, PTS glucose transporter subunit IIA	Unlikely to cause any effect on surrounding genes
46 bp insertion	53256; contig 9	excinuclease ABC subunit UvrA, helix-turn-helix domain-containing protein	Possible disruption of hypothetical protein
204 bp insertion	153775; contig 9	LPXTG cell wall anchor domain-containing protein, pyridoxal phosphate-dependent protein CobC	Should not affect surrounding genes. BLASTed and no significant similarity found for inserted sequence
G -> C	171845; contig 9	cobalamin biosynthesis protein, hypothetical protein	Unlikely to cause any effect on surrounding genes
C -> A	5544; contig 19		

C -> A	5551; contig 19	ADP-forming succinate--CoA ligase subunit beta, hypothetical protein	May impact on expression of succinate--CoA ligase subunit beta
Insertion of 10 bp	121096; contig 20	hypothetical protein, bacterial transcriptional activator domain-containing protein	Potential disruption of expression of hypothetical protein
G -> C	138204; contig 21	discoidin domain-containing protein, DUF1996 domain-containing protein	Likely no effect but possible effect on expression of discoidin domain-containing protein
G -> C	148598; contig 21	S1 family peptidase, elongation factor G	Likely no effect, possible disruption of terminator between genes
C -> G	223; contig 31	none annotated upstream, small ribosomal subunit Rsm22 family protein	Possible effect on expression of small ribosomal subunit Rsm22 family protein
Insertion of 244 bp	7952; contig 37	gamma-glutamyl-gamma-aminobutyrate hydrolase family protein, LysR family transcriptional regulator	Possible effect on terminator between genes
Region of difference	20634 - 20660; contig 39	hypothetical protein, NADH-quinone oxidoreductase subunit NuoN	Possible effect on terminator between genes
G -> T	30033; contig 42	lysylphosphatidylglycerol synthase domain-containing protein, ABC transporter ATP-binding protein	Possible impact on promoter between genes
T -> C	10870; contig 43	3-keto-5-aminohexanoate cleavage protein, redox-sensitive transcriptional activator SoxR	Possible effect on terminator between genes
Region of difference	34722 - 34733; contig 57	iron transporter or lucA/lucC family siderophore biosynthesis protein, none downstream	May effect promoter upstream of iron transporter or lucA/lucC family siderophore biosynthesis protein
G -> T	6643; contig 63	lipoprotein, hypothetical protein	Potential disruption of expression of lipoprotein
C -> G	6665; contig 63		

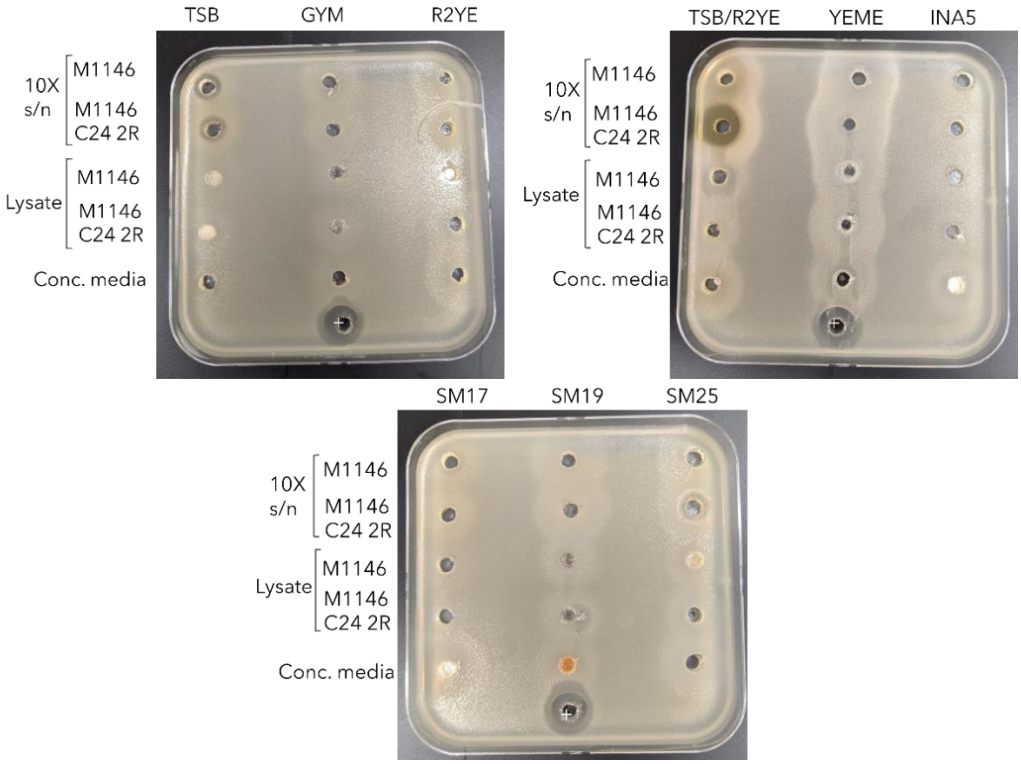


Figure S2.4: Bioassay of *S. coelicolor* M1146 C24 2R using panel of candidate media used for first trials (Figure 2.11). Cultures were grown in liquid medium for nine days before collection and concentration 10X via freeze-drying, Assay was carried out against *B. subtilis* 168, + = 5 μg gentamicin sulfate. First plate (top left) is also used as part of Figure 2.23, but has been included here for comparison to the rest of the media conditions used within this experiment.

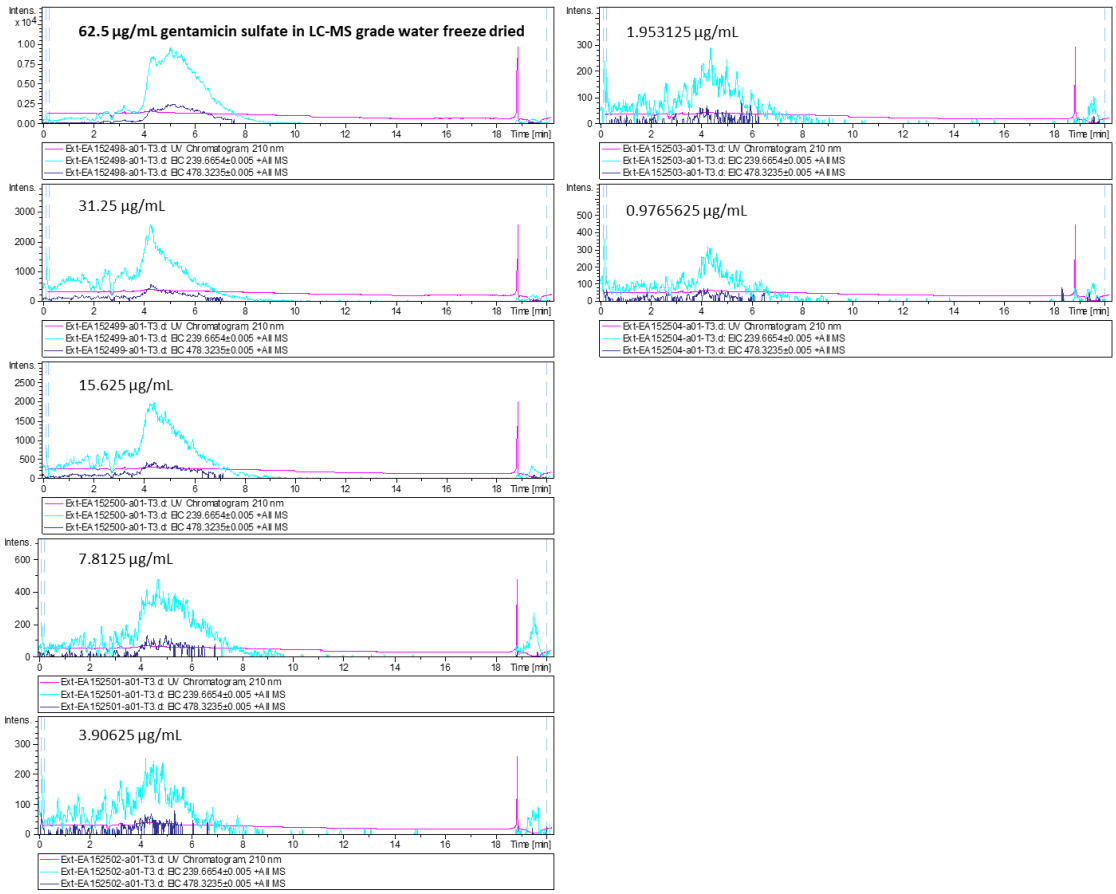


Figure S2.5: Extracted ion chromatograms of gentamicin sulfate standards in water run with a Waters Atlantis T3 column (4.6 x 100 mm, 5 mm particle size) for the separation. Mobile phase: solvent A water:MeCN 90:10, solvent B water:MeCN 10:90, both with 13mM ammonium formate and 0.01% TFA. Dark blue trace = gentamicin C1 $M+H^+$, 478.3235; Light blue trace = gentamicin C1 $M+2H^+$, 239.6654; Purple trace = UV_{210nm}.

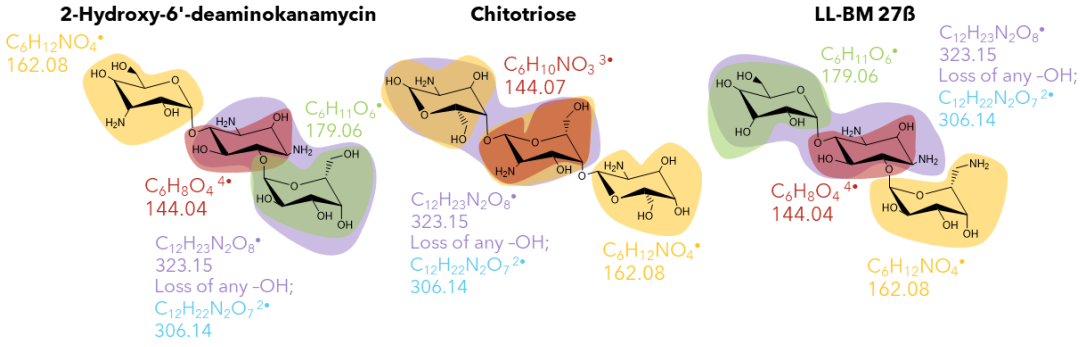


Figure S2.6: Fragmentation analysis of putative species of formula $C_{18}H_{35}N_3O_{13}$.

3. Author Contributions for Chapter 3

Katherine V. Baker was responsible for the construction and testing of pSGCH, the generation and validation of all CRISPR knockout plasmids and strains, the generation and validation of CRISPR knock-in strains, the construction, validation and testing of pSG-WS/pSG-WE/pSG2, characterisation of antibiotic production from *Streptomyces* hosts carrying modified Cluster 24 or minimal aminoglycoside pathways, LC-MS sample preparation and data analysis

Victoria Jackson was responsible for insertion of Cluster 24 into *Streptomyces* hosts and for construction of pSG (Chapter 3.3.2., **Figure S3.6**). Bernhard Kepplinger was responsible for the construction of pSG and pRes (Chapter 3.3.2., **Figure S3.5**, **Figure S3.6**).

Amelia Kelly completed her undergraduate final year project for the degree of BSc (Hons) Microbiology under the supervision of Katherine V. Baker and was responsible for the cloning of pCM4.4C24bi-dc and pCM4.4C24bi-di for the Cas9-mediated insertion of promoter cassettes to Cluster 24 (Chapter 3.3.3. and **Figure S3.8, Figure S3.9**)

Jesús Martín Serrano carried out all running of LC-MS samples (3.3.3., 3.5.2). Jesús Martín Serrano and Francesco Del Carratore carried out LC-MS data processing (Chapter 3.3.3., 3.5.2, **Figure 3.13**).

3. Towards improved selective gentamicin congener production in *Streptomyces* spp.

Katherine V. Baker,[1] Victoria Jackson,[1] Jesus Martin Serrano,[2] Francesco Del Carratore,[1] Bernhard Kepplinger, [3] Amelia Kelly, [1] Olga Genilloud,[2] Nicholas Allenby,[3] Rainer Breitling,[1] Eriko Takano[1]

[1] Manchester Institute of Biotechnology, Manchester Synthetic Biology Research Centre SYNBIOCHEM, Department of Chemistry, The University of Manchester, Manchester, M1 7DN, United Kingdom

[2] Fundación Medina, Av. Conocimiento, Parque Tecnológico de Ciencias de la Salud, 34 - 18016 Granada, Spain

[3] Demuris Limited, The Biosphere, Draymans Way, Newcastle Helix, Newcastle upon Tyne NE4 5BX, United Kingdom

3.1. Abstract

The antimicrobial resistance crisis represents one of the major healthcare threats for current generations; therefore, the combination of production of novel antimicrobials along with rejuvenating our existing antimicrobials is key. For example, one class of antimicrobials, the aminoglycosides, is underused clinically due to the nephrotoxic and ototoxic side effects observed upon administration to patients. In this work, we aimed to generate a set of *Streptomyces* strains capable of producing select gentamicin congeners which could be used for generation of new scaffolds or have previously shown lower cytotoxicity. We used both total cluster refactoring strategies and CRISPR-Cas9 modification of an aminoglycoside cluster from novel *Micromonospora* sp. DEM32671, aiming to produce gentamicins A2 and C1a in *Streptomyces* spp. While we were unable to confirm production of these specific congeners, this work continues describing a pipeline from early heterologous expression towards tailored aminoglycoside production. In addition, we used CRISPR-Cas9 to insert a bidirectional promoter cassette in the main biosynthetic region of Cluster 24 for improvement of antibiotic yield. Together, these two approaches led to a yield improvement allowing further detection of potentially novel aminoglycoside of mass 501.224, providing a platform for

further modification of gentamicin in *S. coelicolor*. If these two approaches were to be combined, we expect production of select congeners to be easier to achieve, allowing for production of a more appealing gentamicin for use clinically.

3.2. Introduction

Antibiotic resistance is one of the largest current global threats to human health, with a 2014 report commissioned by the UK government predicting resistant microbial infections to cause 10 million deaths annually by the year 2050.[1] Amongst the newly-emerging alternative therapies to classical antibiotics (such as phage therapies, immunotherapeutics, fecal microbiota transplants),[2] expanding our library of available compounds is essential to combating antibiotic resistance. The aminoglycosides represent a strong target for a rejuvenation strategy, due to being underexploited currently in high-income countries. Across the class, nephrotoxicity and ototoxicity as side effects are the main factors discouraging clinical use.[3–5] If these toxic side effects could be minimised, it would make the aminoglycosides much more attractive as first-line therapeutics.

Gentamicin is a 4,6-disubstituted 2-deoxystreptamine (2-DOS) aminoglycoside antibiotic. Like the other aminoglycosides, gentamicin can cause severe nephrotoxicity and cochleotoxicity when administered.[6] In both high and low/middle income countries, it is currently reserved for treatment in rare cases, including in treatment of multi-drug resistant infections (particularly tuberculosis)[7] as well as a first-line treatment of newborn babies under high risk of sepsis.[8–11] The primary components of clinically-administered gentamicin are the gentamicin C congeners, which are the end-products of the biosynthetic pathway.[12] As the gentamicin C congeners are challenging to purify out from each other, some pathway intermediates such as gentamicin As, gentamicin B, and gentamicin Xs are also included.[13] Importantly, as gentamicin is derived from microbial cultures, the ratio of each congener varies

hugely based on production location and even batch.[14] The establishment of which gentamicin congeners provide the best antimicrobial activity in comparison to the cytotoxicity is still on-going: Ishikawa et al. (2019) suggested that gentamicin C1a was the least ototoxic congener with *in vitro* studies on otic cell lines and use of an *in vivo* guinea pig model,[15] while O'Sullivan et al. (2020) suggested gentamicin C2b was the least ototoxic in an *in vitro* rat model.[16] O'Sullivan et al. also suggested that gentamicin C2 studies is the most ototoxic congener, a sentiment shared by Kobayashi et al. (2008) in their earlier work to characterise gentamicin C congener otoxicity.[17]

Synthetic biology strategies lend themselves well to the revitalisation of antimicrobials from secondary metabolite biosynthetic gene clusters (BGCs). However, heterologous expression of these gene clusters in more amenable hosts rarely works without further fine-tuning, exemplified in the review by Nah et al. (2017).[18] Bottom-up strategies, or rebuilding strategies, aim to regenerate the minimal biosynthetic region from a gene cluster to avoid interference from native host regulatory systems or cellular burden from maintenance of superfluous genes. This strategy was exemplified by Song et al.'s work (2019), where the spinosad gene cluster was completely reconstructed into 7 synthetic operons. Upon heterologous expression in S*treptomyces albus* J1074, an increase of 328-fold of macrolide spinosad yield was achieved compared to the heterologously-expressed native spinosad cluster.[19] An additional recent example of this can be found in Yan et al.'s work (2018), where an assembly strategy was developed to reassemble synthetic, hybrid myxochromide BGCs. 33 different constructs were generated with variations on the myxochromide biosynthesis pathway, and production of five different myxochromide types was able to be established.[20]

Meanwhile, top-down approaches use smaller, tailored modifications towards changing an intact BGC. A common example of this is the replacement of native promoters for promoter sequences between suited for heterologous host expression. Zhang et al. (2017) have shown a generalised approach for

promoter insertion using CRISPR-Cas9 modification for pigment biosynthetic gene clusters in *Streptomyces*.[21] In the case of the work carried out by Bauman et al. (2019) on streptophenazine biosynthesis, four promoters were inserted into the transcriptionally-silent *spz* BGC by λ-Red recombineering. One bidirectional cassette was sufficient for transcription activation of all cluster genes, and over 100 compounds were yielded from the improved gene cluster upon screening.[22] Recently, Ji et al. (2022) have described a sequential promoter replacement in the daptomycin cluster, to increase total lipopeptide titers up to 1780%.[23] Top-down approaches are not limited towards yield improvement: as heterologously-expressed gene clusters are often transcriptionally silent, small changes to regulatory genes can be made to awaken cluster expression. For example, oviedomycin production was activated in *Streptomyces ansochromogenes* by Xu et al. (2017) by disrupting global regulator AdpA,[24] while in the work of Olano et al. (2014), insertion of strong constitutive promoters alongside overexpression of regulatory genes yielded products from five candidate BGCs. In addition, one of the target regulators was identified in another streptomycete and was inserted into the strain's genome for action on the candidicin and antimycin gene clusters.[25]

In this work, there were two primary aims: further yield improvement from Cluster 24 in a *Streptomyces* heterologous host, and production of selective gentamicin congeners. Gentamicin A2 and gentamicin C1a were the congeners of primary interest to us. Gentamicin A2 represents the first pseudotrisaccharide of the gentamicin pathway,[26] and was of interest as a scaffold for further modification. As 4,6-disubstituted 2-deoxystreptamine aminoglycosides share early biosynthesis pathways, gentamicin A2 may be the easiest target for modification with enzymes from other aminoglycosides of this type. For example, a patent for the gentamicin-kanamycin hybrid aminoglycoside genkamicins was filed in 2015.[27] Meanwhile, gentamicin C1a was shown by Ishikawa et al. in 2019 to be less ototoxic in *in vitro* and *in vivo* studies than other gentamicin congeners,[15] and so this became our target gentamicin C congener in this work.

We also aimed towards two methods of production of these congeners. We tested CRISPR-Cas9 knockout of five genes in three modifications for the production of gentamicin A2 and gentamicin C1a in *S. coelicolor*, though the resultant antimicrobial bioactivity against *B. subtilis* was too low for further characterisation. In addition, we generated a minimal gentamicin biosynthesis pathway based on previously published *in vitro* and *in vivo* characterisation.[28,29] Lastly, we built upon our previously reported primary strategy towards improving aminoglycoside yield by *Streptomyces* heterologous hosts carrying a gene cluster from novel *Micromonospora* sp. DEM32671.[30] Previously, we showed that exposure to increasing concentrations of gentamicin improved both maintenance of this gene cluster over time and antibiotic titres, allowing for first characterisation of the cluster product. In this work, we additionally aimed to further improve yield of this strain by insertion of a bidirectional promoter cassette within a core region of the gene cluster. This should upregulate expression of the majority of cluster ORFs to further improve production. Additionally, after insertion of the promoter cassette we saw a yield improvement which allowed for identification of metabolite of m/z 502.2241, matching an ion that we previously linked to presence of Cluster 24.[25] Though these approaches were not tested in combination, we believe that this work provides a further stepping-stone for production of aminoglycosides in *Streptomyces* heterologous hosts.

3.3. Results and Discussion

3.3.1. Preliminary trials for knockout of key cluster genes, aiming for production of gentamicin A2 and gentamicin C1a

Plasmids for Cas9-based modification of Cluster 24 were generated based on pCM4.4, which was originally constructed by Ye et al. (2020; **Figure 3.1**).[31] pCM4.4ΔgenK and pCM4.4ΔgenJgenK2 **(Figure 3.1)**, knock out i) *genK* and ii) *genJ* and *genK2* respectively, blocking the conversion of gentamicin X2 to G418 and the conversion of JI-20A to gentamicin B. *genK* is also responsible for the conversion of gentamicin A to gentamicin X2 – therefore, removing it completely could impact antibiotic production. *genD1* and *genN* have been shown to preferentially carry out this reaction *in vivo* in *M. echinospora*.[28] If the knockout of *genK* halted production of gentamicin X2 (and therefore the rest of the pathway following), a further modification to insert stronger promoters upstream of *genD1* and *genN* (and therefore over-express them *in vivo*) could additionally be introduced. *genL* has also been recently identified as being involved in the biosynthesis of gentamicin C2b from gentamicin C1a but was not a candidate for deletion in this work. In the industrial gentamicin producer strain *Micromonospora echinospora* DSM 43816, *genL* is present 2.54 Mbp away from the rest of the gene cluster.[28] A homologue of this gene with 89.54% identity was identified by us previously, 2.67 Mbp from Cluster 24 in the *M.* sp. DEM32671 genome.[30] This was therefore not included when the gene cluster was originally captured for heterologous expression.

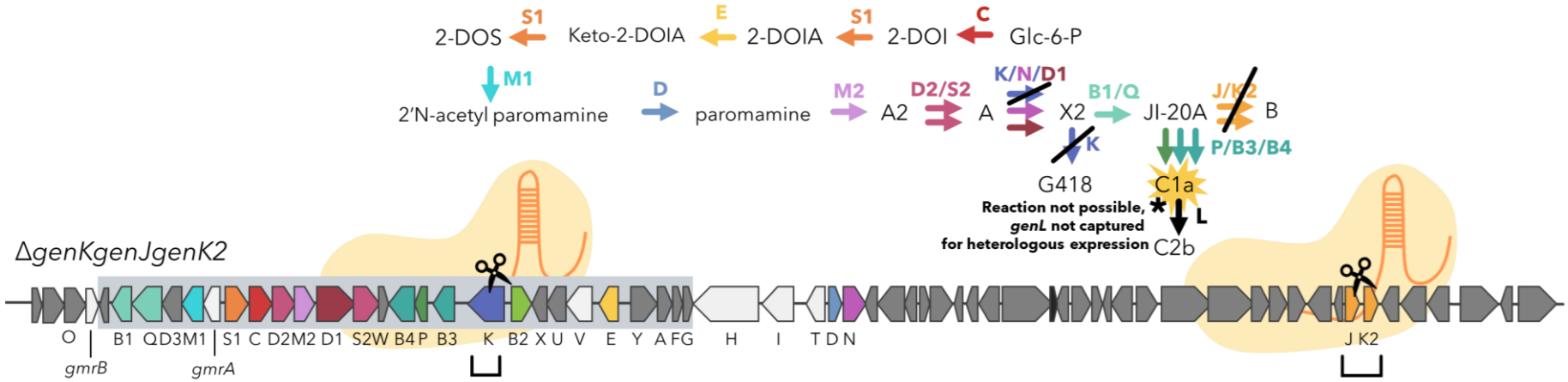


Figure 3.1: Illustrative schematic of the CRISPR-Cas9 modification of Cluster 24 for selective production of gentamicin C1a. Knockout of *genK* disrupts the pathway from gentamicin X2 to triple-methylated intermediate G418, inhibiting production of all other gentamicin C congeners.[28] Disruption of the pathway to gentamicin B from JI-20A is carried out by deletion of *genJ* and *genK2*.

pCM4.4ΔgenK and pCM4.4ΔgenJgenK2 were constructed as described in Materials and Methods **(Figure S3.2)**. The protospacers were first checked for no full sequence homology to the *S. coelicolor* genome or pESAC13-C24 via BLAST, ensuring that the protospacer + protospacer adjacent motif (PAM) sequence (all combinations of 'NGG') were unique. These were then inserted into the pCM4.4 vector to be incorporated into the sgRNA via NEB HiFi assembly **(Table 3.1)**. Homology arms of 1000 bp in size were taken from the sequence of Cluster 24 which flanked the desired deletion (pCM4.4ΔgenK – 25,499 → 26,499, 28,218 → 29,218; pCM4.4ΔgenJgenK2 – 71,413 → 72,413, 74,301 → 75301). Sanger sequencing was used to validate all plasmids had been assembled correctly.

Table 3.1: Protospacers for each CRISPR knockout plasmid generated in this work.

Plasmid	Position of protospacer (3′ end) in Cluster 24	Protospacer sequence (5′ – 3′)
pCM4.4ΔgenD1genS2	19,555	AGGTGTCGATGTCCTTCTGG
pCM4.4ΔgenK	27,883	AACGCGAAGGTGATCCTTGG
pCM4.4ΔgenJgenK2	73,329	CCGTGCTCATTGGATCGCGA

First, knockout of *genK* was carried out on *S. coelicolor* M1152 C24 by conjugation of pCM4.4ΔgenK into the strain and screening for gene deletion **(Figure S3.3)**. Strains originating from two different independent colonies were generated, labelled as *S. coelicolor* M1152 C24 Δ*genK*(2) and *S. coelicolor* M1152 C24 Δ*genK*(3). These strains were cultured in both GYM and R2YE medium over a period of 7 days after inoculation from seed culture, and timepoints taken for measurement of antibiotic production at day 2, day 3, day 5, and day 7 **(Figure 3.2)**. First production of antibiotic from *S. coelicolor* M1152 C24 is visible at day 5 in concentrated culture supernatant and at day 7 in unconcentrated lysate, in R2YE medium only. Bioactivity is also visible in

unconcentrated lysate samples from the *genK* knockout strain *S. coelicolor* M1152 C24 Δ*genK*(2) at day 7 in R2YE medium only. Though this initial bioactivity was low, it remained promising; therefore, we proceeded with use of this strain to knock out *genJ* and *genK2*, responsible for gentamicin B biosynthesis.[32]

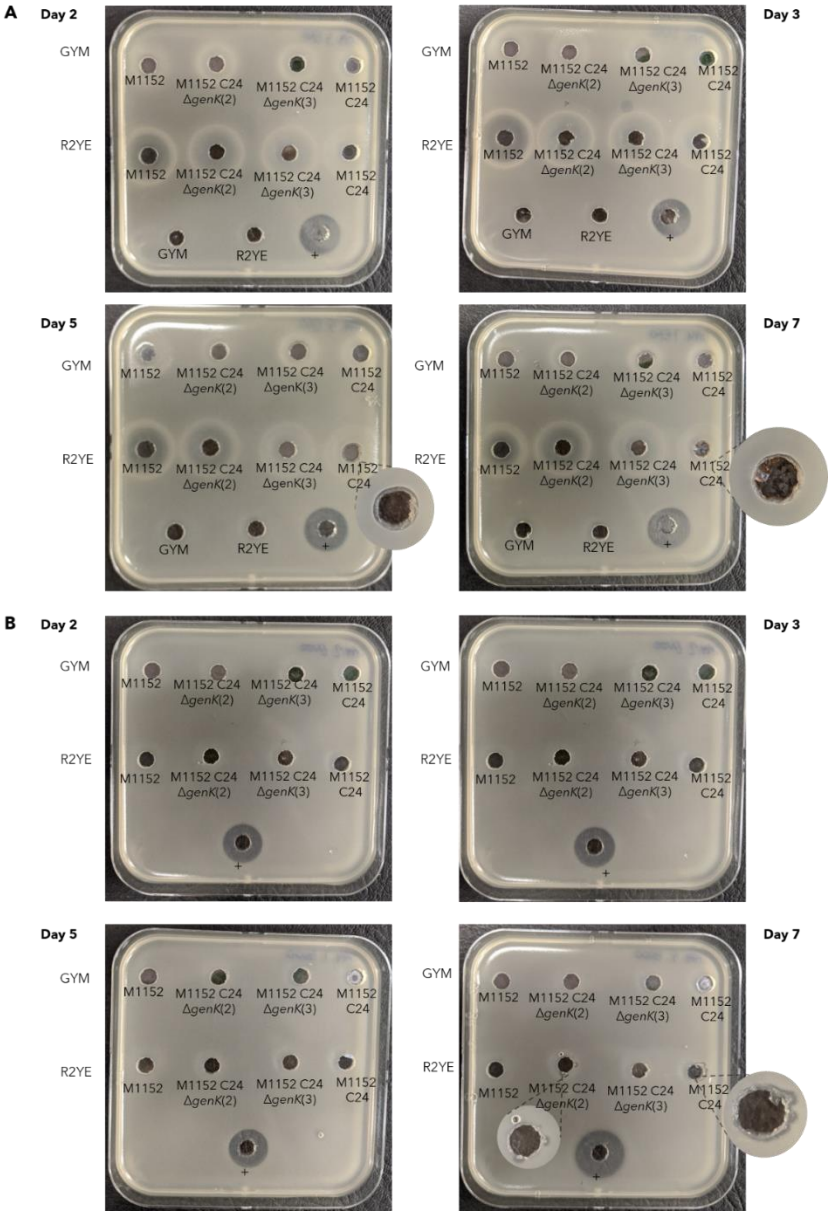


Figure 3.2: Bioassay of samples from CRISPR-modified *S. coelicolor* M1152 C24 strains against *B. subtilis*. A – Bioassay of 10X concentrated culture supernatant. 20 µL of extract was added to each well of an LBA + *B. subtilis* plate. **B** – Bioassay of direct lysate collected through mechanical lysis. 40 µL of direct extract was added to each well of an LBA + *B. subtilis* plate. +ve – 5 µg gentamicin sulfate.

The triple knockout was constructed by further conjugation of pCM4.4ΔgenJgenK2 to *S. coelicolor* M1152 C24 *ΔgenK* **(Figure S3.2)**. After strains were cultured in R2YE medium, extracts were taken for analysis via antimicrobial bioassay **(Figure 3.3).**

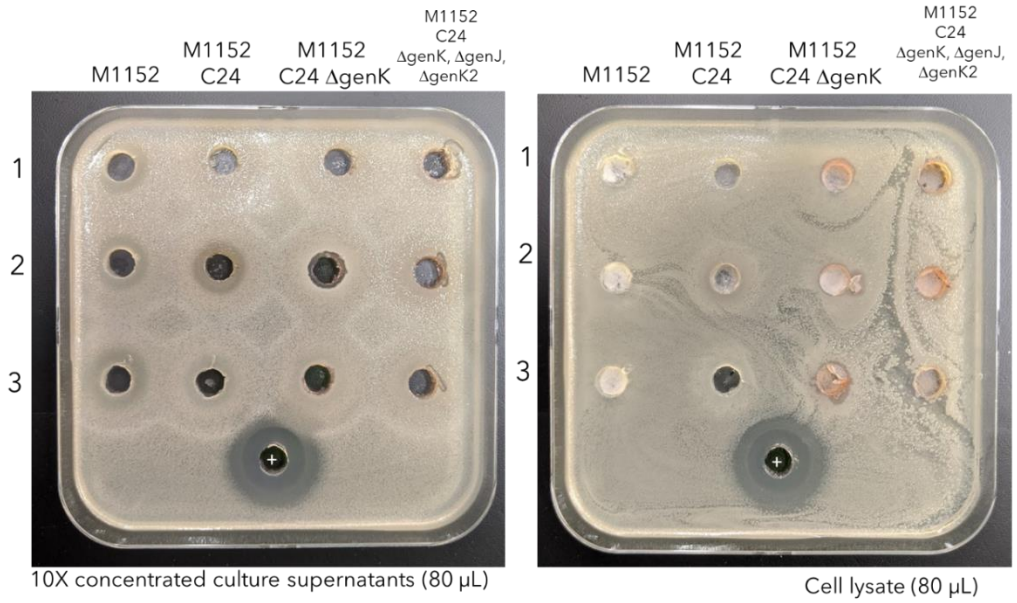


Figure 3.3: Antimicrobial bioassays of *S. coelicolor* M1152 carrying either C24, C24 *ΔgenK*, or C24 *ΔgenK ΔgenJ ΔgenK2*. Cultures in R2YE media were grown for 7 days. Antibiotic activity can be visualised by the inhibition of *B. subtilis* growth. + = gentamicin sulfate, 10 µg.

From this antimicrobial bioassay, it appeared that too little antibiotic was produced for further characterisation. It would be expected that the *genK* knockout strain shows a lower bioactivity, due to the proposed loss of the main antibiotic congeners in the mixture,[28] but no antibiotic activity was seen in this experiment from *S. coelicolor* M1152 C24. Only one of the three *S. coelicolor* M1152 C24 Δ*genK* cultures appeared to show production of a product with antimicrobial effect, and only in the samples concentrated from culture supernatants. The cell lysate fraction did not show convincing antimicrobial 'halo', but while this was unable to be concentrated via freeze-drying due to sample viscosity, it is likely that these metabolites are present intracellularly at concentrations too low to visualise, as activity was seen in the previous experiment. Low antibiotic production was an issue with early experiments

prior to our described approach for yield improvement,[30] which made further characterisation of the product challenging.

At the time of this work, *S. coelicolor* M1152 showed promise as an antibiotic production host, as it was well-suited to the CRISPR-Cas9 systems we had available for cluster modification. However, since this initial work was carried out, we have described the iterative improvement of aminoglycoside yield through creation of a cohort of *S. coelicolor* M1146 C24 mutant strains.[30] These showed an increased antibiotic production phenotype and were able to maintain the BGC due to an increased gentamicin resistance. It is clear that this panel of resistant strains represent a better selection of hosts than the *S. coelicolor* M1152 strain utilised here. The first step for improvement would be to carry out these gene knockouts on the panel of resistant strains, to more accurately see the effect on antibiotic production. With this, and an improved analysis method, we would be able to determine whether these gene knockouts were sufficient for production of the gentamicin C1a congener we aimed for.

Alternatively, since this work was begun, the production of gentamicin C1a has been established by engineering native gentamicin producer *M. purpurea* GK1101 to delete *genK* and *genL* from the gentamicin BGC. Homologous recombination was used to remove these sequences from the genome of *M. purpurea* GK1101, and the strain produced gentamicin C1a selectively and maintained bioactivity.[33] While our work here follows a similar process, the authors of this study did not account for the production of gentamicin B: while this is regarded as a minor component of the mixture,[13] it is clear from O'Sullivan et al.'s work that removal of all contaminating congeners is important for further cytotoxicity testing.[16] The use of a less commonly-used heterologous host organism that is more closely related to the original cluster host is an interesting prospect, and in this case insertion of Cluster 24 to a *Micromonospora* sp. (as both Wei et al. and Chang et al. have

shown two *Micromonospora* sp. to be genetically tractable)[33,34] could assist with the low initial titres seen.

While we used the information provided by Ishikawa et al. (2019) to guide our attempts to produce gentamicin C1a, a recent study by O'Sullivan et al. (2020) has suggested that in fact, gentamicin C2b showed a lower ototoxicity than the other gentamicin C congeners tested. The methods used in these two publications differ slightly: O'Sullivan et al. used cochlear cultures for measurement of outer hair cell survival, and electrophysiological measurement of outer hair cell function (aminoglycoside ototoxicity is caused by the aminoglycosides being taken up through the mechanotransducer (MET) channels of the outer hair cells, as the flow of current across the membrane is impacted).[35,36] The work by Ishikawa et al., however, investigated toxicity by looking into the specific loss of metabolically active cells of two otic cell lines upon exposure to aminoglycosides. Their *in vivo* work focused on compound action potential measurement in live guinea pigs to determine hearing loss, and this was followed by post-mortem measurement of synaptic ribbon decrease in inner hair cells to determine aminoglycoside-linked loss of neurotransmission ability. These differences in methodologies likely account for the differing reports from both works. In O'Sullivan et al.'s work, gentamicin C1a performed the second-best in terms of outer hair cell survival rates, and by replacing gentamicin C1 and C2 in a mix of gentamicin C congeners (final proportions being 65% gentamicin C2b, 28% gentamicin C1a, 7% gentamicin C2a) they were able to increase the outer hair cell survival rate without compromising antibiotic activity. Additionally, this gentamicin C2b/gentamicin C1a/gentamicin C2a mix showed no difference in ototoxicity to pure gentamicin C2b but did show a higher MIC value. For future improvement of our work, establishing a route to gentamicin C2b in a heterologous host would therefore likely be of further value. This could be easily done by expressing *genL* in addition to the modified BGC with *genK*, *genJ* and *genK2* knocked out.[28]

Alongside production of gentamicin C1a, we further aimed to produce a knockout strain that would be capable of producing only gentamicin A2. The plasmid constructed towards this, pCM4.4ΔgenD1genS2, knocks out the *genD1* and *genS2* gene homologues from Cluster 24 **(Figure 3.4)**. Though only *genS2* is required to be knocked out for inhibition of the pathway past the biosynthesis of gentamicin A2, *genD1* was also knocked out at the same time as it is directly located next to *genS2* and has been shown to be unnecessary for biosynthesis of gentamicin A2.[26] Construction was carried out similarly to generation of pCM4.4ΔgenK, albeit with first protospacer insertion via Golden gate assembly before insertion of homology arms **(Figure S3.4)**. 1000 bp homology arms were taken from sequence of Cluster 24 (17,562 → 18,562, 21,488 → 22,488 bp). Sanger sequencing (Eurofins Genomics) was used to validate correct plasmid assembly. While we constructed pCM4.4ΔgenD1genS2 and were able to test this in *S. coelicolor* M1146 C24 **(Figure S3.4)**, we found upon screening that the strain was not modified **(Figure 3.4B)**. Our initial screening efforts also were hampered by the strain losing the BGC during the passaging on solid medium required for curing pCM4.4-based plasmids. Tackling this cluster loss from *S. coelicolor* M1146 was a major effort that we have previously discussed.[30] As with the knockouts of *genK*, *genJ*, and *genK2*, testing this plasmid on the panel of resistant strains that we have now established would be a quick further experiment which should allow for selective gentamicin A2 production. It has been suggested that a byproduct, gentamicin A2e, may be produced by action of GenK on gentamicin A2. This gentamicin A2e differs from gentamicin A2 by addition of a methyl group on the 6′ carbon of the purpurosamine ring.[28] While pCM4.4ΔgenK was originally built to be used in combination with pCM4.4ΔgenJgenK2 for the selective production of gentamicin C1a, though it could be used in combination with pCM4.4ΔgenD1genS2 for removal of the potentially problematic by-product gentamicin A2e.[28] As the same protocol was followed with success to generate the *genK*/*genJ*/*genK2* triple knockout strains, it could additionally be possible that the chosen protospacer was not

favourable for action of Cas9. Changing this would represent a small modification to the pCM4.4ΔgenD1genS2 sequence which could be carried out with a single round of cloning.

The success of the top-down approach to production of selective congeners was not able to be fully realised, largely due to low yield of antibiotic in the heterologous host carrying the BGC at the time of this work. However, if combined with our panel of overproducing strains, we anticipate that production of the selective congeners of interest could be easily achieved.

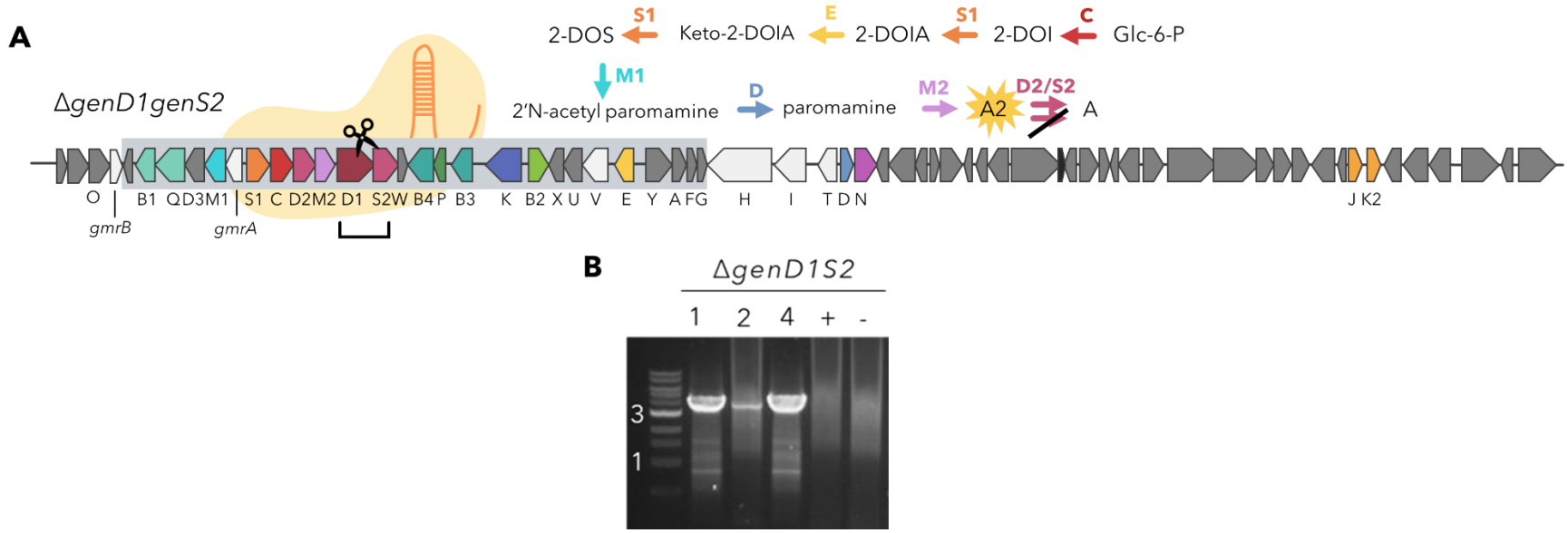


Figure 3.4: A – Illustrative schematic of the CRISPR-Cas9 modification of Cluster 24 for selective production of gentamicin A2. Single knockout of *genS2* is sufficient to inhibit the pathway from gentamicin A2 to gentamicin A. **B –** PCR screen of knockout of *genD1*, *genS2* in *S. coelicolor* M1152. Screen was carried out with genD1S2screenfw and genD1S2screenrev, product of 3909 bp present for unmodified BGC and 873 bp present for modified BGC. + = *S. coelicolor* M1152 C24, - = *S. coelicolor* M1152.

3.3.2. Refactoring for selective gentamicin A2 and gentamicin C1a production

As low yield was proving to be a bottleneck in validating the congeners produced, we next turned towards a total refactoring approach. Our first minimised pathway construct, pSG, contained homologues of the first six genes of the gentamicin pathway (codon-optimised for best expression in *Streptomyces*; homologues of genes from *M. echinospora* DSM 43816 characterised *in vitro* by Park et al.)[26] under control of the tetracycline-inducible expression system (TetR, adapted for expression in *Streptomyces;* **Figure 3.5**).[37] After previous experiments, we anticipated that antibiotic production may be toxic to the heterologous host, and so having the biosynthetic machinery under control of a transcription factor system should alleviate stress especially when first generating the strain. To assist with potential toxicity, a secondary plasmid, pRes, was constructed. This contains *genV* and *gmrA* under control of a medium-strength promoter, *ermEP1*. *genV* has been predicted to encode a transmembrane efflux protein based on amino acid sequence, and *gmrA* is a 16S rRNA methyltransferase which functions to methylate the 16S rRNA at the N^7 of G_{1405}. By including this, the host should be more resistant to gentamicin congeners, while also ensuring that the end-product is not altered in any way that could compromise bioactivity. The homologues of these found in Cluster 24 have a 95.54% (*gmrA*; E = 0.0) and 96.99% (*genV*; E = 0.0) to the corresponding *M. echinospora* genes when compared through BLASTp. In addition, in previous work (Callum Howden, unpublished) *S. coelicolor* carrying Cluster 24 and pRes (and therefore an extra copy of *genV*) showed a larger antibiotic halo than *S. coelicolor* carrying Cluster 24 alone, suggesting a potential increased export of gentamicin from the cell.[38]

We carried out first characterisation of pSG and pRes when also establishing antibiotic production conditions by *S. lividans* TK23, one of our better producer strains earlier in this work.[30] Three solid media (SFM, GYM, and R2YE) and

one liquid medium (GYM) were chosen to attempt to achieve antibiotic activity from *S. lividans* TK23, and all strains were tested with and without 100 ng/mL anhydrotetracycline. This tetracycline analogue was chosen as it has no antibiotic activity and should not have negative impact on *Streptomyces* growth or interfere with antimicrobial bioassay results. Additionally, it has been validated as a functional inducer in *E. coli* using the TetR transcription factor system for gene expression.[39] In this experiment, production of antibiotic from *S. lividans* TK23 C24 was able to be seen in liquid cultures of GYM medium, only without induction with 100 ng/mL anhydrotetracycline. No production was seen by any strains grown on solid medium. Production of antibiotic from *S. lividans* TK23 C24 (and *S. lividans* TK23 C24 pRes) should not be impacted by anhydrotetracycline induction, as none of the genes have been placed under the control of *tcp830*. These early results give indication towards the challenges and the variability of heterologous production, later seen in section 3.3.4. *S. lividans* TK23 pSG pRes showed no halo indicative of antibiotic production in either induced or uninduced conditions. As expected, no antibiotic production was seen from the negative control strain (*S. lividans* TK23).

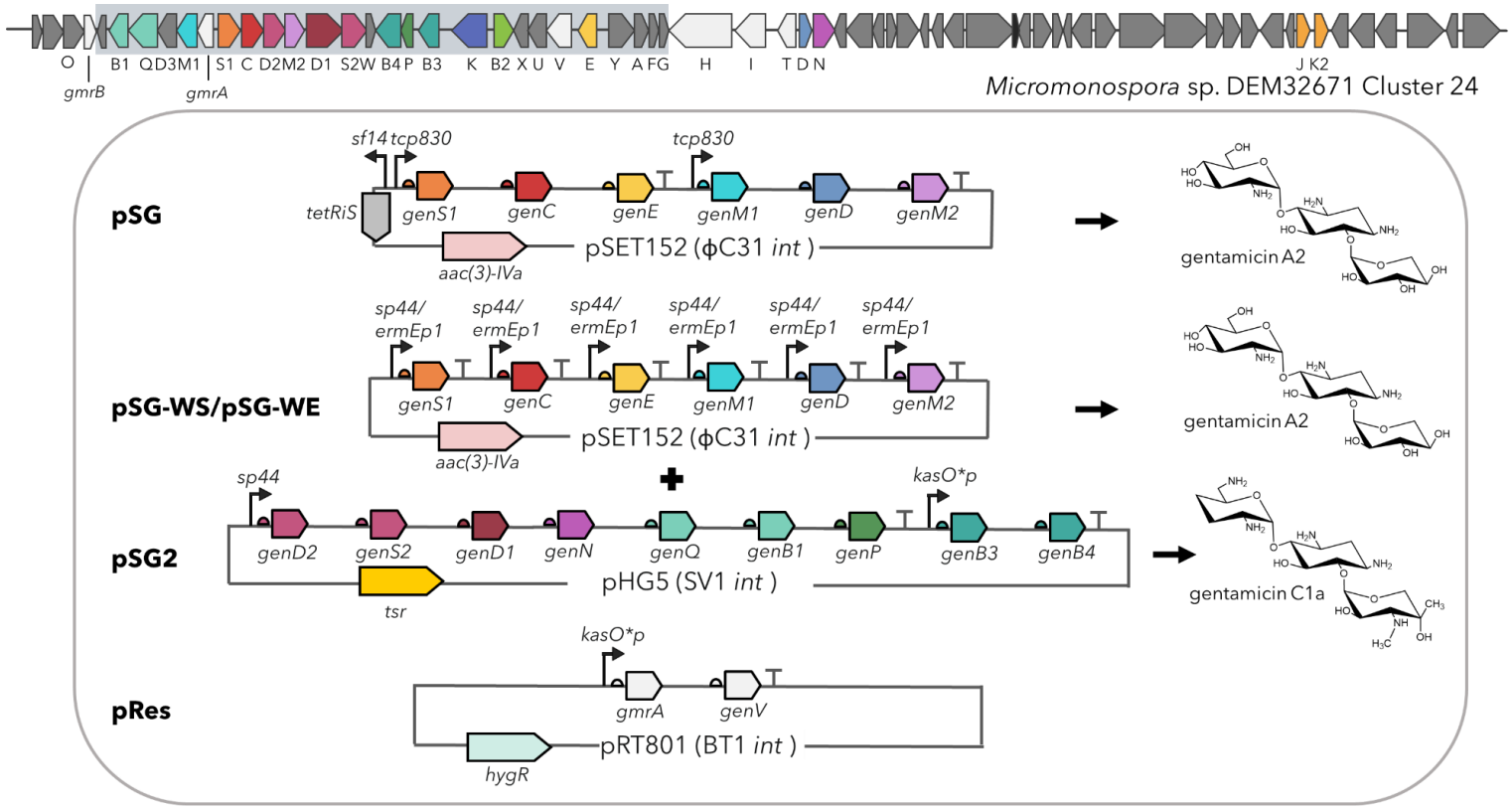


Figure 3.5: Illustrative figure of the minimal pathway constructs generated in this work. pSG consists of the genes for gentamicin A2 biosynthesis under control of tetracycline induction. pSG-WS/pSG-WE variants contain the genes for gentamicin A2 biosynthesis under constitutive control of either strong *sp44* or weaker *ermEp1* promoters. pSG2 is coupled with either pSG, pSG-WS or pSG-WE for production of gentamicin C1a from gentamicin A2. The genes expressed on pRes have no biosynthetic function and instead act as i) a resistance mechanism through expression of *gmrA* (16S rRNA methyltransferase) ii) to export gentamicin from the *Streptomyces* cells through action of *genV*.

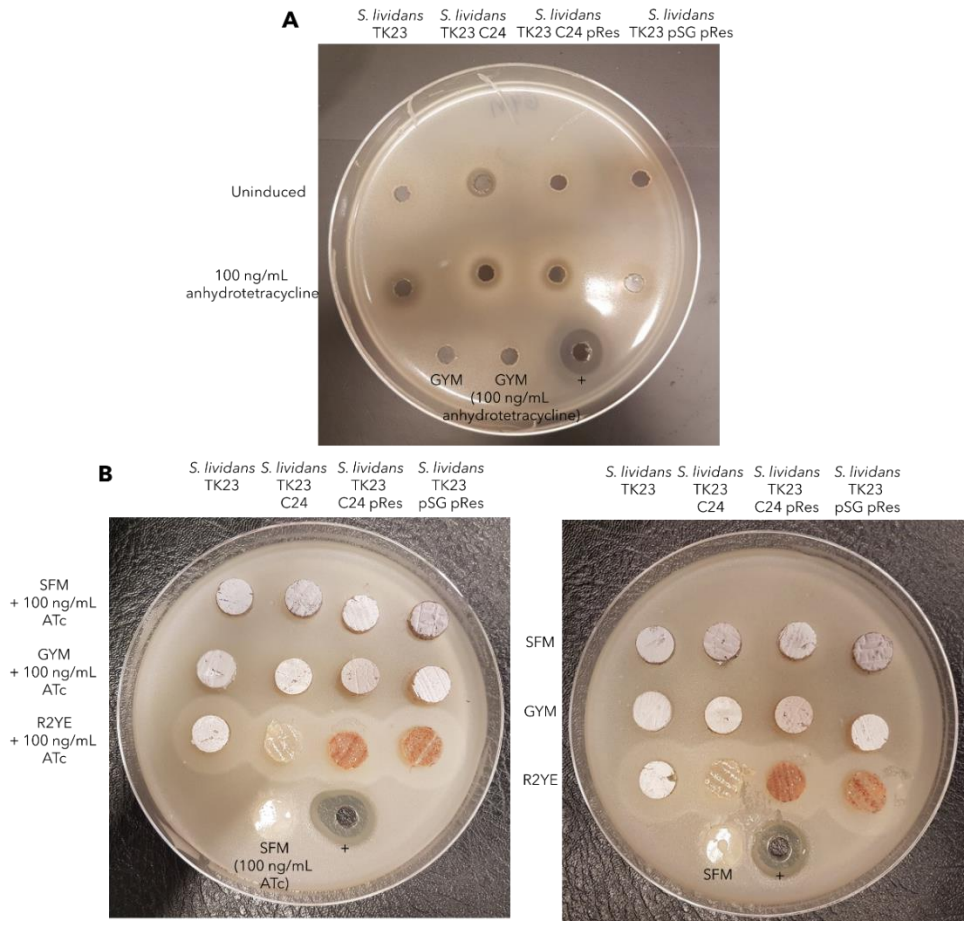


Figure 3.6: First antimicrobial bioassay tests of *S. lividans* TK23 carrying C24, C24 + pRes, or pSG + pRes on solid and liquid media. A – liquid media test of *S. lividans* TK23 carrying C24, C24 + pRes, or pSG + pRes. Strains were grown in GYM for 7 days after first seed culture, with or without 100 ng/mL anhydrotetracycline, and culture supernatants were concentrated 10X before testing 40 μL in each well. GYM with 100 ng/mL anhydrotetracycline was included to check anhydrotetracycline was not causing any inhibition of growth; + = 10 μg of gentamicin sulfate **B** – Scaling up of plug assays. *S. lividans* TK23, *S. lividans* TK23 C24, *S. lividans* TK23 C24 pRes, *S. lividans* TK23 pSG pRes were grown on SFM, R2YE, and GYM media supplemented with 100ng/mL anhydrotetracycline for seven days until plugs were taken and assayed against *B. subtilis*. SFM with 100 ng/mL anhydrotetracycline was included to check anhydrotetracycline was not causing any inhibition of growth; + = 10 μg of gentamicin sulfate.

To establish whether the translation of the cluster genes was occurring in synthetic construct pSG, the terminal gene in each synthetic operon (*genE, genM2*) was tagged at the C-terminus with a polyhistidine tag. The new construct, pSGCH (for pSG C-terminal polyhistidine tag) was conjugated into

S. lividans TK23. A western blot probing for the presence of the His-tagged *genE* and *genM2* was carried out, and bands at 36.1 and 46.7 kDa (for *genE* and *genM2* respectively) would indicate successful detection of expression **(Figure 3.7)**. *S. lividans* TK23 pTE143 was used as a positive control for detection of protein extracted from *Streptomyces*; pTE143 consists of a polyhistidine-tagged ScbR (γ -butyrolactone receptor) under control of the *tipA* promoter.[40] A band at ~28 kDa indicated successful expression of ScbR. While the positive control was able to be visualised under the conditions used, neither *genE* or *genM2* were able to be detected via this western blot. It was therefore hypothesised that the current plasmid design was unsuitable for production of gentamicin A2, and so this should be redesigned.

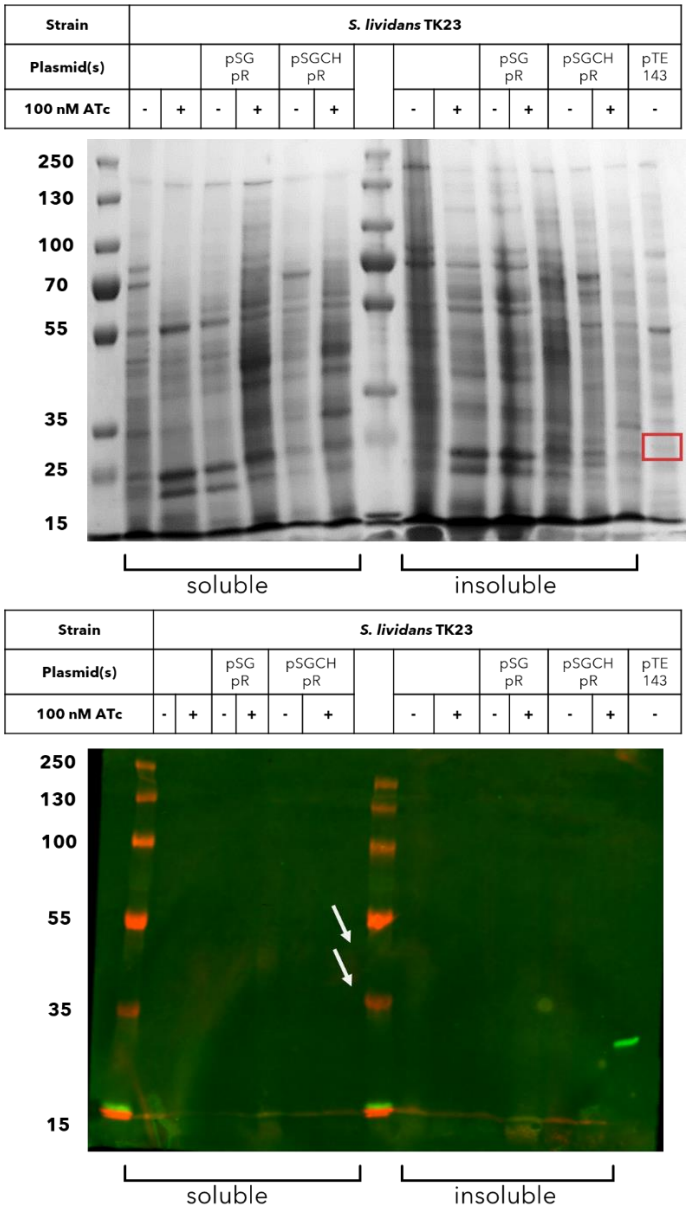


Figure 3.7: Determining protein expression from strains carrying His-tagged pSG, pSGCH. SDS-PAGE and western blot to check translation of *genE*, *genM2* in pSG. - = empty *S. lividans* TK23, pSG pR = pSG pRes, pSGCH pR = pSG with C-terminal His-tagged *genE* and *genM2* pRes. *S. lividans* TK23 pTE143 positive control expresses His-tagged *Sc*bR (~28 kDa). White arrows indicate expected location of GenE and GenM2 if these were able to be detected.

While we were unable to carry out RT-qPCR experiments to experimentally determine this, the hypothesis was that the promoter was not sufficiently strong enough for expression of the terminal genes; as such, the end-product of gentamicin A2 was not able to be produced. Additionally, in an ideal

industrial setting, the strain would not be reliant on additional inducers to achieve compound production. Therefore, we carried out a total redesign of the construct, maintaining the core six genes characterised by Park et al.,[26] but placing each one under control of their own promoters and terminators to form six independent transcriptional units. We selected two different promoters for this based on past experience within the research group to generate two different plasmid variants. *sp44* was the strongest synthetic promoter available at the time of writing of this work, whilst *ermEp1* was a weaker promoter.[41] While homologous recombination due to the repetitive promoter sequences in each construct was a potential issue with the approach, in *Streptomyces* a sequence of > ~500 bp is sufficient for successful recombination.[42] As our promoter sequences were shorter than this, we anticipated any issues of this sort to be minor. With hindsight, a tailored set of six different promoters for each construct with strength determined by RT-qPCR would undoubtedly have provided a better option than our approach here. It remains unclear whether transcription or translation caused a bottleneck in this construct design. While tailoring of ribosome binding sites has been shown to be useful for improvement of gene expression,[43] here we aimed to focus on one aspect of genetic control for optimisation, and maintained native RBS sequences throughout the further constructs designed.

We tested multiple different assembly methods for generation of pSG-WS (pSG with all genes under control of *sp44*) and pSG-WE (pSG with all genes under control of *ermEp1*), including standard Gibson assembly, ligase cycling reaction, and Golden gate assembly. We found substantial difficulties with all of these due to the high G+C% in the DNA sequences, the number of parts required for the assembly, and in the case of ligase cycling reaction and Golden gate assemblies the large difference in part size between synthetic promoter fragments (100 – 200 bp) and the coding sequences (~1 – 2 kb). Eventually, modification of the standard Gibson assembly using flanking UNS (unique nucleotide sequences) described by Torella et al. (2014) allowed for successful assembly.[44] In this case, it was hypothesised that the high G+C% content of the

overlap regions of each fragment may form interfering secondary structure, and inhibit effective assembly. The UNS sequences are a set of ten 40 bp random nucleotide sequences which have been screened to ensure i) G+C content = between 45 and 55%, ii) no start codons, common restriction sites, strong hairpins or predicted bacterial promoters present, iii) low score to the *E. coli* MG1655 genome.[45] Each fragment in the Gibson assembly had UNS sequences added via PCR primer overhangs (UNS_n on the forward primer and UNS_{n+1} on the reverse primer), ensuring that these were the region of overlap for each sequential fragment. While this means the assembly is no longer scarless, in this case the UNS sequences provide insulation between each transcriptional unit and so may be advantageous.

Once constructed, these plasmids were conjugated into *S. coelicolor* M1146, a more favourable host for further modification than *S. lividans* TK23. Upon analysis of culture extracts via bioassay **(Figure 3.8)**, it appeared that no production of bioactive metabolites had occurred. While there are some 'halos' surrounding the wells on the plate, these are also seen in the negative control (*S. coelicolor* M1146, not able to produce antibiotics)[46] and so likely arise from the concentrated growth medium having an effect on *B. subtilis* growth. After investigating further in the literature, we discovered that Li et al. (2018) reported that gentamicin A2 does not have bioactivity against their indicator strain *Bacillus pumilus*.[28] We therefore hypothesised that the lack of production via our bioassays was due to the produced congener being inactive against our *B. subtilis* indicator strain. Here, gene expression analysis would have allowed us to determine how well the construct was functioning, after now understanding that antibiotic activity would not be a readout we could rely on.

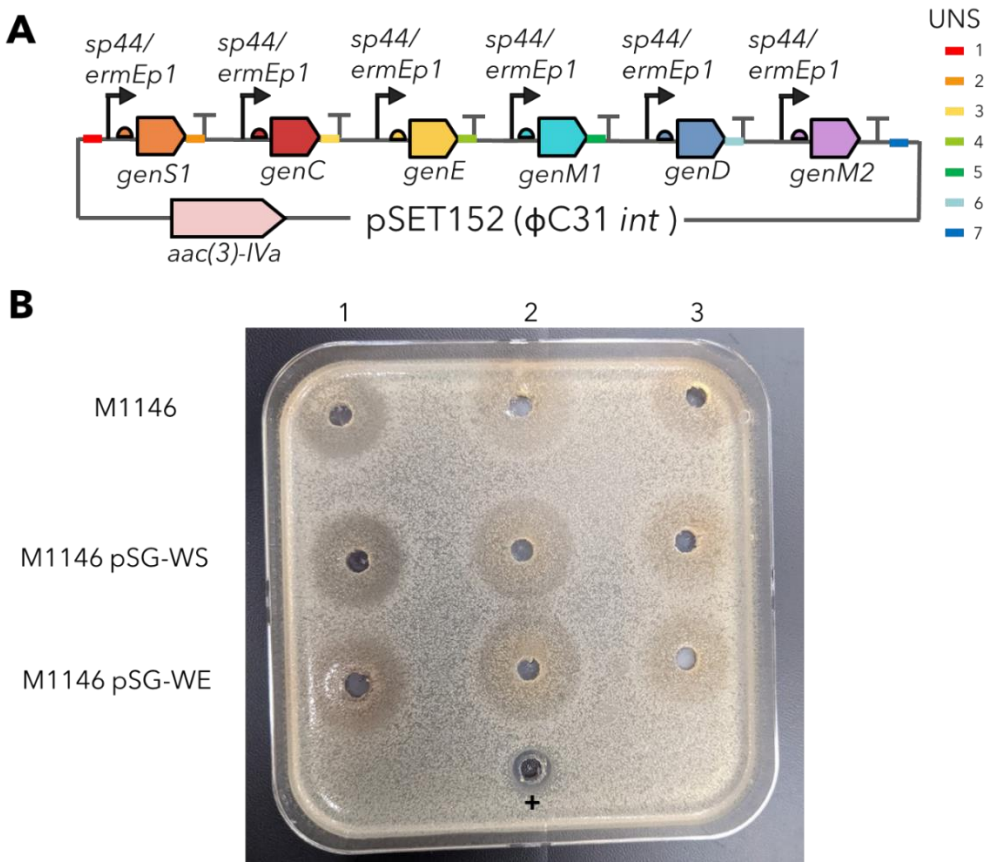


Figure 3.8: Design of pSG-WS/pSG-WE and testing of pSG-WS/pSG-WE in *S. coelicolor* M1146. A – schematic of pSG-WS/pSG-WE with unique nucleotide sequences (UNS) highlighted. **B –** Antimicrobial bioassay of culture extracts from *S. coelicolor* M1146, *S. coelicolor* M1146 pSG-WS, and *S. coelicolor* M1146 pSG-WE. + = 1 μg gentamicin sulfate, indicator strain = *B. subtilis* 168.

While we were unable to test these strains for production of gentamicin A2 via LC-MS, we next decided to establish whether introduction of the further genes required for gentamicin C1a biosynthesis in strains carrying pSG-WS or pSG-WE was sufficient to restore an antibiotic-production phenotype. We constructed pSG2, carrying *sp44-genD2-genS2-genD1-genN-genQ-genB1-genP-fd-kasO*p-genB3-genB4* as an insert, and ensured the antibiotic resistance cassette and integrase were compatible with the existing integrase and antibiotic resistance cassettes in pSG-WS/pSG-WE and pRes **(Figure S3.11)**. pSG2 was co-transformed into *S. coelicolor* M1146 pSG-WS pRes, *S. coelicolor* M1146 pSG-WE pRes, and additionally *S. coelicolor* M1146 pSG pRes. While we had not been able to detect protein expression earlier with the polyhistidine-

tagged variant of *S. coelicolor* M1146 pSG pRes, it had become clear that the lack of antibiotic activity did not necessarily mean that no gentamicin A2 had been produced; therefore, this was also co-transformed with pSG2. Upon conjugation, two distinct colony morphologies (smaller grey colonies and larger white colonies) were identified after approximately 5 days growth at 30°C. This was of interest, as proper spore pigment production from *S. coelicolor* can indicate healthy growth. We were curious whether presence of the pSG2 plasmid alongside the gentamicin A2 biosynthesis plasmids was causing deleterious effects on the strain, and so conducted a screen to test for presence of pSG2 in a subset of colonies from the conjugation plate. Interestingly, plasmid maintenance did not appear to be linked to colony morphology in this case; as such, two colonies from each morphology type were streaked out for spore collection. Only those colonies which were originally grey on these conjugation plates were able to grow upon further passaging, suggesting a possible sickness to the strain. We were therefore unable to continue with *S. coelicolor* M1146 pSG pRes pSG2 (which carried the original tetracycline-inducible gentamicin A2 pathway) as only four exconjugants were yielded and all did not grow upon further passages from the conjugation plate. The TetR system has been described as a leaky inducible system,[47,48] so it is not out of the question that the pSG construct was i) functional and ii) in combination with pSG2, yielded a toxic congener which had a deleterious effect on the host strain.

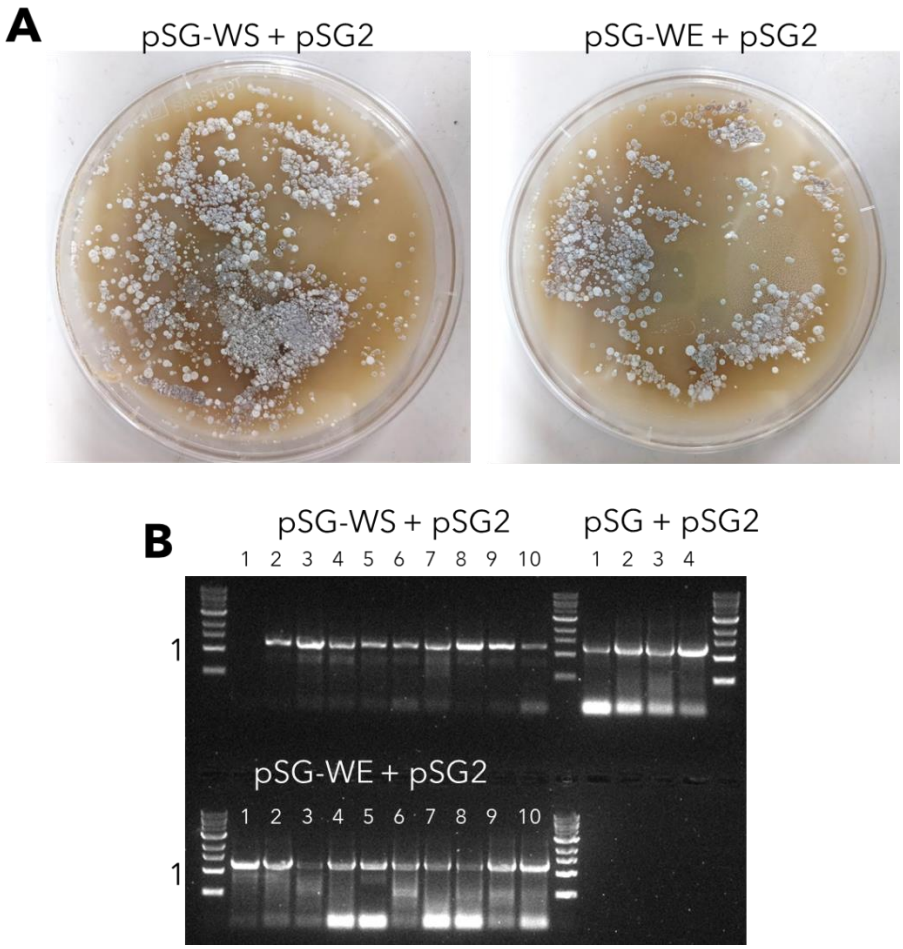


Figure 3.9: Initial screening of *S. coelicolor* M1146 pSG-WS/pSG-WE pSG2 exconjugants. A – Conjugation plates 5 days over initial conjugation, with pSG2 being conjugated into *S. coelicolor* M1146 pSG-WS (left) and *S. coelicolor* M1146 pSG-WE (right). **B** – Colony PCR screen of *S. coelicolor* M1146 pSG-WS pSG2/ *S. coelicolor* M1146 pSG-WE pSG2/ *S. coelicolor* M1146 pSG pSG2. Primers screened for the presence of *genB1*, carried on pSG2 (expected band size = 1151 bp; C1a-genB1rbsfw and C1a-genB1rev). For *S. coelicolor* M1146 pSG-WS pSG2, colonies 1-5 showed smaller grey phenotype, colonies 6-10 showed larger white phenotype; for S. coelicolor M1146 pSG-WE pSG2, colonies 1-2 showed larger grey phenotype, colonies 3-6 showed smaller grey phenotype, colonies 7-10 showed larger white phenotype; for *S. coelicolor* M1146 pSG pSG2, all four colonies screened (not pictured) showed larger white phenotype.

S. coelicolor M1146 pSG-WS pSG2 and *S. coelicolor* M1146 pSG-WE pSG2 were cultured in TSB/R2YE production medium for 7 days prior to concentration of samples and testing via antimicrobial bioassay against *B. subtilis* **(Figure 3.10)**. Two strain lineages were generated from independent colonies from original conjugation, marked 1 and 2 on the bioassay. Interestingly, only one culture,

replicate 2 of *S. coelicolor* M1146 pSG-WS pSG2, showed bioactivity from one strain lineage. This strain contained stronger promoters controlling the gentamicin A2 biosynthesis pathway in comparison to the other variant, *S. coelicolor* M1146 pSG-WE pSG2. Therefore, production should be more likely to be visible in this strain than in the variant with biosynthesis controlled by weaker *ermEP1* promoters. As this is a clear outlier, it is unlikely to represent true bioactivity resulting from the synthetic constructs, though this cannot be confirmed without using analytical methods to determine the metabolites present.

As the strains had shown poor growth upon passaging, it was posited that perhaps a loss of the pathway genes was causing the pathway to be truncated at a certain point in biosynthesis. A first PCR screen was carried out to detect presence of the *genS1-genC* fragment in pSG-WS/pSG-WE, and for the *genQ-genB1* fragment in pSG2. The *genS1-genC* fragment represents the first two steps of biosynthesis of gentamicin; as such, if this is missing, no further biosynthesis is possible.[26] The *genQ-genB1* fragment represents the catalytic step between gentamicin X2 and JI-20A (the immediate precursor to gentamicin C1a).[12] Gentamicin X2 has been shown to have a lower bioactivity than the gentamicin C congeners,[28] and so if this was the only product able to be produced, antibiotic activity may not be visible using antimicrobial bioassay. In *S. coelicolor* M1146 pSG-WS and in *S. coelicolor* M1146 pSG-WS pSG2 samples, 5 of the possible 9 strain replicates carrying pSG-WS appear to have lost the *genS1-genC* fragment. All the spore suspensions used as original liquid culture inoculum, however, maintain this fragment. In *S. coelicolor* M1146 pSG-WE and in *S. coelicolor* M1146 pSG-WE pSG2 samples, only 2 of the possible 9 strain replicates carrying pSG-WS appear to have lost the *genS1-genC* fragment. It therefore may be the case that the high strength of the *sp44* promoter conveys a level of toxicity to the strain, even without production of a bioactive metabolite (*S. coelicolor* M1146 pSG-WS). Meanwhile, no detection of the *genQ-genB1* fragment was possible across any of the strains, including the parental strains used as inoculum for the liquid cultures. It appears that

the passaging on solid medium between original conjugation and generation of spore suspension was sufficient for this fragment to be lost. It was not possible to determine the entire boundary of the lost fragment, but it is likely that the entire pSG2 vector was lost from the strain (as we saw previously with loss of the core biosynthetic region of C24).[30]

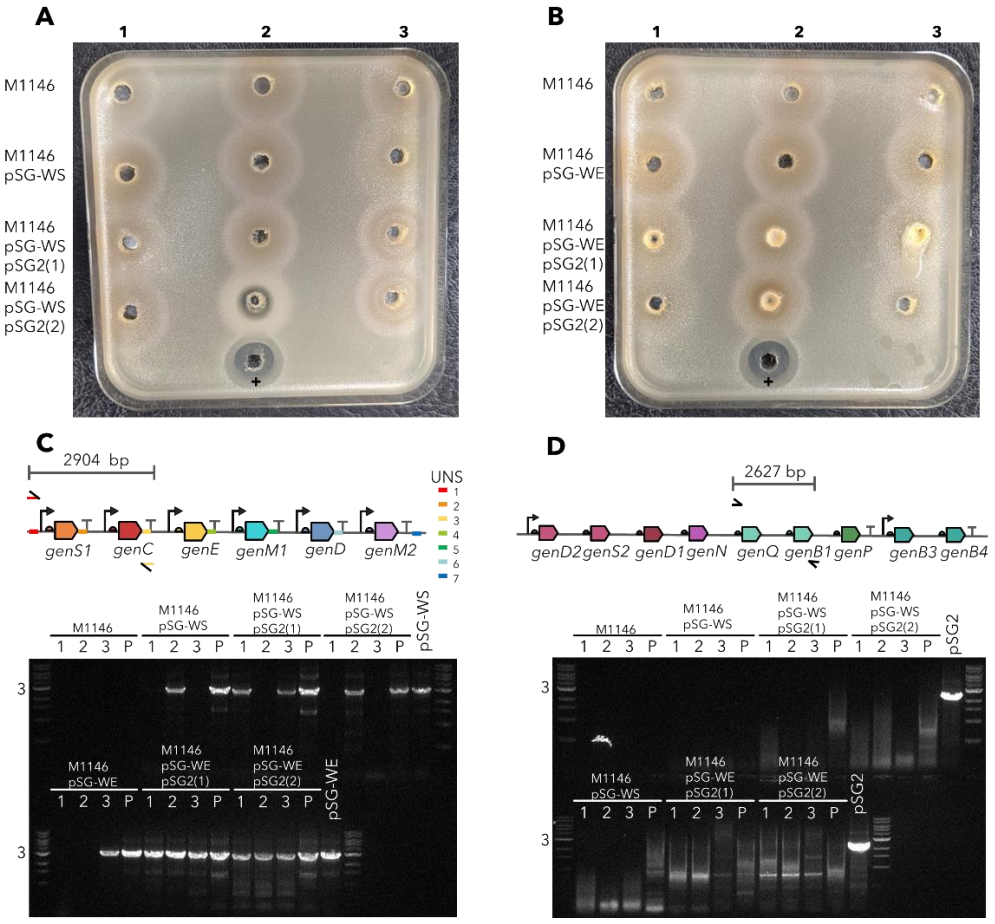


Figure 3.10: Antimicrobial bioassays and subsequent screening of plasmid presence within tested strains carrying minimal pathways. A – Antimicrobial bioassay of culture extracts from *S. coelicolor* M1146, *S. coelicolor* M1146 pSG-WS, and *S. coelicolor* M1146 pSG-WS pSG2 (two test strains). + = 4 μg gentamicin sulfate, indicator strain = *B. subtilis* 168. **B** – Antimicrobial bioassay of culture extracts from *S. coelicolor* M1146, *S. coelicolor* M1146 pSG-WE, and *S. coelicolor* M1146 pSG-WE pSG2 (two test strains). + = 4 μg gentamicin sulfate, indicator strain = *B. subtilis* 168. **C** - Screening of *S. coelicolor* M1146 pSG-WS/pSG-WE pSG2 cultures for the presence of *genS1* and *genC* using primers UNS1F and UNS3R. A band at 2904 bp suggests maintenance of these two genes within the strain. Strain used as inoculum for each culture is marked 'P'; 1-3 refer to samples taken from replicate liquid cultures after 7 days of growth. Purified pSG-WS or pSG-WE were used as template as a positive control for the PCR. **D** – Screening of *S. coelicolor* M1146 pSG-WS/pSG-WE pSG2 cultures for the presence of *genS1* and *genC* using primers C1a-genQrbsfw and C1a-genB1rev. A band at 2627 bp suggests maintenance of these two genes within the strain. Strain used as inoculum for each culture is marked 'P'; 1-3 refer to samples taken from replicate liquid cultures after 7 days of growth. Purified pSG2 was used as template as a positive control for the PCR.

Full sequencing of the new biosynthetic regions would allow for determination of which genes were maintained, and which were lost. This would also allow for an estimation of which genes cause a higher degree or lower degree of toxicity to the host strain, assuming that genes with a highly toxic effect will be less well tolerated. It can be assumed that any genes allowing for the production of bioactive product are more likely to be lost from the strain, and this would include the *genN-genD1-genB1-genQ-genP-genB3-genB4* fragment which forms the majority of the pSG2 insert.

It is clear that the lack of resistance of the strain (in comparison to the panel of resistant strains carrying C24 we previously generated)[30] may have been causative of this loss of pathway genes. Testing the function of pRes as an aid to increasing aminoglycoside resistance would be a useful next step. This could be done through carrying out comparative growth experiments on *S. coelicolor* carrying pRes and without, with increasing concentrations of gentamicin sulfate to determine the MIC for both strains. We originally aimed to further improve upon pRes by a further iteration by i) changing out the *ermEp* promoter for a stronger promoter, either *kasO*p* or *sp44* ii) adding an extra region into the plasmid from Cluster 24 (*genH-genI-genT*). Another 16S rRNA methylase, *gmrB*, had also been identified as a potential candidate, but Wan et al. (2018) report that gene knockout experiments suggest that it makes little change in levels of resistance of *M. purpurea* G1008 to gentamicin.[49] By sequence comparison, *genH* is predicted to be a putative gentamicin exporter (closest match a putative gentamicin exporter of 99.12% identity containing conserved MFS transporter domains, *M. echinospora*; Accession: AGB13922.1), *genI* a putative gentamicin exporter (closest match a putative gentamicin exporter of 99.68% identity, *M. echinospora*; Accession: AGB13923.1) and *genT* a putative 16S rRNA (cytidine(1402)-2'-O)-methyltransferase (95.51% identity with predicted rRNA small subunit methyltransferase I domains, *M. echinospora*; Accession: WP_088981615.1) therefore, these genes represent targets with predicted function similar to the *gmrA* and *genV* already present in pRes. While introduction of the biosynthetic and existing pRes plasmids into

the mutant *S. coelicolor* M1146 strains (generated as part of our previous work)[30] may provide a high enough level of resistance to aminoglycosides to not require these extra genes in the minimal pathway, this may be a useful further option in case resistance proves to be a bottleneck in increasing titres.

3.3.3. Insertion of a constitutive promoter cassette to Cluster 24 using CRISPR-Cas9 yields an increase in bioactivity

To further increase antibiotic yield for characterisation, it was decided to use CRISPR-Cas9 to insert a promoter cassette into the putative core biosynthetic region of Cluster 24. This had been tested on the parental strain *S. coelicolor* M1146 C24 prior to generating the resistant strains. However, upon conjugation of the plasmid into *S. coelicolor* M1146 C24, no colonies were able to be obtained and it was hypothesised that replacement of the promoters may become lethal in strains not resistant enough to gentamicin.

The best region of insertion was selected based on visual inspection of the BGC and subsequent prediction of promoter presence via BPROM **(Figure 3.11)**.[50] BPROM is a web tool that identifies similarities between the target sequence and known *E. coli* promoter sequences, to predict -10 and -35 sites RNA-polymerase binding sites within a specific genomic region. The chosen site of insertion should positively impact transcription levels of 12 ORFs of the 34 homologous to those with biosynthetic relevance in the *M. echinospora* DSM 43816 gentamicin gene cluster.

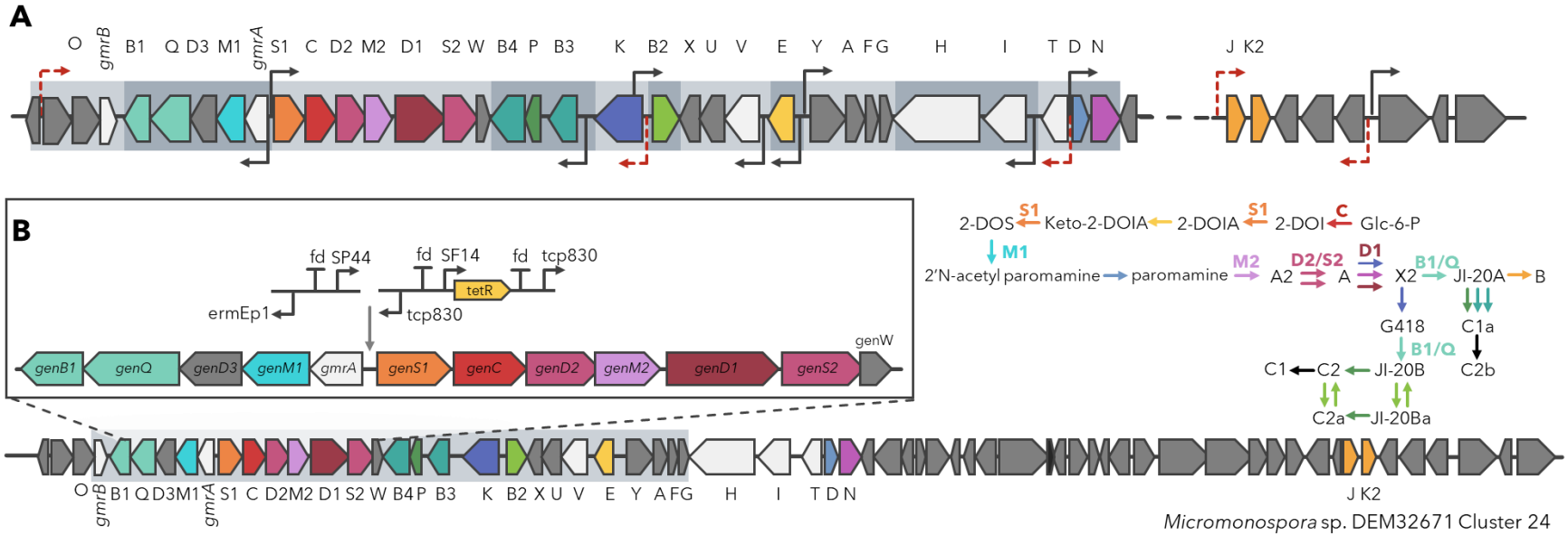


Figure 3.11: Schematic of promoter locations in Cluster 24. A – predicted promoters in Cluster 24 using BPROM.[50] Promoters successfully detected are marked in black; regions where promoters are likely to be present to allow for cluster function are marked in dashed red line. Predicted operons are marked in alternating light and dark grey. **B** – site of insertion of designed promoter cassettes to Cluster 24. Gene homologues likely to be impacted by the promoter insertion and implicated in gentamicin biosynthesis are marked with the last letters of their name on the biosynthetic pathway.

For the constitutive promoter cassette, two promoters were selected based on strength and prior experience using these in *S. coelicolor*. *sp44* was derived from the coelimycin BGC promoter *kasO** by Bai et al. (2015) and is regarded as one of the strongest synthetic promoters currently available for *Streptomyces*.[41,51] It was chosen for the expression of homologues of very early pathway enzymes: *genS1* and *genC* carry out the first and second steps of gentamicin biosynthesis respectively.[26] A weaker promoter, *ermEp1,* was chosen as a second promoter as it had no sequence similarity to *sp44*; additionally, it is a well-documented promoter for engineering of *Streptomyces*.[41] The cassette for insertion was designed to incorporate a 30 bp sequence upstream of both *gmrA* and *genS1* to ensure maintenance of the native ribosome binding site, alongside each promoter and the *fd* terminator. The terminator was insulated on each side with 198 bp and 160 bp random sequence screened for presence of transcriptional elements prior to insertion. For the inducible promoter cassette, the TetRiS system was utilised, due to being used previously within our group.[37] Both *genB1-genQ-genD3-genM1-gmrA* and *genS1-genC-genD2-genM2-genD1-genS2-genW* putative operons were placed under control of the *tcp830* promoter, and *tetR* was placed under control of constitutive promoter *sf14*, flanked by terminator sequences to avoid read-through transcription of the Cluster 24 operons. The cassette was cloned into the pCM4.4 vector flanked by respective homology arms to yield pCM4.4C24bi-dc **(Figure S3.8; Figure S3.9)** and CRISPR-Cas9-mediated modification was carried out as described by Ye et al.[31] As with the CRISPR knockout plasmids, BLAST was used to identify a protospacer with unique sequence within both the *S. coelicolor* genome and pESAC13-C24 (all combinations of NGG unique; 5′ – GCAATAACTTTAAGTACACG – 3′). Homology arms were taken from 12,305 bp – 13,305 bp and 13,544 bp – 14,544 bp for *gmrA* and *genS1* flanking regions respectively.

The growth of the strain after conjugation of the promoter cassette insertion plasmids was immediately affected, with conjugation of pCM4.4C24bi-dc resulting in only three colonies and of pCM4.4C24bi-di resulting in only two

colonies. After streaking these out, it was determined by PCR screening that all three colonies with pCM4.4C24bi-dc conjugated had successfully inserted the new constitutive promoter fragment, and the generation of the strain was confirmed by Sanger sequencing **(Figure S3.9)**. Unfortunately, a strain carrying the inducible promoter cassette inserted into Cluster 24 was unable to be verified. Cultures of the strain with inserted constitutive promoter cassette were grown with standard production conditions identical to those used previously; samples concentrated by freeze-drying and then bioassayed against *B. subtilis* **(Figure 3.12)**. A clear halo (marked with *) suggesting production of antibiotic by *S. coelicolor* M1146 C24 2R carrying the inserted promoter cassette was seen in one of three replicates. From this bioassay, titre of aminoglycoside in 250 mL shake flasks (when comparing to standards of gentamicin sulfate of known concentration) still appears to be on the μg/mL scale. It appears this is an increase in comparison to the parental strain;[30] however this cannot be confirmed without using a dilution factor and sample volume where bioactivity from the parental strain can be visualised on the same bioassay plate.

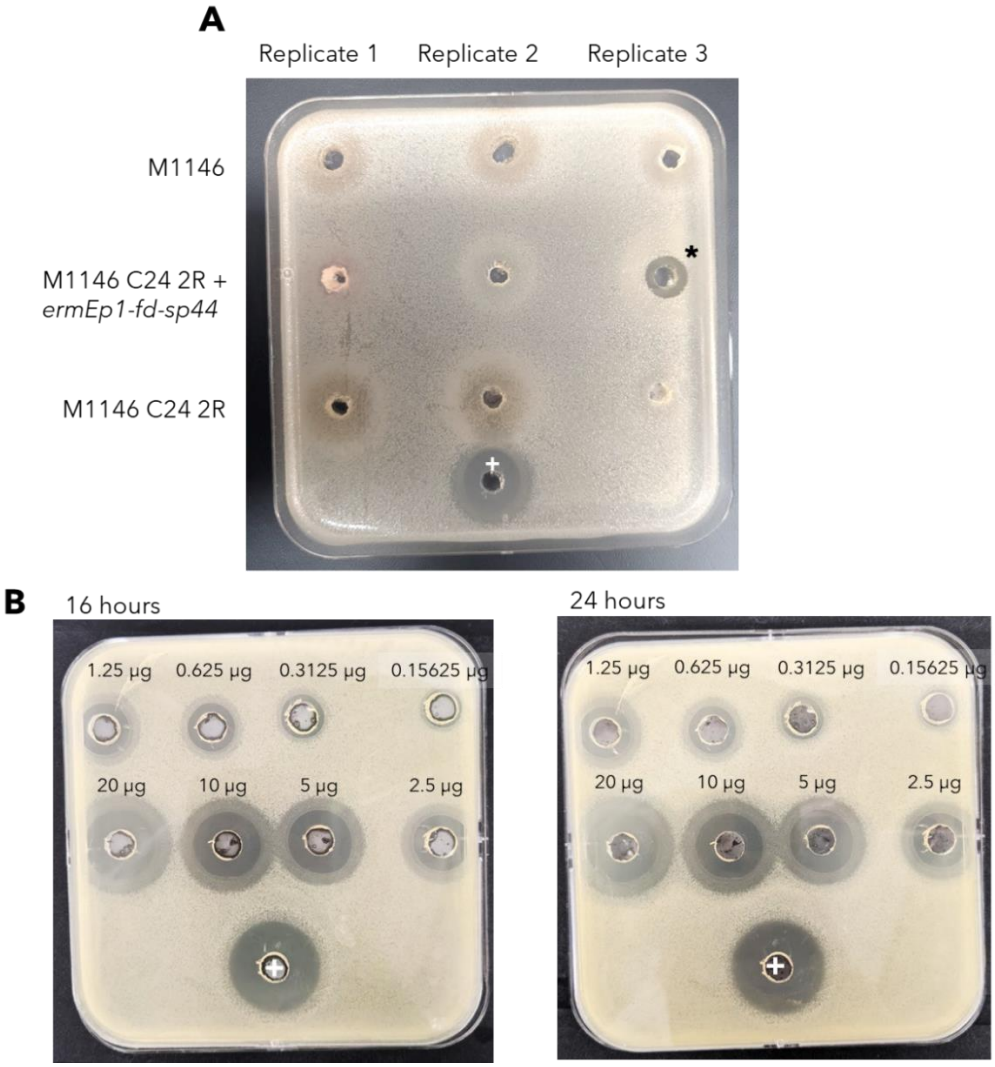


Figure 3.12: Antimicrobial bioassays to estimate production yield of strains with constitutive promoter cassette inserted between *genS1* and *gmrA*. A – Bioassay of *S. coelicolor* M1146 C24 2R and *S. coelicolor* M1146 C24 2R + promoter cassette against *B. subtilis* 168. Samples were concentrated 25X by freeze drying and 50 µL added to each well. + = gentamicin sulfate, 5 µg. **B –** Serial dilutions of gentamicin sulfate in sterile ddH_2O, bioassayed against *B. subtilis* in LB agar and photographed twice at 16 and 24 hours of growth at 37°C. + = 10 µg of chloramphenicol.

These samples were again analysed by LC-MS to determine the cause of antibiotic activity **(Figure 3.13)**. Three different chromatographic methods were tested using three different columns: Waters Atlantis T3, Zorbax SB-C8, and Waters Xbridge Amide. After scanning all samples for all 86 aminoglycosides in the Dictionary of Natural Products,[52] the only aminoglycoside-related mass identified was of m/z 502.2241, as we described

previously (Chapter 2).[30] This mass was only detected from the asterisked sample in Figure 3.12, which was the only the sample with visible bioactivity against *B. subtilis*. Additionally, the same mass was only detectable in the chromatograms using the Waters Xbridge Amide column, used in hydrophilic interaction liquid chromatography (HILIC; Figure 3.13). HILIC has been shown to be a better option than standard reversed-phase chromatography for detection of aminoglycosides: as ion-pairing reagents are required for the highly polar aminoglycosides to be retained on reversed-phase column matrices, this can lead to contamination of instruments over time and cause a loss of sensitivity of the mass spectrometer.[53–55] However, it is interesting that detection of the ion of m/z 502.2241 was only possible on the amide column when it was able to be detected on the C18-based Waters Atlantis T3 column previously with different samples (Chapter 2). The peak shape is not as standard in this case due to the low intensity of the product ion, and the entire chromatogram retains some noise due to the complexity of the sample. Most polar species in the TSB/R2YE would have been retained upon passing through the C18 SPE to clean up the sample, and few other purification techniques without derivatization of gentamicin or related aminoglycosides are available.[56–58]

In our previous work (Chapter 2), we were also able to detect aminoglycoside precursors 2-deoxystreptamine, 2-deoxy-scyllo-inososе, 2-deoxy-scyllo-inosamine and paromamine in the samples with bioactivity from strains carrying Cluster 24.[30] Here, we were not able to detect any of these in any sample, including the sample with bioactivity.

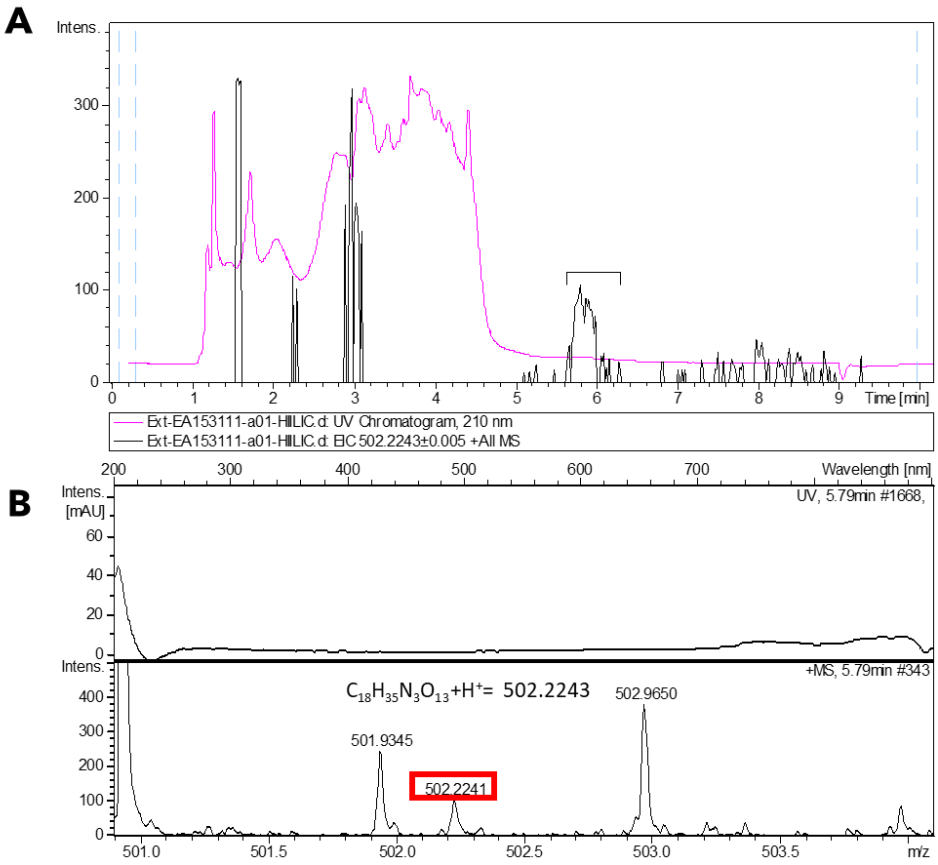


Figure 3.13: LC-MS chromatograms of species of m/z 502.2241, predicted to be representative of $C_{18}H_{35}N_3O_{13}$ + H (monoisotopic mass = 502.2248). A – chromatogram of total run of *S. coelicolor* M1146 C24 2R + promoter cassette sample with bioactivity. Purple trace is UV chromatogram (210 nM), black chromatogram is extracted ion chromatogram at masses 502.2243 ± 0.005. **B** – Zoomed in chromatogram of peak highlighted in A. Mass of interest is highlighted in red.

To fully characterise the entire cluster output, it would be necessary for LC-MS method development to be prioritised, considering the low concentrations of gentamicin we expect are being produced in this work. The low yield is also coupled with a possible ion suppression effect from the media components, meaning that concentrations of gentamicin up to and including 62.5 μg/mL in the TSB/R2YE media used for culturing were unable to be detected (no higher concentrations tested). Conversely, we were able to detect gentamicin at concentrations as low as 0.977 μg/mL when this was dissolved in HPLC-grade water (Chapter 2; no lower concentrations tested).[30] As previously mentioned, purification of aminoglycosides is challenging due to

their hydrophilicity.[59] It is likely that a balancing of i) appropriate LC-MS methodology with ii) fine-tuning suitable media components (which do not have an ion suppression effect on aminoglycosides) would be required. Nevertheless, it is interesting that this mass was again detected by strains carrying Cluster 24, suggesting this may be a true output of the cluster. The production of this compound, and bioactivity, was only seen in one of three replicates. However, within this work, we were unable to further investigate this inconsistency in production levels and note that until the cause of this is established, production may not be reproducible between culture flasks. Additionally, we were unable to fully establish whether there were any masses attributed to non-aminoglycosides which were present only in this bioactive sample or across all samples from strains carrying the cluster. This information would be of interest for further production strategies.

Insertion of promoter cassettes has previously been shown in many cases to be useful to awaken or upregulate BGC expression;[21,60,61] in the work carried out by Bauman et al. (2019), insertion of a single bi-directional promoter cassette was enough for all biosynthetic genes in their cluster to be transcribed.[22] If the host strain of choice is genetically amenable, insertion of a constitutive promoter cassette can also assist in cases where mutating for over-production is not easily screened (for example, the white to violet colour change seen by Estévez et al. of their streptomycin-resistant, fredericamycin-overproducing *S. albus* strain).[62] In our work, use of RT-qPCR to investigate how many cluster genes had expression 'switched on' by the insertion of promoter cassette would have been highly valuable.

To further increase the titre of antibiotic present, from our strains carrying the inserted promoter cassette, experiments could be carried out to determine the expression levels of each Cluster 24 gene of *M.* sp. DEM32671 through RT-qPCR. This would confirm the operons predicted *in silico* by BPROM, and allow for estimation of relative promoter strength for each operon. Matching the experimentally-determined operon strength to that of the new promoters

inserted would allow for a more elegant design. By ensuring the ratio of the cluster proteins remains as similar to the native host as possible, it may be more likely that production yields can be optimised. Pathway balancing in this way was previously shown to be important in the production of Taxol precursors in *E. coli*: strains with higher production of the oxygenated diterpenes of interest were shown with targeted RT-qPCR to have a lower expression of a key synthase earlier in the biosynthetic pathway.[63] Recently, Ji et al. (2022) tested different promoter strengths at various points in the daptomycin pathway, and found that combinations of strong and weak promoters yielded higher product titres than only using strong promoters. While insertion of an inducible promoter was not able to yield a useful strain in this work, this approach could be revisited knowing that the promoter-insertion strategy is successful for Cluster 24. Instead, a more tightly-regulated inducible expression system, such as the resorcinol and cumate inducible systems described by Horbal et al. (2014), or synthetic inducible systems described by Ji et al. (2019) could be utilised for this purpose.[64,65]

Another strategy forwards for yield improvement could lie in a co-culture system, as gentamicin precursor 2-deoxy-scyllo-inosose has been shown recently to be well-produced in in *B. subtilis* (38 g/L best achieved yield after improvement of culture conditions and testing alternative enzyme candidates).[66] The stark difference in growth rate between a *Bacillus* and *Streptomyces* strain could make such a co-culture system incredibly challenging to construct, but could improve the biosynthetic potential of the *Streptomyces* strain significantly. *E. coli* co-culture for sharing of cellular burden has been well described;[67–69] along with balancing the growth rate, the main challenge here would be to ensure i) enough 2-deoxy-scyllo-inosose would be able to be taken up by the *Streptomyces* strain; ii) the produced aminoglycoside congeners would not have an inhibitory effect on the growth of the *Bacillus* strain iii) the balancing of media components was suitable for growth and production from both strains.

3.4. Conclusion

In this work we attempted production of gentamicin C1a and gentamicin A2 congeners selectively, along with an improvement of antibiotic production by *Streptomyces* strains carrying Cluster 24 of novel *Micromonospora M.* sp. DEM32671. Our selective congener production experiments were not able to be fully realised due to i) low initial yield of antibiotic ii) loss of minimal pathway genes through culturing, likely due to the strain's resistance to aminoglycosides being too low for sustained production. With our additional work in this area,[30] we have established routes to yield improvement and increased strain resistance to aminoglycosides; therefore, with future work a road towards these selective congeners can still be realised. Additionally, we tested a strategy towards improvement of antibiotic yield by inserting a promoter cassette into Cluster 24 for upregulation of the predicted early biosynthetic genes. While this production was not reproducible across all replicates, we were again able to continue to characterise one output of Cluster 24 as species of proposed formula $C_{18}H_{35}N_3O_{13}$, and have generated a strain ready for application of our methods to produce select congeners of gentamicin.

3.5. Materials and Methods

3.5.1. Strains and plasmids

All strains and plasmids used and generated in this work are listed in **Table 3.2**.

Table 3.2: Bacterial strains and plasmids used in this work.

Bacterial strain	Genotype	Use	Reference
Escherichia coli NEB5α	*fhuA2 (argF-lacZ)U169 phoA glnV44 80 (lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	Cloning	[70]
E. coli ET12567 pUZ8002	*Fdam13::Tn9 dcm6 hsdM hsdR zjj202::Tn10 recF143 galK2 galT22 ara14 lacY1 xyl5 leuB6 thi1 tonA31 rpsL136 hisG4 tsx78 mtl1 glnV44* pUZ8002	Conjugation of plasmids into *Streptomyces*	[71]
Bacillus subtilis 168	*trpC2*	Indicator strain for bioassays	[72,73]
Streptomyces lividans TK23	*spc-1* SLP2- SLP3-	Expression of Cluster 24-derived plasmids	[42,74]
Streptomyces coelicolor M1152	*S. coelicolor* M145 *Δact Δred Δcda Δcpk rpoB*[C1298T])	Expression and modification of Cluster 24-derived plasmids	[46]
Streptomyces coelicolor M1146	*S. coelicolor* M145 *Δact Δred Δcda Δcpk*	Expression of Cluster 24-derived plasmids	[46]
Streptomyces coelicolor M1152 C24	*S. coelicolor* M145 *Δact Δred Δcda Δcpk rpoB*[C1298T]) C24	Modification of Cluster 24	[30]

Streptomyces coelicolor M1152 C24 Δ*genK*	*S. coelicolor* M145 *Δact Δred Δcda Δcpk rpoB*[C1298T]) C24[Δ*genK*]	Production of gentamicin C1a	This work
Streptomyces coelicolor M1152 C24 Δ*genK ΔgenJ ΔgenK2*	*S. coelicolor* M145 *Δact Δred Δcda Δcpk rpoB*[C1298T]) C24[Δ*genK ΔgenJ ΔgenK2*]	Production of gentamicin C1a	This work
Streptomyces coelicolor M1146 C24 2R	*S. coelicolor* M145 *Δact Δred Δcda Δcpk* C24 (92 mutations, detailed in **Chapter 2**)	Modification of Cluster 24	[30]
Streptomyces coelicolor M1146 C24 2R + *ermEp1-fd-sp44*	*S. coelicolor* M145 *Δact Δred Δcda Δcpk* C24(*ermEp1-fd-sp44*)	Modification of Cluster 24	This work
Plasmid	**Backbone**	**Insert**	**Reference**
pSG	pSET152 (ϕC31, Apra^R)	*tetR-SF14-tcp830-genS1-genC-genE-mmr(5g)-tcp830-genM1-genD-genM2-fd*	[38]
pSGCH	pSET152 (ϕC31, Apra^R)	*tetR-SF14-tcp830-genS1-genC-genE (C-terminal 6xHis)-mmr(5g)-tcp830-genM1-genD-genM2 (C-terminal 6xHis)-fd*	This work
pSG-WS	pSET152 (ϕC31, Apra^R)	*sp44-genS1-fd-80bp-sp44-genC-fd-80bp-sp44-genE-fd-80bp-sp44-genM1-fd-80bp-sp44-genD-fd-80bp-sp44-genM2-fd*	This work
pSG-WE	pSET152 (ϕC31, Apra^R)	*ermEp1-genS1-fd-80bp- ermEp1-genC-fd-80bp- ermEp1-genE-fd-80bp-ermEp1-genM1-fd-*	This work

		80bp- ermEp1-genD-fd-80bp- ermEp1-genM2-fd	
pSG2	pTE1364 (ϕSV1, modified marker Kan -> Thio; original pHG5)	*sp44-genD2-genS2-genD1-genN-genQ-genB1-genP-sbiB-kasO*p-genB3-genB4-fd*	This work
pRes	pT801 (ϕBT1, Hyg^R)	*ermEp-gmrA-genV*	[38]
pCM4.4	pCRISPomyces 2.0 (replicative, $Apra^R$)	UNS2-*fd-ermE**-spCas9-*fd*-gapdhp-gRNA-tracr-ori-UNS6	[31]
pCM4.4ΔgenK	pCM4.4 (replicative, $Apra^R$)	UNS6-DNA repair template-UNS7	This work
pCM4.4ΔgenJgenK2	pCM4.4 (replicative, $Apra^R$)	UNS6-DNA repair template-UNS7	This work
pCM4.4C24bi-dc	pCM4.4 (replicative, $Apra^R$)	UNS6-half DNA repair template-*ermEp1-fd-sp44-* half DNA repair template-UNS7	This work
pCM4.4C24bi-di	pCM4.4 (replicative, $Apra^R$)	UNS6- half DNA repair template-*tcp830-fd-tetR-sf14-tcp830-* half DNA repair template-UNS7	This work
pTE143	pSET152 (ϕC31, $Apra^R$)	*ptipA-scbR(C-terminal polyhistidine tag)-fd*	[40]

3.5.2. Media and cultivation conditions

S. coelicolor was grown on SFM agar (2% agar, 2% mannitol, 2% soy flour) for collection of spore suspensions. Strains carrying Cluster 24 or pSG2 were maintained on thiostrepton (50 μg/mL), strains carrying pSET152-derivative plasmids were maintained on apramycin (50 μg/mL), strains carrying pRes were maintained on hygromycin (50 μg/mL). For tests with strains carrying

pSG, 100 ng/mL anhydrotetracycline was used as inducer. All plates were incubated at 30°C.

For *Streptomyces* liquid cultures, 50 mL of media was used in a siliconized 250 mL Erlenmeyer flask. 10 mM diameter coil springs were inserted to ensure dispersed growth. Seed cultures were grown by inoculation of tryptone soya broth (TSB; 1.7% pancreatic digest of casein, 0.3% enzymatic digest of soya bean, 0.5% sodium chloride, 0.25% K_2HPO_4, 0.5% glucose; obtained as pre-mixed powder from Oxoid cat. no. CM0129) with ~ 10^8 CFU spore suspension for 48 hours at 180 rpm and 30°C. For production cultures, 1 mL of seed culture was used for inoculation to TSB/R2YE production medium (1:1 ratio of unautoclaved TSB and previously autoclaved R2YE Media A (10.3% sucrose, 1% glucose, 1.12% $MgCl_2.6H_2O$, 0.025% K_2SO_4, 0.01% Difco casamino acids, 0.5% Difco yeast extract).[42] GYM medium (0.4% glucose, 0.4% yeast extract, 1% malt extract) and R2YE medium (10.3% sucrose, 1% glucose, 1.12% $MgCl_2.6H_2O$, 0.025% K_2SO_4, 0.01% Difco casamino acids, 0.5% Difco yeast extract, 0.05% KH_2PO_4, 0.57% TES Buffer, 0.3% $CaCl_2.2H_2O$, 0.3% L-proline, 0.02% NaOH, 0.2% trace elements solution) was also used for production tests when specified.[42] Liquid cultures were grown without antibiotics. For tests with strains carrying pSG, 100 ng/mL anhydrotetracycline was used as inducer.

E. coli NEB5α, *E. coli* ET12567 pUZ8002, and *B. subtilis* 168 were cultivated in Lysogeny Broth Miller (LB) (Formedium; 1% NaCl, 1% tryptone, 0.5% yeast extract) at 37°C and 180 rpm. Lysogeny Broth Miller (LB) agar (Formedium; as above but with 1.5% agar) was used as a solid medium. Where hygromycin was used, Lysogeny Broth Lennox was used (0.5% NaCl) to ensure proper antibiotic function.

3.5.3. Construction of minimal pathway plasmids

Unless stated otherwise, parts for the assembly of minimal pathways were generated via PCR (PrimeSTAR Max DNA polymerase Takara; cat. no. R045A) using the following cycle: initial denaturation 98°C for 2 min; amplification (35X cycles) 98°C for 10s, 55°C for 15s, 72°C for 45s. PCR primers are described in **Table S3.1**.

For pSG, the two three-gene operons (*tcp830-genS1-genC-genE-mmr(5g)*, *tcp830-genM1-genD-genM2-fd*) were ordered as synthetic DNA and pSET152 was used as vector template.

For pSGCH, pSG was used as template for the two three-gene operons and pSET152 was used as vector template.

For pSG-WS and pSG-WE, pSG was used as template for each coding sequence and primers included an overhang to either synthetic 'promoter block' or 'promoter' sequences. Promoter blocks (deftempspacer-*fd*-80bpspacer-*sp44*/*ermEp1*-deftempspacer; where 'deftempspacer' is 'spacer of defined temperature', 70°C) and promoters (deftempspacer-*sp44*/*ermEp1*-deftempspacer) were ordered as synthetic DNA from Integrated DNA Technologies (IDT); these were amplified with specific overhang on the reverse primer for each gene in pSG-WS/pSG-WE with the following cycle initial denaturation 98°C for 2 min; amplification (35X cycles) 98°C for 10s, 55°C for 15s, 72°C for 5s. A nested primer PCR with the forward primer for promoter block/promoter (promfw/promblkfw) and the respective reverse primer for each coding sequence was carried out using the following cycle: initial denaturation 98°C for 2 min; amplification (35X cycles) 98°C for 10s, 55°C for 15s, 72°C for 50s. UNS sequences were added through further PCR to each promoter/promoter block-coding sequence fusion in order UNS_n – fused promoter/promoter block-coding sequence – UNS_{n+1}.

For pSG2, each gene was individually amplified from purified *S. lividans* TK23 C24 gDNA, aside from *genB1* and *kasO*p-genB3-genB4-fd* which were ordered

as synthetic DNA from Thermo Fisher GeneArt. Additionally, pSG2 used a pTE1364 backbone which had the kanamycin resistance cassette swapped for a thiostrepton resistance cassette to avoid potential modification of the aminoglycoside product.

Plasmids were assembled according to the manufacturer's protocol with the In-Fusion® HD Cloning Plus (Takara, cat. no. 638909; pSG only) or NEBuilder® HiFi DNA Assembly Master Mix (all other plasmids; NEB, cat. no. E5520S) and transformed into chemically-competent *E. coli* NEB® 5*α* (NEB, cat. no. C2987H) following standard manufacturer's protocol. Correct assembly was confirmed with Sanger sequencing.

3.5.4. Construction of pCM4.4-based plasmids for CRISPR knockouts and knock-ins

For construction of pCM4.4C24bi-dc, pCM4.4C24bi-di, and pCM4.4ΔgenD1genS2, protospacer was inserted into 'empty' pCM4.4 with overhangs compatible with Golden gate assembly. Protospacers for each plasmid are described in text; as Cas9 modification site was the same for both pCM4.4C24bi-dc and pCM4.4C24bi-di only one precursor plasmid was generated. Golden gate assembly was performed with BbsI (NEB) and T4 DNA Ligase (NEB) with the following cycle: 10X cycles of 37°C and 16°C alternating (10 min each), then for heat inactivation of the ligase 50°C (5 min), then for heat inactivation of BbsI 65°C (20 min). After transformation using chemically competent *E. coli* NEB® 5*α* (NEB, cat. no. C2987H) following standard manufacturer's protocol, successful insertion (and loss of *lacZ* gene) was determined by blue-white screening and further Sanger sequencing. Variants of pCM4.4 with these protospacers inserted but without homology arms are termed pCM4.4C24bi-dp and pCM4.4ΔgenD1genS2p for ease of understanding. For pCM4.4ΔgenK and pCM4.4ΔgenJgenK2, protospacer was included into primers for HiFi assembly, omitting the first round of cloning.

Unless otherwise stated, amplification of fragments for HiFi assembly to insert the homology arms for repair template was carried out using PrimeSTAR Max DNA polymerase (Takara; cat. no. R045A) with the following cycle: initial denaturation 98°C for 2 min; amplification (35X cycles) 98°C for 10s, 55°C for 15s, 72°C for 40s.

For pCM4.4ΔgenD1genS2 assembly was carried out as a 4-part reaction: Fragment 1 (aac3(IV)-pSG5 rep-oriT-traJ-UNS2) was amplified with pCM4.4 HA site UNS7 fw and dLancI-II 1 rev; Fragment 2 (UNS2-*fd*-*ermE**-spCas9-*fd*-gapdhp-protospacer-gRNA-tracr-*ori*-UNS6) was amplified with Cas9 UNS2 fw and pCM4.4 HA site UNS6 rev; Fragment 3 (UNS6-5′ homologous arm) was amplified with genD1S2frag1fw and genD1S2frag1rev; Fragment 4 (3′ homologous arm-UNS7) was amplified with genD1S2frag2fw and genD1S2frag2rev. Fragments 1, 2 and 3 used pCM4.4ΔgenD1genS2p as template, Fragments 4 and 5 used genomic DNA extracted from *S. coelicolor* M1146 C24.

For pCM4.4ΔgenK assembly was carried out as a 5-part reaction: Fragment 1 (aac3(IV)-pSG5 rep-oriT-traJ-UNS2) was amplified with pCM4.4 HA site UNS7 fw and dLancI-II 1 rev; Fragment 2 (UNS2-*fd*-*ermE**-spCas9-*fd*-gapdhp) was amplified with Cas9 UNS2 fw and protospacer OLrevgenK; Fragment 3 (protospacer-gRNA-tracr-*ori*-UNS6) was amplified with protospacer OLfwgenK and pCM4.4 HA site UNS6rev; Fragment 4 (UNS6-5′ homologous arm) was amplified with genKfrag1fw and genKfrag1rev; Fragment 5 (3′ homologous arm-UNS7) was amplified with genKfrag2fw and genKfrag2rev. Fragments 1, 2 and 3 used pCM4.4 as template, Fragments 4 and 5 used genomic DNA extracted from *S. coelicolor* M1146 C24.

For pCM4.4ΔgenJgenK2 assembly was carried out as a 5-part reaction: Fragment 1 (aac3(IV)-pSG5 rep-oriT-traJ-UNS2) was amplified with pCM4.4 HA site UNS7 fw and dLancI-II 1 rev; Fragment 2 (UNS2-*fd*-*ermE**-spCas9-*fd*-gapdhp) was amplified with Cas9 UNS2 fw and protospacer OLrevgenJK2; Fragment 3 (protospacer-gRNA-tracr-*ori*-UNS6) was amplified with

protospacer OLfwgenJK2 and pCM4.4 HA site UNS6rev; Fragment 4 (UNS6-5′ homologous arm) was amplified with genJK2frag1fw and genJK2frag1rev; Fragment 5 (3′ homologous arm-UNS7) was amplified with genJK2frag2fw and genJKf2rag2rev. Fragments 1, 2 and 3 used pCM4.4 as template, Fragments 4 and 5 used genomic DNA extracted from *S. coelicolor* M1146 C24.

For pCM4.4C24bi-di assembly was carried out as a 5-part reaction: Fragment 1 (aac3(IV)-pSG5 rep-oriT-traJ-UNS2) was amplified with pCM4.4 HA site UNS7 fw and dLancI-II 1 rev; Fragment 2 (UNS2-*fd-ermE**-spCas9-*fd*-gapdhp) was amplified with Cas9 UNS2 fw and pCM4.4 HA site UNS6 rev; Fragment 3 (UNS6-5′ homologous arm) was amplified with gmrAUNS6fw and grmAtetrev; Fragment 4 (*tcp830-fd-tetR-sf14-tcp830*) was first amplified as two halves (t1, t2) due to repetitive sequences at each end using primers tetblock1fw/tetblock1rev and tetblock2fw/tetblock2rev, before carrying out nested primer PCR with primers tetblockflankfw and tetblockflankrev and t1+t2 as template; Fragment 5 (3′ homologous arm-UNS7) was amplified with genS1tetfw and genS1UNS7rev. Fragments 1 and 2 used pCM4.4C24bi-dp as template, Fragment 3 was ordered from Integrated DNA Technologies as a synthetic gene fragment (gBlock) whilst Fragments 4 and 5 used genomic DNA extracted from *S. coelicolor* M1146 C24. Fragment 4 sub-fragments t1 and t2 used the following cycle for PCR: initial denaturation 98°C for 2 min; amplification (35X cycles) 98°C for 10s, 55°C for 15s, 72°C for 10s.

For pCM4.4C24bi-dc assembly was carried out as a 5-part reaction: Fragment 1 (*aac3(IV)-pSG5 rep-oriT-traJ*-UNS2) was amplified with pCM4.4 HA site UNS7 fw and dLancI-II 1 rev; Fragment 2 (UNS2-*fd-ermE**-spCas9-*fd*-gapdhp) was amplified with Cas9 UNS2 fw and pCM4.4 HA site UNS6 rev; Fragment 3 (UNS6-5′ homologous arm) was amplified with gmrAUNS6fw and gmrAconstrev; Fragment 4 (*ermEp1-fd-SP44*) was amplified with ESconstbidampfw and ESconstbidamprev; Fragment 5 (3′ homologous arm-UNS7) was amplified with genS1constfw and genS1UNS7rev. Fragments 1 and 2 used pCM4.4C24bi-dp as template, Fragment 3 was ordered from

Integrated DNA Technologies as a synthetic gene fragment (gBlock) whilst Fragments 4 and 5 used genomic DNA extracted from *S. coelicolor* M1146 C24.

Agarose gel electrophoresis was used to confirm correct amplification of fragments. Fragments were treated with DpnI (NEB) for 18 hours (37°C) prior to assembly. Assembly was carried out according to the manufacturer's protocol with NEBuilder® HiFi DNA Assembly Master Mix (NEB, cat. no. E5520S) and transformed into chemically competent *E. coli* NEB® 5α (NEB, cat. no. C2987H) following standard manufacturer's protocol. After purification of plasmid using a QIAprep Spin Miniprep Kit (Qiagen, cat. no. 27106X4), correct construction was verified by Sanger sequencing.

3.5.5. Conjugation of plasmids to *Streptomyces* spp.

Competent cells of *E. coli* ET12567 pUZ8002 were obtained using a standard calcium chloride protocol and transformation carried out with standard heat-shock method.[75] Conjugation from *E. coli* ET12567 pUZ8002 (+ plasmid of interest) into *S. coelicolor* or *S. lividans* was carried out as described by Kieser et al. in Practical Streptomyces Genetics.[42] Screening for CRISPR-mediated knockouts or knock-ins or insertion of minimal pathway plasmids was done by using PCR with Terra polymerase (Takara; cat. no. 639270) on collected single colonies (initial denaturation 98°C for 2 min; amplification (35X cycles) 98°C for 10s, 68°C for 3 min) using primers as described for each construct in **Table S3.1**. After amplification, samples were checked via agarose gel and where relevant (CRISPR-mediated knockouts or knock-ins), modification was confirmed by Sanger sequencing.

3.5.6. Preparation of liquid samples for bioassay

Entire cultures were frozen at (-20°C) before thawing and centrifuging at 7000 xg for 5 minutes. An aliquot of culture supernatant was taken into a 50 mL Corning tube before snap-freezing in liquid nitrogen and further freezing at -

80°C for one hour. Samples were freeze dried for 22 hours and then resuspended in sterile ddH_2O to the concentration stated for each experiment.

3.5.7. Antibiotic bioassay

Bioassay was carried out against indicator strain *B. subtilis* (grown in LB-Miller broth, at 37°C, at 180 rpm, for 16 hours before subculturing again in the same medium). 50 mL agar inoculated with 50 µL *B. subtilis* culture (OD_{600} = 0.6) was used for each 120 mm square bioassay plate. To analyse levels of antibiotic production in samples from liquid culture, a 13 mM diameter hole was cut into the indicator plate and 50 µL of concentrated culture extract was added to each. The plate was incubated for 16 hours at 37°C before photographing.

3.5.8. Characterisation of cluster output

Samples for LC-MS were prepared by filtering culture supernatant through a C18 SPE cartridge (Thermo Scientific 60108-701, 2000 mg bed weight, 15 mL column capacity) for partial clean-up and were then snap-frozen in liquid nitrogen before freeze drying. LC-MS was carried out using three chromatographic set-ups:

i) Waters Atlantis T3 column (4.6 x 100 mm, 5 mm particle size) in an Agilent 1200 Series Rapid Resolution LC coupled to a Bruker maXis HR-qTOF mass spectrometer. Mobile phase: solvent A water:AcN 90:10, solvent B water:AcN 10:90, both with 13mM ammonium formate and 0.01% TFA. The gradient composition was: 100% A for 12 minutes, 90% A 10% B for 6 seconds, 100% B for 4 minutes 54 seconds, 100% A for 4 minutes. The flow rate was 0.5 mL/min throughout.

ii) Zorbax SB-C8 column (2.1 x 30 mm, 3.5 mm particle size) in an Agilent 1200 Series Rapid Resolution LC coupled to a Bruker maXis HR-qTOF mass spectrometer. Mobile phase: solvent A water:AcN 90:10, solvent B water:AcN 10:90, both with 13mM

ammonium formate and 0.01% TFA. The gradient composition was: 90% A 10% B for 6 minutes, 100% B for two minutes, 90% A 10% B for 2 minutes. The flow rate was 0.3 mL/min throughout.

iii) Waters Xbridge Amide column (2.1 x 100 mm, 3.5 mm particle size) in an Agilent 1200 Series Rapid Resolution LC coupled to a Bruker maXis HR-qTOF mass spectrometer. Mobile phase: solvent A water:AcN 90:10, solvent B water:AcN 10:90, both with 13mM ammonium formate and 0.01% TFA. The gradient composition was: 100% B for 5 minutes, 100% A for 3 minutes, 100% B for 3 minutes. The flow rate was 0.3 mL/min throughout.

For manual inspection of MS data, mzMatch/PeakML were used for the selection of i) peaks present only in samples from strains carrying C24; ii) selection of peaks of mass similar to adducts of gentamicin or gentamicin intermediates across all samples.[76]

3.5.9. SDS-PAGE and western blotting

Seed cultures were inoculated with ~10^8 CFU of chosen spores and grown for 3 days. After centrifugation, pellet was suspended in 0.2 volume lysis buffer (100 mM Tris-HCl pH 8.5, 20 mM sodium chloride, 1 mM DTT + 1 cOmplete tablet per 25 mL) and sonicated for two minutes on, 15s off, for ten minutes total sonication time (50% cycle duty, 50% power). Cultures were centrifuged at 15,000 xg for 30 minutes at 4°C. Supernatant was removed and taken as soluble fraction, pellet was resuspended in 0.2 volume of ddH_2O and taken as insoluble fraction. Samples to be analysed by SDS-PAGE were prepared with a 1:1 ratio of sample to loading dye (100 mM Tris-Cl (pH 6.8), 4% (w/v) SDS (sodium dodecyl sulfate; electrophoresis grade), 0.2% (w/v) bromophenol blue, 20% (v/v) glycerol, 200 mM β-mercaptoethanol). These were denatured (95°C, 10 minutes) and loaded and run on a precast Mini-PROTEAN® TGX Stain-Free™ gel (Bio-Rad, Cat. No. #456-8046) at 200V for 38 minutes. A molecular weight ladder (PageRuler Plus Prestained Protein Ladder, Thermo

Scientific, Cat. No. 26619) was also run for comparison to estimate the size of protein bands. Gels were stained with InstantBlue™ Protein Stain (Expedeon, Cat. No. 1SB1L) for 1 hour and imaged with a Gel Doc™ EZ Gel Documentation System. A duplicate gel was transferred to a nitrocellulose membrane using the Trans-Blot Turbo Transfer System (25V, 30 minutes) and western blotted using the iBind™ Western System according to manufacturer's instructions (Novex, Cat. No. SLF1000; primary antibody = Mouse Monoclonal Anti-polyHistidine antibody (Sigma, Cat. No. H1029), secondary antibody = Goat anti-Mouse IgG H&L (IRDye® 800CW) (Abcam, Cat. No. ab216772)). Western blots were imaged using a LI-COR Odyssey® Sa and Image Studio.

1. J. O'Neill, "The Review on Antimicrobial Resistance. Antimicrobial Resistance: Tackling a Crisis for the Health and Wealth of Nations. UK Government Report," 2014.

2. M. Kumar *et al.*, "Futuristic Non-antibiotic Therapies to Combat Antibiotic Resistance: A Review," *Frontiers in Microbiology*, vol. 12, no. 609459, 2021, doi: 10.3389/FMICB.2021.609459.

3. E. Selimoglu, "Aminoglycoside-Induced Ototoxicity," *Current Pharmaceutical Design*, vol. 13, no. 1, pp. 119–126, 2007, doi: 10.2174/138161207779313731.

4. S. J. McWilliam, D. J. Antoine, R. L. Smyth, and M. Pirmohamed, "Aminoglycoside-induced nephrotoxicity in children.," *Pediatric Nephrology*, vol. 32, no. 11, pp. 2015–2025, 2017, doi: 10.1007/s00467-016-3533-z.

5. J. M. Lopez-Novoa, Y. Quiros, L. Vicente, A. I. Morales, and F. J. Lopez-Hernandez, "New insights into the mechanism of aminoglycoside nephrotoxicity: An integrative point of view," *Kidney International*, vol. 79, no. 1. pp. 33–45, 2011. doi: 10.1038/ki.2010.337.

6. S. J. Kohlhepp, M. O. Loveless, P. W. Kohnen, D. C. Houghton, W. M. Bennett, and D. N. Gilbert, "Nephrotoxicity of the constituents of the gentamicin complex.," *The Journal of Infectious Diseases*, vol. 149, no. 4, pp. 605–14, 1984, doi: 10.1093/infdis/149.4.605.

7. J. A. Caminero, G. Sotgiu, A. Zumla, and G. B. Migliori, "Best drug treatment for multidrug-resistant and extensively drug-resistant tuberculosis," *The Lancet Infectious Diseases*, vol. 10, no. 9, pp. 621–629, 2010, doi: 10.1016/S1473-3099(10)70139-0.

8. L. A. Grohskopf, W. C. Huskins, R. L. Sinkowitz-Cochran, G. L. Levine, D. A. Goldmann, and W. R. Jarvis, "Use of antimicrobial agents in United States neonatal and pediatric intensive care patients," *Pediatric Infectious Disease Journal*, vol. 24, no. 9, pp. 766–773, 2005, doi: 10.1097/01.INF.0000178064.55193.1C.

9. G. M. Pacifici, "Clinical pharmacokinetics of aminoglycosides in the neonate: A review," *European Journal of Clinical Pharmacology*, vol. 65, no. 4, pp. 419–427, 2009, doi: 10.1007/S00228-008-0599-Y/

10. G. M. Musiime, A. C. Seale, S. G. Moxon, and J. E. Lawn, "Risk of gentamicin toxicity in neonates treated for possible severe bacterial infection in low- and middle-income countries: Systematic Review," *Tropical Medicine & International Health*, vol. 20, no. 12, pp. 1593–1606, 2015, doi: 10.1111/TMI.12608.

11. E. J. Onyango *et al.*, "Evaluation of the clinical practice of aminoglycoside use in paediatric patients in Kenya: findings and implications for lower-middle income countries," *JAC-Antimicrobial Resistance*, vol. 2, no. 1, 2020, doi: 10.1093/JACAMR/DLZ087.

12. J. Guo *et al.*, "Specificity and Promiscuity at the Branch Point in Gentamicin Biosynthesis," *Chemistry & Biology*, vol. 21, no. 5, pp. 608–618, 2014, doi: 10.1016/j.chembiol.2014.03.005.

13. A. F. Vydrin, I. v Shikhaleev, V. L. Makhortov, N. N. Shcherenko, and N. v Kolchanova, "Component Composition of Gentamicin Sulfate Preparations," *Pharmaceutical Chemistry Journal*, vol. 37, no. 8, pp. 52–54, 2003.

14. L. O. White, A. Lovering, and D. S. Reeves, "Variations in Gentamicin C1, C1a, C2, and C2a Content of Some Preparations of Gentamicin Sulphate Used Clinically as Determined by High-Performance Liquid Chromatography," *Therapeutic Drug Monitoring*, vol. 5, no. 1, pp. 123–126, 1983, doi: 10.1097/00007691-198303000-00014.

15. M. Ishikawa *et al.*, "Lower ototoxicity and absence of hidden hearing loss point to gentamicin C1a and apramycin as promising antibiotics for clinical use," *Scientific Reports*, vol. 9, no. 1, p. 2410, 2019, doi: 10.1038/s41598-019-38634-3.

16. M. E. O'Sullivan *et al.*, "Dissociating antibacterial from ototoxic effects of gentamicin C-subtypes," *PNAS*, vol. 117, no. 51, pp. 32423–32432, Dec. 2020, doi: 10.1073/pnas.2013065117.

17. M. Kobayashi, M. Sone, M. Umemura, T. Nabeshima, T. Nakashima, and S. Hellström, "Comparisons of cochleotoxicity among three gentamicin compounds following intratympanic application," *Acta Oto-Laryngologica*, vol. 128, no. 3, pp. 245–249, 2008, doi: 10.1080/00016480701558948.

18. H.-J. Nah, H.-R. Pyeon, S.-H. Kang, S.-S. Choi, and E.-S. Kim, "Cloning and Heterologous Expression of a Large-sized Natural Product Biosynthetic Gene Cluster in *Streptomyces* Species," *Frontiers in Microbiology*, vol. 8, no. MAR, p. 394, 2017, doi: 10.3389/fmicb.2017.00394.

19. C. Song *et al.*, "Enhanced Heterologous Spinosad Production from a 79-kb Synthetic Multioperon Assembly," *ACS Synthetic Biology*, vol. 8, no. 1, pp. 137–147, 2019, doi: 10.1021/ACSSYNBIO.8B00402/

20. F. Yan *et al.*, "Synthetic biology approaches and combinatorial biosynthesis towards heterologous lipopeptide production," *Chem Sci*, vol. 9, no. 38, pp. 7510–7519, 2018, doi: 10.1039/C8SC02046A.

21. M. M. Zhang *et al.*, "CRISPR-Cas9 strategy for activation of silent *Streptomyces* biosynthetic gene clusters," *Nature Chemical Biology*, vol. 13, no. 6, pp. 607–609, 2017, doi: 10.1038/NCHEMBIO.2341.

22. K. D. Bauman *et al.*, "Refactoring the Cryptic Streptophenazine Biosynthetic Gene Cluster Unites Phenazine, Polyketide, and Nonribosomal Peptide Biochemistry," *Cell Chemical Biology*, vol. 26, no. 5, pp. 724-736.e7, 2019, doi: 10.1016/j.chembiol.2019.02.004.

23. C. H. Ji, H. Kim, H. W. Je, H. Kwon, D. Lee, and H. S. Kang, "Top-down synthetic biology approach for titer improvement of clinically important antibiotic daptomycin in *Streptomyces roseosporus*," *Metabolic Engineering*, vol. 69, pp. 40–49, 2022, doi: 10.1016/J.YMBEN.2021.10.013.

24. J. Xu, J. Zhang, J. Zhuo, Y. Li, Y. Tian, and H. Tan, "Activation and mechanism of a cryptic oviedomycin gene cluster via the disruption of a global regulatory gene, adpA, in *Streptomyces ansochromogenes*.," *The Journal of Biological Chemistry*, vol. 292, no. 48, pp. 19708–19720, 2017, doi: 10.1074/jbc.M117.809145.

25. C. Olano *et al.*, "Activation and identification of five clusters for secondary metabolites in *Streptomyces albus* J1074.," *Microbial Biotechnology*, vol. 7, no. 3, pp. 242–56, 2014, doi: 10.1111/1751-7915.12116.

26. J. W. Park *et al.*, "Genetic dissection of the biosynthetic route to gentamicin A2 by heterologous expression of its minimal gene set.," *PNAS*, vol. 105, no. 24, pp. 8399–404, 2008, doi: 10.1073/pnas.0803164105.

27. Y. Sun, X. Jian, and Z. Deng, "Preparation and applications of a series of novel aminoglycoside compounds, and construction method of high-yield engineering strain for producing novel aminoglycoside compounds," Chinese Patent CN106432382B, Apr. 01, 2019

28. S. Li *et al.*, "Methyltransferases of gentamicin biosynthesis," *PNAS,* vol. 115, no. 6, pp. 1340–1345, 2018, doi: 10.1073/pnas.1711603115.

29. S. Li *et al.*, "Mechanistic Insights into Dideoxygenation in Gentamicin Biosynthesis," *ACS Catalysis*, vol. 11, no. 19, pp. 12274–12283, 2021, doi: doi.org/10.1021/acscatal.1c03508.

30. K. V. Baker *et al.*, "Chapter 2 - Establishing *M.* sp. DEM32671 as a producer of aminoglycosides through heterologous expression in Streptomyces spp.," PhD Thesis Chapter, University of Manchester, Manchester, 2022.

31. S. Ye, B. Enghiad, H. Zhao, and E. Takano, "Fine-tuning the regulation of Cas9 expression levels for efficient CRISPR-Cas9 mediated recombination in

Streptomyces," *Journal of Industrial Microbiology & Biotechnology*, vol. 47, no. 4–5, pp. 413–423, 2020, doi: 10.1007/s10295-020-02277-5.

32. Y. H. Ban *et al.*, "Complete reconstitution of the diverse pathways of gentamicin B biosynthesis," *Nature Chemical Biology*, vol. 15, pp. 295–303, 2019, doi: 10.1038/s41589-018-0203-4.

33. Z. Wei, X. Shi, R. Lian, W. Wang, W. Hong, and S. Guo, "Exclusive production of gentamicin C1a from *micromonospora purpurea* by metabolic engineering," *Antibiotics*, vol. 8, no. 4, 2019, doi: 10.3390/antibiotics8040267.

34. Y. Chang *et al.*, "Overproduction of gentamicin B in industrial strain *Micromonospora echinospora* CCTCC M 2018898 by cloning of the missing genes genR and genS," *Metabolic Engineering Communications*, 2019, doi: 10.1016/j.mec.2019.e00096.

35. A. A. Vu *et al.*, "Integrity and regeneration of mechanotransduction machinery regulate aminoglycoside entry and sensory cell death," *PLoS One*, vol. 8, no. 1, 2013, doi: 10.1371/JOURNAL.PONE.0054794.

36. A. Alharazneh *et al.*, "Functional hair cell mechanotransducer channels are required for aminoglycoside ototoxicity," *PLoS One*, vol. 6, no. 7, 2011, doi: 10.1371/JOURNAL.PONE.0022347.

37. A. Rodríguez-García, P. Combes, R. Pérez-Redondo, M. C. A. Smith, and M. C. M. Smith, "Natural and synthetic tetracycline-inducible promoters for use in the antibiotic-producing bacteria *Streptomyces*," *Nucleic Acids Research*, vol. 33, no. 9, pp. e87–e87, 2005, doi: 10.1093/NAR/GNI086.

38. C. Howden, "Generating New Antibiotics by Cloning Antibiotic Biosynthetic Gene Clusters," Thesis (MSc), University of Newcastle, Newcastle, 2017.

39. R. Lutz and H. Bujard, "Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements.," *Nucleic Acids Research*, vol. 25, no. 6, p. 1203, 1997, doi: 10.1093/NAR/25.6.1203.

40. D. D'Alia, "Regulation of antibiotic production by ncRNA and signaling molecules in *Streptomyces coelicolor*," PhD thesis, University of Groningen, Groningen, 2011.

41. M. Myronovskyi and A. Luzhetskyy, "Native and engineered promoters in natural product discovery," *Natural Product Reports*, vol. 33, no. 8, pp. 1006–19, 2016, doi: 10.1039/c6np00002a.

42. T. Kieser, M. J. Bibb, M. J. Buttner, K. F. Chater, and D. A. Hopwood, "Practical *Streptomyces* Genetics," *John Innes Foundation*, p. 529, 2000. doi: 10.4016/28481.01.

43. L. Horbal, T. Siegl, and A. Luzhetskyy, "A set of synthetic versatile genetic control elements for the efficient expression of genes in Actinobacteria," *Scientific Reports*, vol. 8, p. 491, 2018, doi: 10.1038/s41598-017-18846-1.

44. J. P. Torella, F. Lienert, C. R. Boehm, J. H. Chen, J. C. Way, and P. A. Silver, "Unique nucleotide sequence-guided assembly of repetitive DNA parts for synthetic biology applications," *Nature Protocols*, vol. 9, no. 9, pp. 2075–2089, 2014, doi: 10.1038/nprot.2014.145.

45. J. P. Torella, C. R. Boehm, F. Lienert, J. H. Chen, J. C. Way, and P. A. Silver, "Rapid construction of insulated genetic circuits via synthetic sequence-guided isothermal assembly," *Nucleic Acids Research*, vol. 42, no. 1, pp. 681–689, 2014, doi: 10.1093/NAR/GKT860.

46. J. P. Gomez-Escribano and M. J. Bibb, "Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters," *Microbial Biotechnology*, vol. 4, no. 2, pp. 207–215, 2011, doi: 10.1111/j.1751-7915.2010.00219.x.

47. R. Bertram and W. Hillen, "The application of Tet repressor in prokaryotic gene regulation and expression," *Microbial Biotechnology*, vol. 1, no. 1, p. 2, 2008, doi: 10.1111/J.1751-7915.2007.00001.X.

48. H. H. Huang, C. Seeger, U. Helena Danielson, and P. Lindblad, "Analysis of the leakage of gene repression by an artificial TetR-regulated promoter in cyanobacteria," *BMC Res Notes*, vol. 8, no. 1, Sep. 2015, doi: 10.1186/S13104-015-1425-0.

49. Y-.F. Wan, R. Lian, W-.R. Hong, and X-.A. Shi, "Study on the function of resistant genes *gmrB* and *gmrA* in gentamicin biosynthetic gene cluster," *Chinese Journal of Antibiotics*, vol. 43, no. 6, pp. 688–695, 2018.

50. V. Solovyev and A. Salamov, "Automatic annotation of microbial genomes and metagenomic sequences. In metagenomics and its applications in agriculture, biomedicine and environmental studies (Ed. R.W. Li).," *Nova Science Publishers*, 2011.

51. C. Bai *et al.*, "Exploiting a precise design of universal synthetic modular regulatory elements to unlock the microbial natural products in Streptomyces," *PNAS*, vol. 112, no. 39, pp. 12181–12186, 2015, doi: 10.1073/PNAS.1511027112.

52. *Dictionary of Natural Products*. Taylor & Francis Group, 2000.

53. R. Ishii, M. Horie, W. Chan, and J. MacNeil, "Multi-residue quantitation of aminoglycoside antibiotics in kidney and meat by liquid chromatography with tandem mass spectrometry," *Food Additives & Contaminants: Part A: Chemistry, Analysis, Control, Exposure & Risk Assessment*, vol. 25, no. 12, pp. 1509–1519, 2008, doi: 10.1080/02652030802189740.

54. P. Kumar, A. Rubies, R. Companyó, and F. Centrich, "Hydrophilic interaction chromatography for the analysis of aminoglycosides," *Journal of Separation Science*, vol. 35, no. 4, pp. 498–504, 2012, doi: 10.1002/JSSC.201100860.

55. C. Chiaochan, U. Koesukwiwat, S. Yudthavorasit, and N. Leepipatpiboon, "Efficient hydrophilic interaction liquid chromatography–tandem mass spectrometry for the multiclass analysis of veterinary drugs in chicken muscle," *Analytica Chimica Acta*, vol. 682, no. 1–2, pp. 117–129, 2010, doi: 10.1016/J.ACA.2010.09.048.

56. E. Kaale, C. Govaerts, J. Hoogmartens, and A. van Schepdael, "Mass spectrometric study to characterize thioisoindole derivatives of aminoglycoside antibiotics," *Rapid Commun Mass Spectrom*, vol. 19, no. 20, pp. 2918–2922, 2005, doi: 10.1002/RCM.2152.

57. R. Sardella *et al.*, "Optimized Extraction of Amikacin from Murine Whole Blood," *Molecules*, vol. 26, no. 3, 2021, doi: 10.3390/MOLECULES26030665.

58. S. B. Turnipseed, S. B. Clark, C. M. Karbiwnyk, W. C. Andersen, K. E. Miller, and M. R. Madson, "Analysis of aminoglycoside residues in bovine milk by liquid chromatography electrospray ion trap mass spectrometry after derivatization with phenyl isocyanate," *J Chromatogr B Analyt Technol Biomed Life Sci*, vol. 877, no. 14–15, pp. 1487–1493, May 2009, doi: 10.1016/J.JCHROMB.2009.03.025.

59. F. Farouk, H. M. E. Azzazy, and W. M. A. Niessen, "Challenges in the determination of aminoglycoside antibiotics, a review," *Anal Chim Acta*, vol. 890, pp. 21–43, 2015, doi: 10.1016/J.ACA.2015.06.038.

60. D. Montiel, H.-S. Kang, F.-Y. Chang, Z. Charlop-Powers, and S. F. Brady, "Yeast homologous recombination-based promoter engineering for the activation of silent natural product biosynthetic gene clusters", PNAS, vol. 112, no. 29, pp. 8953-8, 2015, doi: 10.1073/pnas.1507606112.

61. T. H. Eyles, N. M. Vior, and A. W. Truman, "Rapid and Robust Yeast-Mediated Pathway Refactoring Generates Multiple New Bottromycin-Related

Metabolites," ACS Synthetic Biology, vol. 7, no. 5, pp. 1211-1218, 2018, doi: 10.1021/acssynbio.8b00038.

62. M. R. Estévez *et al.*, "Novel Fredericamycin Variant Overproduced by a Streptomycin-Resistant *Streptomyces albus* subsp. *chlorinus* Strain," *Marine Drugs*, vol. 18, no. 6, 2020, doi: 10.3390/MD18060284.

63. B. W. Biggs *et al.*, "Overcoming heterologous protein interdependency to optimize P450-mediated Taxol precursor synthesis in *Escherichia coli*," *PNAS*, vol. 113, no. 12, pp. 3209–3214, 2016, doi: 10.1073/PNAS.1515826113.

64. L. Horbal, V. Fedorenko, and A. Luzhetskyy, "Novel and tightly regulated resorcinol and cumate-inducible expression systems for *Streptomyces* and other actinobacteria," *Applied Microbiology and Biotechnology*, vol. 98, no. 20, pp. 8641–8655, 2014, doi: 10.1007/S00253-014-5918-X.

65. C. H. Ji, H. Kim, and H. S. Kang, "Synthetic Inducible Regulatory Systems Optimized for the Modulation of Secondary Metabolite Production in *Streptomyces*," *ACS Synthetic Biology*, vol. 8, no. 3, pp. 577–586, 2019, doi: 10.1021/ACSSYNBIO.9B00001.

66. J. H. Lim *et al.*, "Enhanced biosynthesis of 2-deoxy-scyllo-inososе in metabolically engineered *bacillus subtilis* recombinants," *Frontiers in Microbiology*, vol. 9, 2018, doi: 10.3389/fmicb.2018.02333.

67. X. Wang, Z. Li, L. Policarpio, M. A. G. Koffas, and H. Zhang, "De novo biosynthesis of complex natural product sakuranetin using modular co-culture engineering," *Applied Microbiology and Biotechnology*, vol. 104, no. 11, pp. 4849–4861, Jun. 2020, doi: 10.1007/S00253-020-10576-1.

68. X. Wang, A. Shao, Z. Li, L. Policarpio, and H. Zhang, "Constructing *E. coli* Co-Cultures for De Novo Biosynthesis of Natural Product Acacetin," *Biotechnology Journal*, vol. 15, no. 9, p. 2000131, Sep. 2020, doi: 10.1002/BIOT.202000131.

69. J. A. Jones *et al.*, "Complete Biosynthesis of Anthocyanins Using *E. coli* Polycultures," *mBio*, vol. 8, no. 3, pp. e00621-17, Jul. 2017, doi: 10.1128/mBio.00621-17.

70. B. P. Anton and E. A. Raleigh, "Complete Genome Sequence of NEB 5-alpha, a Derivative of *Escherichia coli* K-12 DH5α.," *Genome Announc*, vol. 4, no. 6, Nov. 2016, doi: 10.1128/genomeA.01245-16.

71. D. J. MacNeil, K. M. Gewain, C. L. Ruby, G. Dezeny, P. H. Gibbons, and T. MacNeil, "Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector," *Gene*, vol. 111, no. 1, pp. 61–68, 1992, doi: 10.1016/0378-1119(92)90603-M.

72. P. R. Burkholder and N. H. Giles, “Induced biochemical mutations in *Bacillus subtilis*.,” *Am J Bot*, vol. 34, no. 6, pp. 345–8, 1947.

73. F. Kunst *et al.*, “The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*.,” *Nature*, vol. 390, no. 6657, pp. 249–256, 1997, doi: 10.1038/36786.

74. O. Hara and T. Beppu, “Mutants blocked in streptomycin production in *Streptomyces griseus* - the role of A-factor.,” *J Antibiot (Tokyo)*, vol. 35, no. 3, pp. 349–58, 1982.

75. J. Sambrook and D. W. Russell, *Molecular cloning: a laboratory manual.* Cold Spring Harbour Laboratory, 2001.

76. R. A. Scheltema, A. Jankevics, R. C. Jansen, M. A. Swertz, and R. Breitling, “PeakML/mzMatch: A File Format, Java Library, R Library, and Tool-Chain for Mass Spectrometry Data Analysis,” *Analytical Chemistry*, vol. 83, no. 7, pp. 2786–2793, 2011, doi: 10.1021/AC2000994.

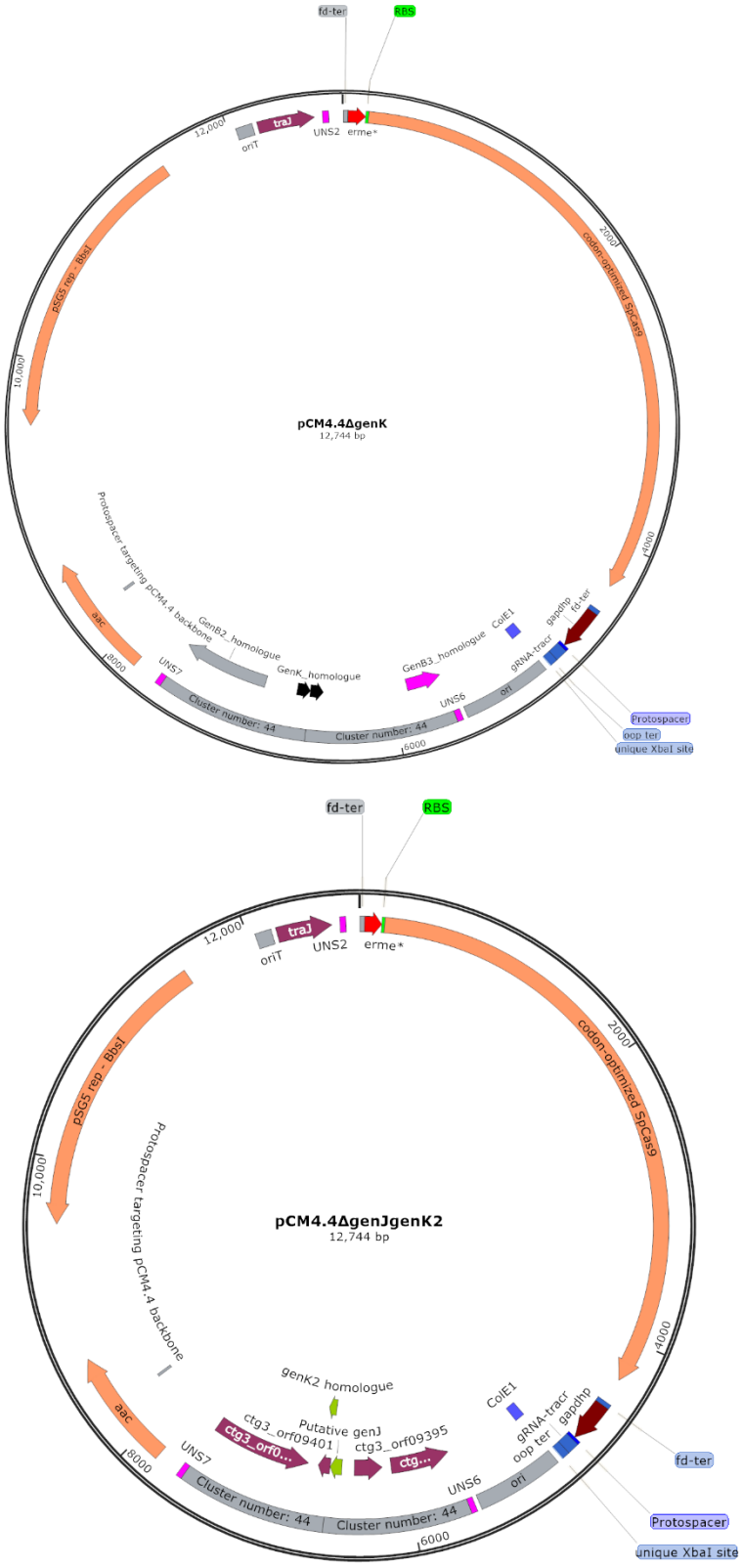


Figure S3.1: Plasmid maps for pCM4.4ΔgenK and pCM4.4ΔgenJgenK2.

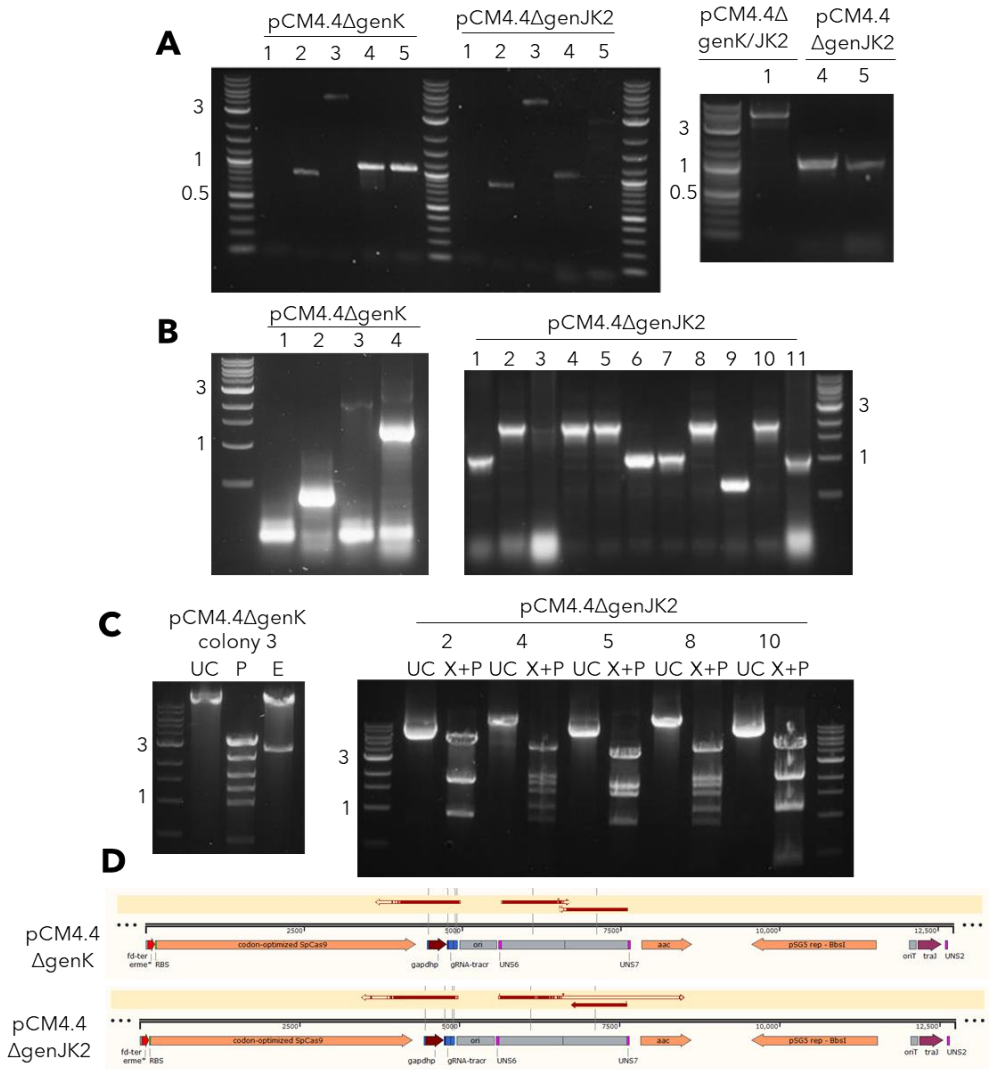


Figure S3.2: Assembly of pCM4.4ΔgenK, pCM4.4ΔgenJgenK2. A – Agarose gel of PCR amplification of constituent parts for HiFi assembly. Expected bands sizes = 1 – 5042 bp, 2 – 853 bp, 3 – 4910 bp, 4 – 1060 bp/1058 bp (pCM4.4ΔgenK/pCM4.4ΔgenJgenK2 respectively), 5 – 1059 bp. Ladder = NEB 1Kb Plus **B –** Colony PCR of pCM4.4ΔgenK, pCM4.4ΔgenJgenK2 using UNS6F, UNS7R primers. pCM4.4ΔgenK colony 3 showed positive result, pCM4.4ΔgenJgenK2 2, 4, 5, 8, 10 showed positive result and were taken forward for further screening (expected band size for positive result = 2000 bp). **C** – Restriction digest screening of pCM4.4ΔgenK, pCM4.4ΔgenJgenK2 screened with either PstI/EcoRI (*ΔgenK*) or XhoI + PstI (*ΔgenJ ΔgenK2*). UC denotes uncut plasmid. pCM4.4ΔgenK colony 3, pCM4.4ΔgenJgenK2 14 and 18 were positives (expected band sizes pCM4.4ΔgenK = EcoRI digest: 10,216 bp, 2528 bp. PstI digest: 3355 bp, 3078 bp, 2202 bp, 1558 bp, 1212 bp, 920 bp, 419 bp. Expected band sizes pCM4.4ΔgenJgenK2 = XhoI + PstI digest: 4570 bp, 2149 bp, 1863 bp, 1558 bp, 1166 bp, 920 bp, 419 bp. **D** – positive sequencing for pCM4.4ΔgenK, pCM4.4ΔgenJgenK2.

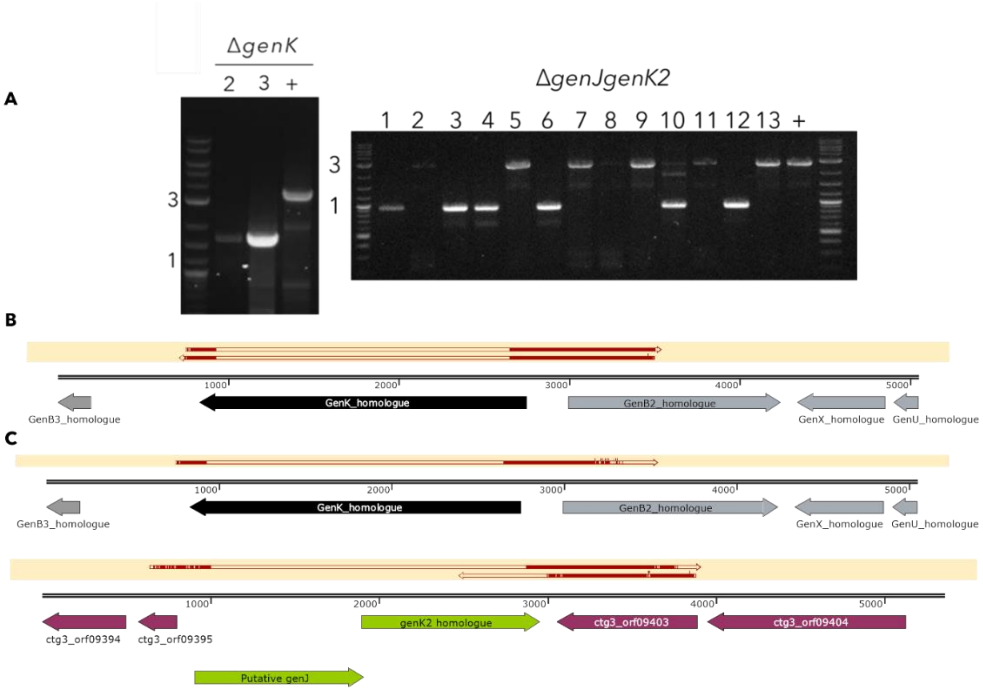


Figure S3.3: Screening of *S. coelicolor* M1152 C24 /pCM4.4ΔgenK/pCM4.4ΔgenJK2 exconjugants for modification. A – Agarose gel of first round screening for modification. *ΔgenK* colonies 2, 3; *ΔgenJΔgenK2* colony 1, 3, 4, 6, 10, 12 showed positive for the modification (band sizes of 1640, 1159 bp respectively). **B** – sequencing of *S. coelicolor* M1152 C24 *ΔgenK*(2) with primers genKscreenfw and genKscreenrev. **C** – sequencing of S. coelicolor M1152 C24 *ΔgenK(2)ΔgenJΔgenK2* (derived from colony 1 in Figure S3.3A) for maintenance of *genK*-knockout (top; genKscreenfw primer) and *genJ*, *genK2*-knockout (bottom; genJscreenfw, genJscreenrev primers).

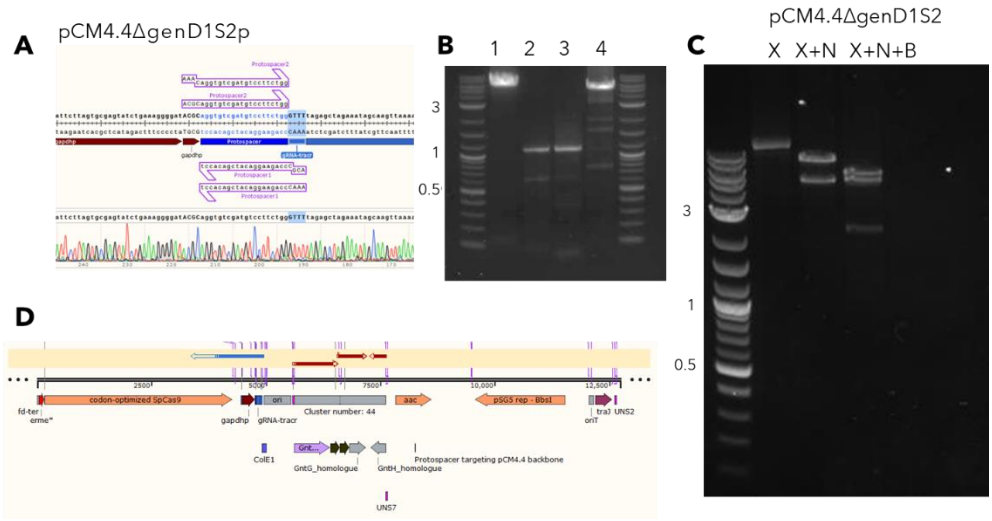


Figure S3.4: Assembly of pCM4.4ΔgenD1genS2. A – Positive sequencing for protospacer insertion. **B –** Agarose gel of PCR amplification of constituent parts for HiFi assembly. Expected band sizes = 1 – 6180 bp, 2 – 1060 bp, 3 – 1060 bp, 4 – 5042 bp. Ladder = NEB 1Kb Plus. **C –** Restriction digest screening of colony obtained from assembly – lane 1 = XbaI single digest – expected band size = 12745 bp; lane 2 = XbaI and NheI digest – expected band sizes = 7832, 4913 bp; lane 3 = XbaI, NheI and BbsI digest – expected band sizes = 5636, 4913, 2196 bp. **D –** Positive sequencing of pCM4.4ΔgenD1genS2 homology arms and protospacer, highlighting correct assembly.

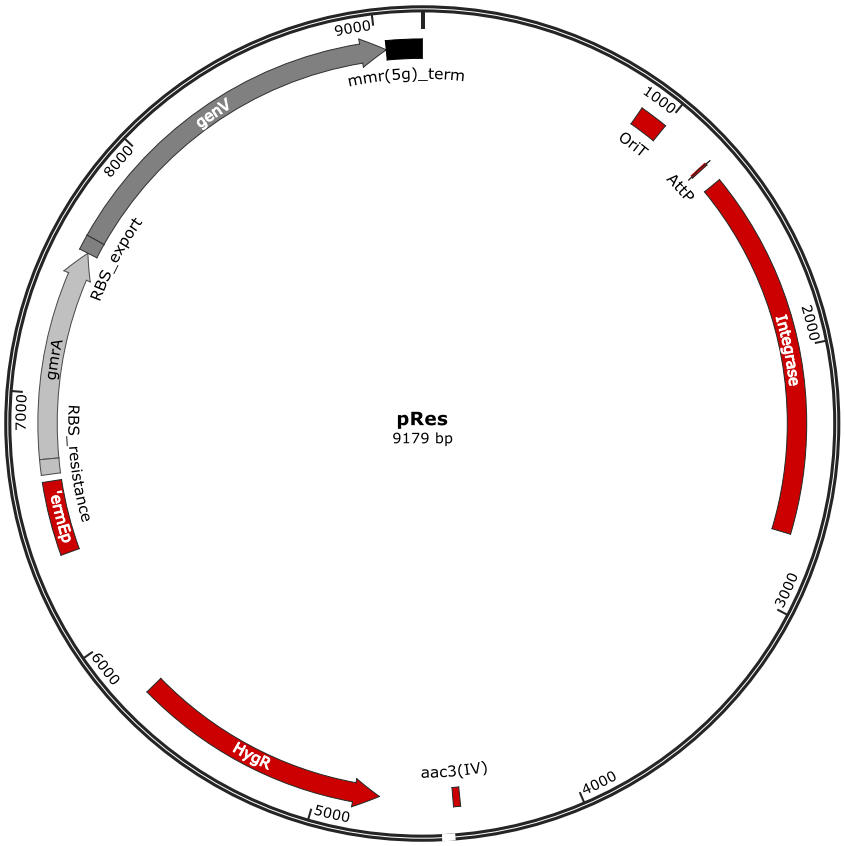


Figure S3.5: Plasmid map of pRes, consisting of *gmrA* and *genV* homologues from Cluster 24 under control of *ermEp*.

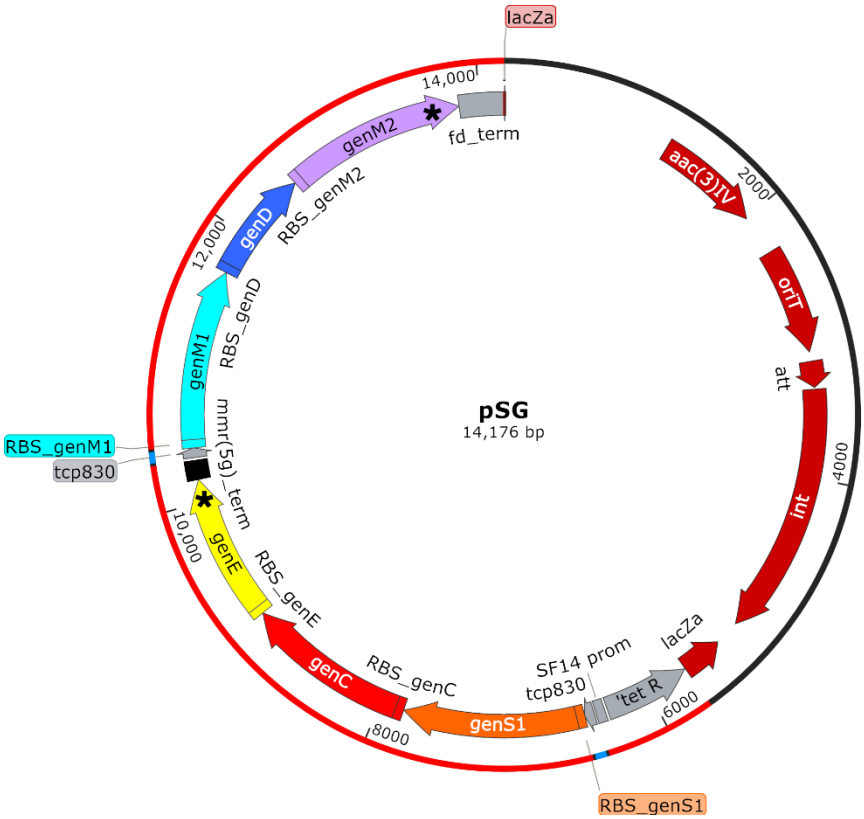


Figure S3.6: Plasmid map of pSG, insert consisting of *tetR-SF14-tcp830-genS1-genC-genE-mmr(5g)-tcp830-genM1-genD-genM2-fd*. Sites where histidine tags are inserted at the 3′ end of *genE* and *genM2* are highlighted with asterisk (*).

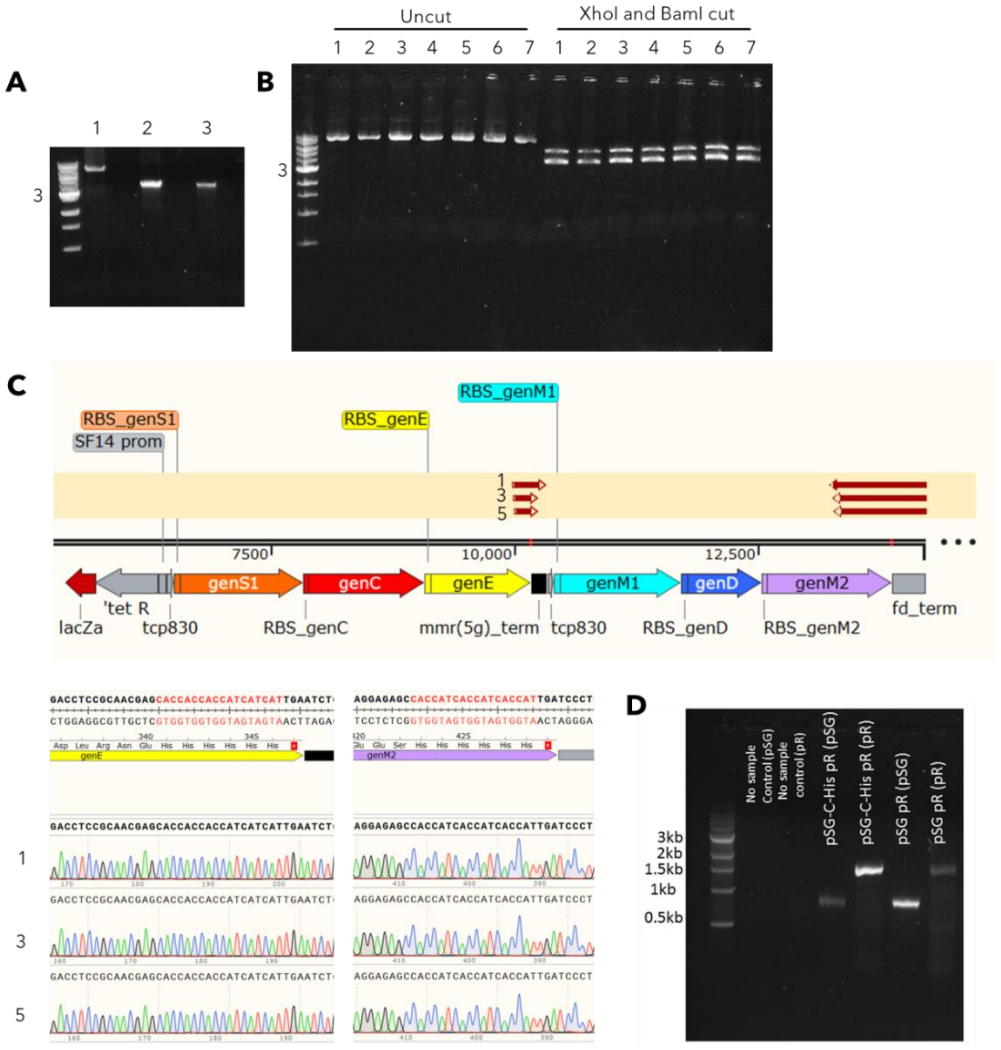


Figure S3.7: Construction and validation of pSGCH. A – PCR reactions for each piece of the plasmid to be assembled via HiFi assembly – expected sizes = 1 – 6.8kb, 2 – 3.8 kb, 3 – 3.8 kb. **B –** Restriction digest of plasmid DNA purified from 7 colonies after assembly. A positive result is indicated by the presence of bands at 6 kb, 3.8 kb, and 750 bp. **C –** Positive sequencing of assembled plasmid with C-terminal His-tags successfully added to *genE* and *genM2*. **D –** Screening for the presence of pSG in *S. lividans* TK23. Positive bands for presence of pSG = 800 bp; positive bands for presence of pRes = 1.5 kb.

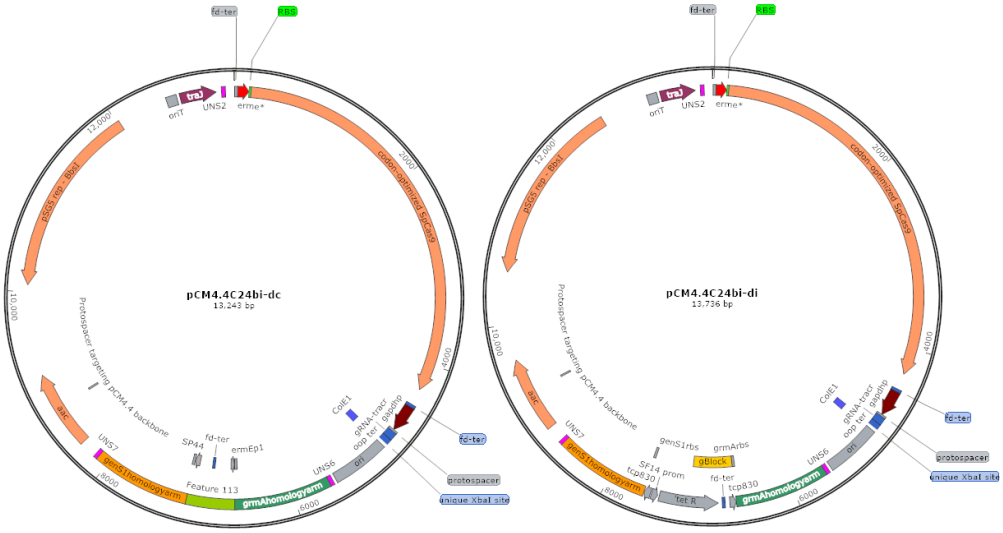


Figure S3.8: Plasmid map for pCM4.4C24bi-dc and pCM4.4C24bi-di.

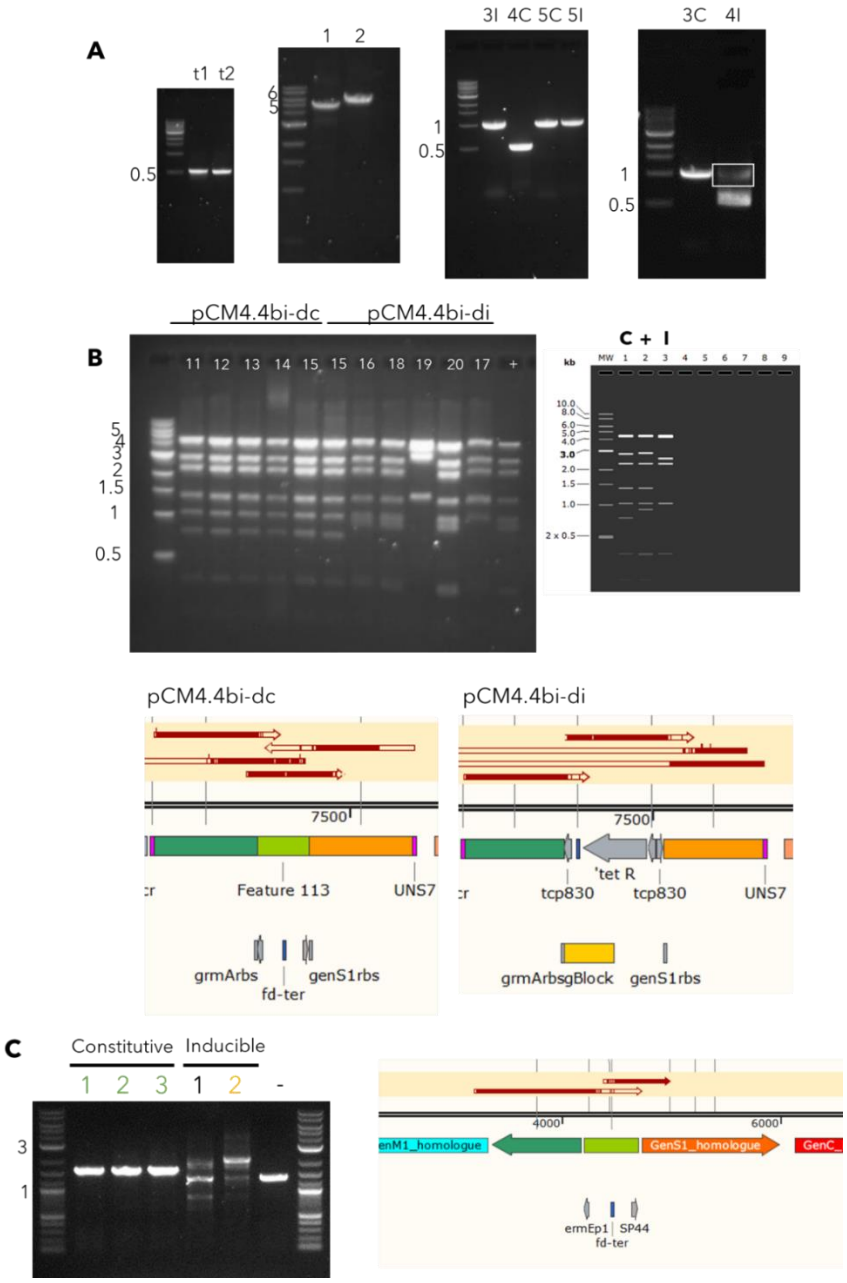


Figure S3.9: Construction of pCM4.4C24bi-dc and pCM4.4C24bi-di, and validation of constitutive promoter cassette insertion to *S. coelicolor* M1146 C24 2R. A – Amplification of constituent fragments for assembly after generation of pCM4.4C24bi-dp. **B –** Validation of assembly. PstI digest of candidate pCM4.4C24bi-dc and pCM4.4C24bi-di assemblies. Expected band sizes for pCM4.4C24bi-dc = 3921 bp, 3011 bp, 2202 bp, 1558 bp, 1212 bp, 920 bp, 419 bp. Expected band sizes for pCM4.4C24bi-di = 3722 bp, 3703 bp, 2202 bp, 1558 bp, 1212 bp, 920 bp, 419 bp. Sequencing for both plasmids was confirmed. **C –** validation of insertion of constitutive promoter cassette to *S. coelicolor* M1146 C24 2R. Colonies were screeened after conjugation with pCM4.4bi-dscreenfw and pCM4.4bi-dscreenrev, a size increase of 260 bp (1586 bp product length) indicated successful assembly. Sequence of the insert region was confirmed by Sanger sequencing.

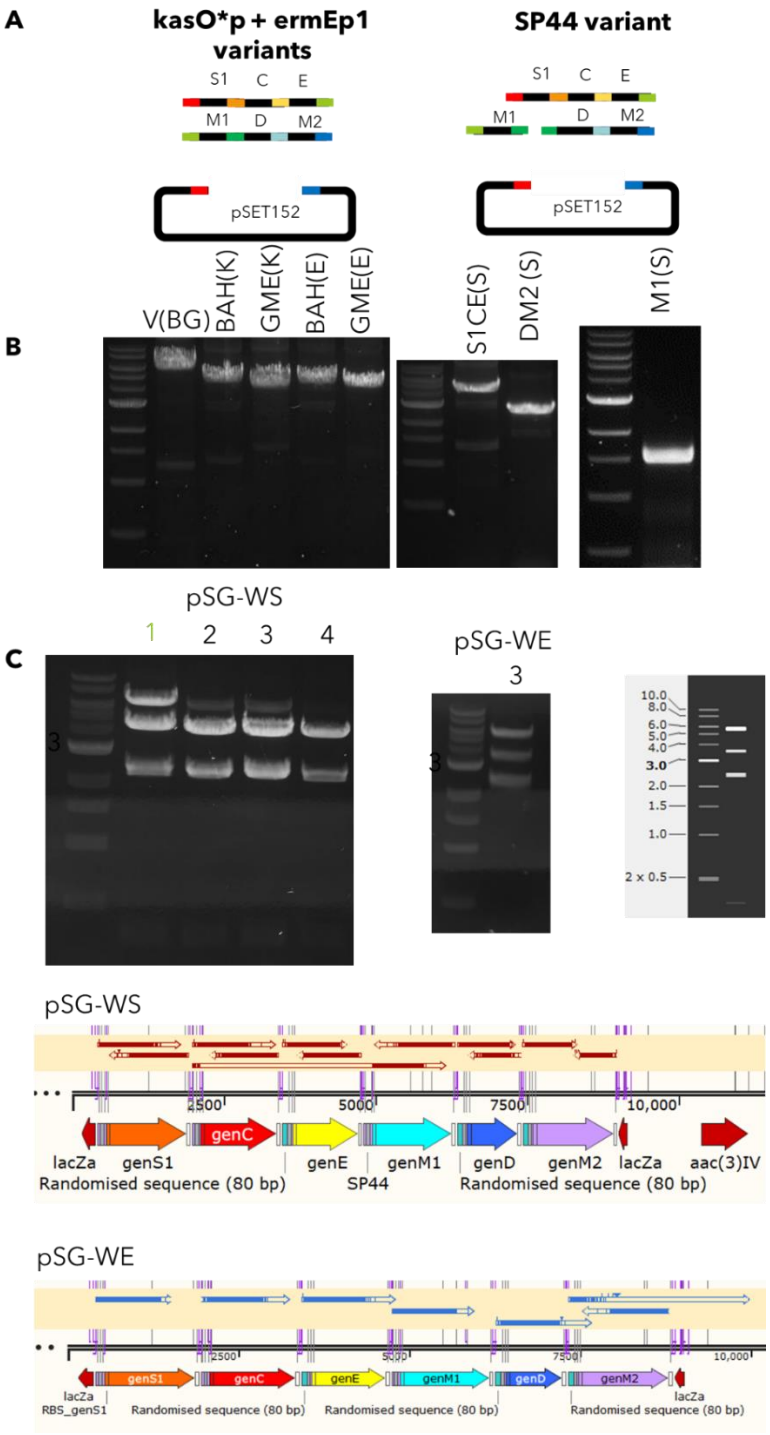


Figure S3.10: Construction and validation of pSG-WS and pSG-WE. A – After first assembly to generate half-plasmids, second round of assembly was done to stitch together fragments as shown **B –** PCR of parts for each full assembly. (S) denotes *sp44* variant, (E) denotes *ermEP1* variant, (K) denotes abandoned *kasO*p* variant. **C –** Validation of correct assembly. NcoI screening of putative correct assemblies. Expected band size = 5765 bp, 3503 bp, 2362 bp, 2308 bp, 318 bp. Correct assemblies were confirmed by Sanger sequencing.

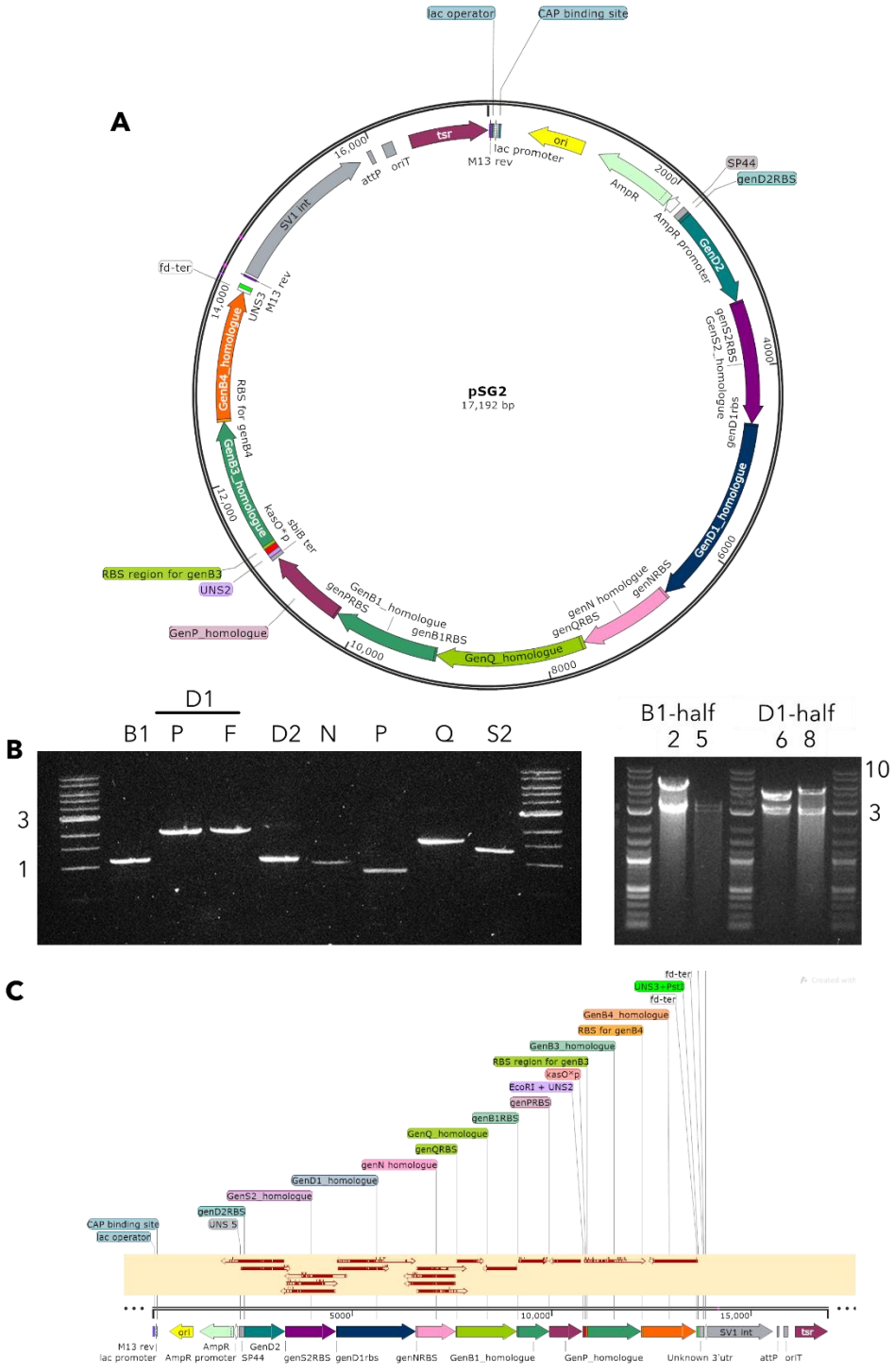


Figure S3.11: Construction and validation of pSG2, consisting of *sp44-genD2-genS2-genD1-genN-genQ-genB1-genP-sbiB-kasO*p-genB3-genB4-fd*. **A** – plasmid map of pSG2. **B** – PCR reaction to generate parts for assemblies, validation of half assemblies (i) *sp44-genD2-genS2-genD1*, ii) *genN-genQ-genB1-genP*) via restriction digest (EcoRI + SpeI; expected sizes B1-half 6844 bp, 3146 bp, D1-half 6690 bp, 3146 bp. **C** – validation of full assembly via Sanger sequencing (primers used were assembly primers for each coding sequence detailed in **Table S3.1.**

Table S3.1: Primers used in this study for construction and validation of plasmid constructs.

pSGCH			
Primer name	**Primer sequence 5′ – 3′**	**Product Length**	**Specific Function**
pSG_C_Htag_GA_1	AGCCACCATCACCATCACCATTGATCCCTGCAA GCCTCAGC	6763 bp	Amplification of pSET152 with 5′ overhang containing 6xHis tag
pSG_C_Htag_GA_2	GCCAACAGCACTGCAGATCT		
pSG_C_Htag_GA_3	GGTCAAGGCGTAGGTCTACG	3775 bp	Amplification of genS1-genC-genE with 3′ overhang containing 6xHistag
pSG_C_Htag_GA_4	AGATTCAATGATGATGGTGGTGGTGCTCGTTGC GGAGGTCGAA		
pSG_C_Htag_GA_5	CACCACCACCATCATCATTGAATCTGGGGGAAC GCCGC	3690 bp	Amplification of genM1-genD-genM2 with 3′ overhang containing 6xHistag
pSG_C_Htag_GA_6	AGGGATCAATGGTGATGGTGATGGTGGCTCTCC TCCATCAGGG		
All pCM4.4-based plasmids			
pCM4.4 HA site UNS7 fw	CAAGACGCTGGCTCTGACATTTCCGCTACTGAA CTACTCGACGCTCAGTGGAACGAAAAC	5743 bp	Amplification of Fragment 1 of pCM4.4 plasmids with protospacer pre-inserted
dLancI-II 1 rev	GCTTGGATTCTGCGTTTGTTTCCGTCTACGAAC TCCCAGCGGGACGTGCTTGGCAATCA		
Cas9 UNS2 fw	GCTGGGAGTTCGTAGACGGAAACAAACGCAGAA TCCAAGCCATGCGCTCCATCAAGAA	5041 bp	Amplification of Fragment 2 of pCM4.4 plasmids with protospacer pre-inserted
pCM4.4 HA site UNS6 rev	GTATGTGACCGTAGAGTATTCTTAGGTGGCAGC GAACGAGCAGACCCCGTAGAAAAGA		
pCM4.4protospacers eqrev	GCGTCGATTTTTGTGATGCT	-	Sequencing of pCM4.4-based plasmids
pCM4.4repairfwd	TCCTTTGATCTTTTCTACGGGG		
pCM4.4repairrev	GTTTTCGTTCCACTGAGCGT		

UNS6F	CTCGTTCGCTGCCACCTAAGAATACTCTACGGT CACATAC	2120 bp	Colony PCR of pCM4.4-based plasmid assemblies
UNS7R	CGAGTAGTTCAGTAGCGGAAATGTCAGAGCCAG CGTCTT		
pCM4.4C24bi-dc and pCM4.4C24bi-di			
pCM4.4bi-dconstprotfw	ACGCGCAATAACTTTAAGTACACG	-	Annealed oligos for protospacer insertion
pCM4.4bi-dconstprotrev	AAACCGTGTACTTAAAGTTATTGC		
gmrAUNS6fw	CTCGTTCGCTGCCACCTAAGAATACTCTACGGT CACATACTGGCCAGGCTGCGGGACAACGCC	1060 bp	Amplification of Fragment 3 for assembly of pCM4.4C24bi-dc
grmAconstrev	TTGGTAGGATCCAGCGGGTAGATCTTTCGGAGG ATTCGATGACGACATCTGTG		
ESconstbidampfw	TACCCGCTGGATCCTACCAACCGGC	499 bp	Amplification of Fragment 4 for assembly of pCM4.4C24bi-dc
ESconstbidamprev	TACCTACACCAGACTTTACAACACCG		
genS1constfw	GTAAAGTCTGGTGTAGGTAACGGTTGCCGCAGT CTGGCTT	1060 bp	Amplification of Fragment 5 for assembly of pCM4.4C24bi-dc
genS1UNS7rev	CGAGTAGTTCAGTAGCGGAAATGTCAGAGCCAG CGTCTTCGATCTCGTCGTCTTCGATCCGACCGG		
tetblock1fw	CGTCATCGAATCCTCCGAAAGATCCGACGTACG CCCAATATCTCTATCACTGATAGG	546 bp	Amplification of ½ Fragment 4 for pCM4.4C24bi-di (nested primer PCR with tetblockflankfw and tetblockflankrev)
tetblock1rev	CGGCACCCGCCCAACAGAGAAACAGTACGAAAC CCTGGAAAATCAGC		
Tetblock2fw	GCTGATTTTCCAGGGTTTCGTACTGTTTCTCTG TTGGGCGGGTGCCG	538 bp	Amplification of ½ Fragment 4 for pCM4.4C24bi-di (nested primer PCR with tetblockflankfw and tetblockflankrev)
Tetblock2rev	AAGCCAGACTGCGGCAACCGTCGACGTACGCCC AATATCTCTATCACTGATAGG		
tetblockflankfw	CGTCATCGAATCCTCCGAAAGATC	1037 bp	Amplification of Fragment 4 from t1 and t2 (pCM4.4C24bi-di)
tetblockflankrev	AAGCCAGACTGCGGCAACCGT		

grmAtetrev	AGTGATAGAGATATTGGGCGTACGTCGGATCTTTCGGAGGATTCGATGACGACATCTGTG	1060 bp	Amplification of Fragment 3 (used with gmrAUNS6fw; pCM4.4C24bi-di)
genS1tetfw	AGATATTGGGCGTACGTCGACGGTTGCCGCAGTCTGGCTT	1060 bp	Amplification of Fragment 5 (used with genS1UNS7rev; pCM4.4C24bi-di)
pCM4.4bi-d_screenfw	GGAAAGCCCAGGGTCAGCA	1586 bp (pCM4.4C24bi-dc), 2079 bp (pCM4.4C24bi-di)	Screening for successful insertion of promoter cassette in pCM4.4C24bi-dc and pCM4.4C24bi-di
pCM4.4bi-d_screenrev	AGAAGGGTCACTTTCACGTCCTC		
pCM4.4ΔgenK			
protospacer OLrevgenK	CCAAGGATCACCTTCGCGTTGCGTATCCCCTTTCAGATACT	4910 bp	To be used with Cas9 UNS2 fw for generation of Fragment 2
protospacer OLfwgenK	AACGCGAAGGTGATCCTTGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAG	853 bp	To be used with pCM4.4 HA site UNS6 rev for generation of Fragment 3
genKfrag1fw	CTCGTTCGCTGCCACCTAAGAATACTCTACGGTCACATACCAACGTCGACGCGAACGC	1060 bp	Amplification of Fragment 4
genKfrag1rev	GATGCATGGTGGAAATCGGAGACGAGCGGATCGGCCGG		
genKfrag2fw	ATCCGGCCGATCCGCTCGTCTCCGATTTCCACCATGCATCC	1059 bp	Amplification of Fragment 5
genKfrag2rev	CGAGTAGTTCAGTAGCGGAAATGTCAGAGCCAGCGTCTTGACGCTCGCCTCGGTGACC		
genKscreenfw	GACGAAGCGGCGCTGGATC	1640 bp	Validation of genK knockout Streptomyces strains
genKscreenrev	TGCCGTAGGGGTCAGAGCTG		
pCM4.4ΔgenJgenK2			
protospacer OLfwgenJK2	CCGTGCTCATTGGATCGCGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAG	4910 bp	To be used with pCM4.4 HA site UNS6 rev for generation of Fragment 2

protospacer OLrevgenJK2	TCGCGATCCAATGAGCACGGGCGTATCCCCTTTCAGATACT	853 bp	To be used with Cas9 UNS2 fw for generation of Fragment 3
genJK2frag1fw	CTCGTTCGCTGCCACCTAAGAATACTCTACGGTCACATACGCCGGCTCGTGGAGACGC	1058 bp	Amplification of Fragment 4
genJK2frag1rev	GGTCTCGCCCTGGTCGACGACTGTCGGGGTTGTCGGTG		
genJK2frag2fw	CACCGACAACCCCGACAGTCGTCGACCAGGGCGAGACCGG	1059 bp	Amplification of Fragment 5
genJK2frag2rev	CGAGTAGTTCAGTAGCGGAAATGTCAGAGCCAGCGTCTTGGGCGCCCGAGGGCGT		
genJscreenfw	GCCCCGGTCGATCCTGCC	1121 bp	Validation of genJ, genK2 knockout Streptomyces strains
genJscreenrev	GGCACAGACCCACAGGAGGC		
pCM4.4ΔgenD1genS2			
genD1S2pfw	ACGCAGGTGTCGATGTCCTTCTGG	-	Annealed oligos for protospacer insertion
genD1S2prev	AAACCCAGAAGGACATCGACACCT		
genD1S2frag1fw	CTCGTTCGCTGCCACCTAAGAATACTCTACGGTCACATACCACCCAGCACCCGACCA	1060 bp	Amplification of Fragment 3
genD1S2frag1rev	CCGCTCCCGGATCGCCGGCCTTCTCCGGGCCCTGCACC		
genD1S2frag2fw	CAGGTGCAGGGCCCGGAGAAGGCCGGCGATCCGGGAGCG	1060 bp	Amplification of Fragment 4
genD1S2frag2rev	CGAGTAGTTCAGTAGCGGAAATGTCAGAGCCAGCGTCTTGATCTACAAGACGGCGCACG		
		873 bp	Validation of genD1, genS2 knockout in Streptomyces strains
genD1S2screenfw	CGTCGATCATCTCGCTCTCGAC		
genD1S2screenrev	CCGGACGTTGAAGTCGACCGT		
pSG-WS/pSG-WE			
gtmBoverlapprom	CCAGTAATGGGGGTTCAATACACTCCATTACTTATTACTCCCGGAAATTAGCC	1355 bp	

gtmBrev	TTAGCCCCGGAGCTCGCG		Amplification of genS1 with overhang to attach promoter (deftempspacer-sp44/ermEp1-deftempspacer)
gtmAoverlapprombl k	GGCTAAGGCCTGAAGACTTATCGCCCCCTTCGC ACCACACCTCATATA	1256 bp	Amplification of genC with overhang to attach promoter block (deftempspacer-fd-80bpspacer-sp44/ermEp1-deftempspacer)
gtmArev	TCAGCCGTCGGCGGCGCCG		
gacHoverlapprombl k	GGCTAAGGCCTGAAGACTTATCGCCGTATCATA TCACAGGGCGTCCCCT	1108 bp	Amplification of genE with overhang to attach promoter block (deftempspacer-fd-80bpspacer-sp44/ermEp1-deftempspacer)
gacHrev	TCACTCGTTGCGGAGGTCGA		
gtmGoverlapprom	CCAGTAATGGGGGTTCAATACACTCCAAGGCGC CACGGCGTTC	1334 bp	Amplification of genM1 with overhang to attach promoter (deftempspacer-sp44/ermEp1-deftempspacer)
gtmGrev	TCACTTGCCGGACGCG		
gtmMoverlappromb lk	GGCTAAGGCCTGAAGACTTATCGCCATAATAGC TCCTACGATTACCATTCG	847 bp	Amplification of genD with overhang to attach promoter block (deftempspacer-fd-80bpspacer-sp44/ermEp1-deftempspacer)
gtmMrev	TCAGGCGGAGCTACCGAGG		
gtmEoverlapprombl k	GGCTAAGGCCTGAAGACTTATCGCCGACGCGCA ACAACGGC	1355 bp	Amplification of genM2 with overhang to attach promoter block (deftempspacer-fd-80bpspacer-sp44/ermEp1-deftempspacer)
gtmErev	ACCGATACAATTAAAGGCTCCTTTTGGAGCCTT TTTTTTTGGAGATTTTTCAGCTCTCCTCCATCA GGGA		
promfw	TCTACGGGTCTCCAGGTCGACTCTCTACAGGGA TGCATTACA	108 bp	Forward primer for deftempspacer-sp44/ermEp1-deftempspacer
promoverlapgtmB	GGCTAATTTCCGGGAGTAATAAGTAATGGAGTG TATTGAACCCCCATTACTGG	108 bp	Reverse primer for deftempspacer-sp44/ermEp1-deftempspacer (genS1 overhang)

promoverlapgtmG	GAACGCCGTGGCGCCTTGGAGTGTATTGAACCCCCATTACTGG	108 bp	Reverse primer for deftempspacer- sp44/ermEp1-deftempspacer (genM1 overhang)
promblkfw	TCTACGGGTCTCGTGAACGCAACAACGGGTCTGA	280 bp	Forward primer for deftempspacer-fd-80bpspacer-sp44/ermEp1-deftempspacer
promblkoverlapgtmA	TATATGAGGTGTGGTGCGAAGGGGGCGATAAGTCTTCAGGCCTTAGCC	280 bp	Reverse primer for deftempspacer-fd-80bpspacer-sp44/ermEp1-deftempspacer (genC overhang)
promblkoverlapgacH	AGGGGACGCCCTGTGATATGATACGGCGATAAGTCTTCAGGCCTTAGCC	280 bp	Reverse primer for deftempspacer-fd-80bpspacer-sp44/ermEp1-deftempspacer (genE overhang)
promblkoverlapgtmM	CGAATGGTAATCGTAGGAGCTATTATGGCGATAAGTCTTCAGGCCTTAGCC	280 bp	Reverse primer for deftempspacer-fd-80bpspacer-sp44/ermEp1-deftempspacer (genD overhang)
promblkoverlapgtmE	GCCGTTGTTGCGCGTCGGCGATAAGTCTTCAGGCCTTAGCC	280 bp	Reverse primer for deftempspacer-fd-80bpspacer-sp44/ermEp1-deftempspacer (genM2 overhang)
promgtmBfwUNS1	CATTACTCGCATCCATTCTCAGGCTGTCTCGTCTCGTCTCTAGAGCGGTCTCCTCTACAG	1559 bp	Addition of UNS1 and UNS2 flanking sequences to deftempspacer-sp44/ermEp1-deftempspacer-genS1
fdgtmBrevUNS2	CTTGGATTCTGCGTTTGTTTCCGTCTACGAACTCCCAGCAAAATCTCCAAAAAAAAAGGC	280 bp	
promblkgtmAUNSfw2	GCTGGGAGTTCGTAGACGGAAACAAACGCAGAATCCAAGCTATTCCCCACGGGACAGGT	1514 bp	Addition of UNS2 and UNS3 flanking sequences to deftempspacer-fd-80bpspacer- sp44/ermEp1-deftempspacer-genC
fdgtmArevUNS3	GACCTTGATGTTTCCAGTGCGATTGAGGACCTTCAGTGCAAAATCTCCAAAAAAAAAGGC		
promblkgacHUNSfw3	GCACTGAAGGTCCTCAATCGCACTGGAAACATCAAGGTCGATTCCCCACGGGACAGGT	1361 bp	Addition of UNS3 and UNS4 flanking sequences to deftempspacer-fd-80bpspacer- sp44/ermEp1-deftempspacer-genE
fdgacHrevUNS4	ACTTTGCGTGTTGTCTTACTATTGCTGGCAGGAGGTCAGAAAATCTCCAAAAAAAAAGGC		
promgtmGfwUNS4	CTGACCTCCTGCCAGCAATAGTAAGACAACACGCAAAGTCTAGAGCGGTCTCCTCTACAG	1557 bp	Addition of UNS4 and UNS5 flanking sequences to deftempspacer-sp44/ermEp1-deftempspacer-genM1
fdgtmGrevUNS5	TCTAACGGACTTGAGTGAGGTTGTAAAGGGAGTTGGCTCAAAATCTCCAAAAAAAAAGGC		

promblkgtmMUNSfw5	GAGCCAACTCCCTTTACAACCTCACTCAAGTCCGTTAGAGATTCCCCACGGGACAGGT	1101 bp	Addition of UNS5 and UNS6 flanking sequences to deftempspacer-fd-80bpspacer- sp44/ermEp1-deftempspacer-genD
fdgtmMrevUNS6	ATGTGACCGTAGAGTATTCTTAGGTGGCAGCGAACGAGAAAATCTCCAAAAAAAAAGGC		
promblkgtmEUNSfw6	CTCGTTCGCTGCCACCTAAGAATACTCTACGGTCACATACATTCCCCACGGGACAGGT	1608 bp	Addition of UNS5 and UNS6 flanking sequences to deftempspacer-fd-80bpspacer- sp44/ermEp1-deftempspacer-genM2
fdgtmErevUNS7	GAGTAGTTCAGTAGCGGAAATGTCAGAGCCAGCGTCTTGAAAATCTCCAAAAAAAAAGGC		
pSET152revUNS1	GAGACGAGACGAGACAGCCTGAGAATGGATGCGAGTAATGTGCAGCACATCCCCCTTT	5795 bp	Amplification of pSET152 with UNS1 and UNS4 overhangs; for cloning of first half of UNS-gentamicin A2 construct (pSET152revUNS1 can be used in combination with pSET152fwUNS7 for cloning of entire construct)
pSET152fwUNS4	CTGACCTCCTGCCAGCAATAGTAAGACAACACGCAAAGTCAGGCGATTAAGTTGGGTAAC		
pSET152revUNS4	GACTTTGCGTGTTGTCTTACTATTGCTGGCAGGAGGTCAGTGCAGCACATCCCCCTTT	5795 bp	Amplification of pSET152 with UNS4 and UNS7 overhangs; for cloning of second half of UNS-gentamicin A2 construct (pSET152fwUNS7 can be used in combination with pSET152revUNS1 for cloning of entire construct)
pSET152fwUNS7	CAAGACGCTGGCTCTGACATTTCCGCTACTGAACTACTCGAGGCGATTAAGTTGGGTAAC		
UNS1F	CATTACTCGCATCCATTCTCAGGCTGTCTCGTCTCGTCTC	2904 bp	Screening for presence of genS1-genC fragment of pSG-WS/pSG-WE
UNS3R	CGACCTTGATGTTTCCAGTGCGATTGAGGACCTTCAGTGC		
pSG2			
pTE1364-UNS5RSP44	CACCGCACAGCATGTTGTCAAAGCAGAGACGGTTCGAATGTGAACCTCTAACGGACTTGAGTGAGGTTGTAAAGGGAGTTGGCTC	5487 bp	Amplification of pTE1364 at UNS5 site, partial sp44 sequence included in overhang
pTE1364D1F	CCGGGGCAGGTAGCCGCTGACGATCAAGGAGTGCCGCC	5487 bp	Amplification of pTE1364 with pTE1364-UNS5RSP44, overhang specific for genD1

pTE1364PF	CCAACTGCTGGACGAATTTCTCTGACGATCAAGGAGTGCCGCC	5487 bp	Amplification of pTE1364 with pTE1364-UNS5RSP44, overhang specific for genP
pTE1364f-UNS3	GCACTGAAGGTCCTCAATCGCACTGGAAACATCAAGGTCGCGATCAAGGAGTGCCGCCGGCCTCGGC	5527 bp	Amplification of pTE1364 with pTE1364-UNS5RSP44, overhang specific for UNS3 (addition of genB3/genB4)
C1a-genD2rbs0.5SP44	GTCTCTGCTTTGACAACATGCTGTGCGGTGTTGTAAAGTCTGGTGTAGGTAGGACACGGTGGGTGCTGCCGATG	1113 bp	Amplification of genD2 with overhangs for assembly
C1a-D2rev	TACCTGCCCCGGCGAGTCAGGCATTCATCTCCACTTTCCG	5487 bp	Amplification of pTE1364 with pTE1364-UNS5RSP44, overhang specific for genD1
C1a-genS2rbsfw	GGAAAGTGGAGATGAATGCCTGACTCGCCGGGGCAGGTAG	1325 bp	Amplification of genS2 with overhangs for assembly
C1a-genS2rev	CTCAGGACTCCTCCATGAGGGATCATAGGCTCTTCTTCAGCGCC		
C1a-D1fw	GGCGCTGAAGAAGAGCCTATGATCCCTCATGGAGGAGTCCTGAGATGACCGTCAC	2043 bp /2045 bp	Amplification of genD1 with overhangs for assembly (C1a-genD1rev for half-assembly, C1a-genD1revN for full assembly)
C1a-genD1rev	CTACTCCCGAGAGTCGGTTTCAGCGGCTACCTGCCCCGG		
C1a-genD1revN	CGCTACTCCCGAGAGTCGGTTTCAGCGGCTACCTGCCC		
C1a-genNfwD1	GGGCAGGTAGCCGCTGAAACCGACTCTCGGGAGTAGCG	1025 bp/1059 bp	Amplification of genN with overhangs for assembly (C1a-genNfwSP44 for half-assembly, C1a-genNfwD1 for full assembly)
C1a-genNfwSP44	GTCTCTGCTTTGACAACATGCTGTGCGGTGTTGTAAAGTCTGGTGTAGGTAAACCGACTCTCGGGAGTAGCGC		
C1a-genNrev	GACGCGGGCGGTGTCGTCGGTCAGCCCCGCAGAAGCCGGTC		
C1a-genB1rbsfw	ATCCCGACTTCTCGCTGTGAGGAACGAATCGAGCGGAGTG	1151 bp	Amplification of genB1 with overhangs for assembly

C1a-genB1rev	CCCTGTCGTTCGTACGCCTCAGGCTTCCGCCCATTGTGCA		
C1a-genQrbsfw	GACCGGCTTCTGCGGGGCTGACCGACGACACCGCCCGCGT	1588 bp	Amplification of genQ with overhangs for assembly
C1a-genQrev	CACTCCGCTCGATTCGTTCCTCACAGCGAGAAGTCGGGAT		
C1a-genPrbsfw	GCACAATGGGCGGAAGCCTGAGGCGTACGAACGACAGGGC	872 bp/855 bp	Amplification of genP with overhangs for assembly (C1a-genPrev for half-assembly, C1a-Prev-UNS2 for full assembly)
C1a-genPrev	CTGGTCGAGCTGGACGGTCAGAGAAATTCGTCCAGCAGTTGGATGTAG		
C1a-Prev-UNS2	GCTTGGATTCTGCGTTTGTTTCCGTCTACGAACTCCCAGCTCAGAGAAATTCGTCCAGCAGTTGG		
UNS2F	GCTGGGAGTTCGTAGACGGAAACAAACGCAGAATCCAAGC	2923 bp	Amplification of UNS2-kasO*p-genB3-genB4-fd-UNS3 (with UNS3R)

4. Author Contributions for Chapter 4

Katherine V. Baker was responsible for the identification of queuosine co-clustering in aminoglycoside gene clusters, the calculation of amino acid percentages in genome ORFs, the creation of heatmaps determining regions enriched in specific amino acids, the design, construction and testing of riboswitch constructs, and the design, construction and testing of *ΔgenA ΔgenF ΔgenG* knock-out strain.

Timothy Kirkwood was responsible for the initial creation of the workbook for calculation of amino acid percentages in genome ORFs (used for Figure 4.4, Table S4.1, detailed in sections 4.3.1., 4.5.3.).

4. Investigating the role of queuosine biosynthesis in the context of aminoglycoside biosynthetic gene clusters

Katherine V. Baker,[1] Timothy D. J. Kirkwood,[1] Nicholas Allenby,[2] Rainer Breitling,[1] Eriko Takano[1]

[1] Manchester Institute of Biotechnology, Manchester Synthetic Biology Research Centre SYNBIOCHEM, Department of Chemistry, The University of Manchester, Manchester, M1 7DN, United Kingdom

[2] Demuris Limited, The Biosphere, Draymans Way, Newcastle Helix, Newcastle upon Tyne NE4 5BX, United Kingdom

4.1. Abstract

Applying control to biological systems is a cornerstone of the field of synthetic biology but can be challenging when engineering new genetic circuits in heterologous hosts. Riboswitches are ligand-inducible or ligand-repressible gene expression regulators on the RNA level and can be a useful alternative to inducible promoter systems when creating genetic circuits. The pre-queuosine-1 class of riboswitches has been well-studied in the past but has rarely been described in actinomycetes. We identified that pre-queuosine-1 biosynthesis genes appears to co-cluster with certain 2-deoxystreptamine-based aminoglycoside biosynthetic gene clusters. We aimed to improve our understanding of whether pre-queuosine-1 may have an impact on cluster gene expression through binding to riboswitches and causing a repressive effect. We studied intergenic regions of a target biosynthetic gene cluster (Cluster 24 from *M.* sp. DEM32671) *in silico* and determined that the region upstream of putative queuosine biosynthesis genes may contain a pre-queuosine-1-linked riboswitch. We then tested this region in *Streptomyces coelicolor* by coupling the putative riboswitch region to an mCherry reporter, but found no significant mCherry reporter repression in the presence of pre-queuosine-1. Knock-out of predicted key queuosine biosynthesis genes *genA*, *genF*, and *genG* from Cluster 24 also appeared to increase antibiotic production,

though further characterisation is required to determine the full impact. We additionally tested two alternative pre-queuosine-1-responsive riboswitches in *S. coelicolor*, aiming to create a toolbox of pre-queuosine-1-responsive riboswitches for use in synthetic biology. One candidate riboswitch from *Mycobacterium abscessus* subsp. *abscessus* showed promise but requires further fine-tuning to be used in the construction of genetic circuits for Actinomycetes.

4.2. Introduction

In recent years, greater attention has been drawn towards the value of non-coding RNA in biological systems. In prokaryotes, it is believed that between 6 and 14% of the genome is a region of non-coding DNA.[1] Non-coding RNA as a group has been shown to have a multitude of functions in prokaryotes, including in pathogenicity,[2,3] defence against viruses (a key example being the guide RNAs of prokaryote defence system CRISPR-Cas9),[4] housekeeping,[5–7] and control of gene expression.[8–11]

Riboswitches function as a non-coding-RNA-based method of transcriptional and translational control. They are generally present in the 5′ untranslated region (5′-UTR) of messenger RNAs in bacteria, and function natively as small-molecule recognition systems for the control of translation.[9] Their action is two-fold: first, the recognition of a cognate ligand or ligands, and secondly, the respective conformational change in folding of the mRNA to either inhibit or allow for transcription or translation to occur. Riboswitches tend to inhibit these processes upon addition of ligand, due to the formation of complex secondary structure which can either conceal the ribosome binding site or form transcriptional terminators.[12] As of 2019 there were 28 experimentally-characterised classes of riboswitches in bacteria, mainly under the control of RNA-derived compounds such as nucleotide derivatives, coenzymes, and signalling molecules.[13]

Riboswitches are often identified as regulatory elements directly preceding genes for the biosynthesis of their associated ligand. Some examples of these include the riboswitches upstream of vitamin B1 and B12 biosynthesis genes, nucleic acid precursors biosynthesis genes and amino acid biosynthesis genes.[14,15] Riboswitch classes are relatively well conserved; this is hypothesised to be due to the lack of variety in available monomers to create regions for ligand-binding (four nucleotide bases in mRNA, in comparison to the much larger variety of amino acids available for binding of small molecules to proteins).[8] Consequently, computational analyses are suitable for identification of novel candidate riboswitches. A pipeline for the large-scale discovery of non-coding RNA hypothesised to form secondary structure in intergenic regions was described in 2019 by Stav et al. Starting from 2807 fully sequenced prokaryotic genomes, they identified a putative novel riboswitch class responsive to a biosynthesis intermediate of thiamine pyrophosphate.[16]

Queuosine, a 7-deazaguanoside-based nucleobase, has been identified in most kingdoms of life (yeasts and archaea being the main exceptions; *Thermus thermophilus*, once thought not to have genes for queuosine biosynthesis had these identified in 2015).[17–21] At the wobble-base position of anticodons with sequence 'GUN' (coding for aspartic acid, tyrosine, asparagine, and histidine), queuosine intermediate pre-queuosine-1 (preQ1) is incorporated and then further modified to queuosine to improve translational fidelity **(Figure 4.1)**.[22]

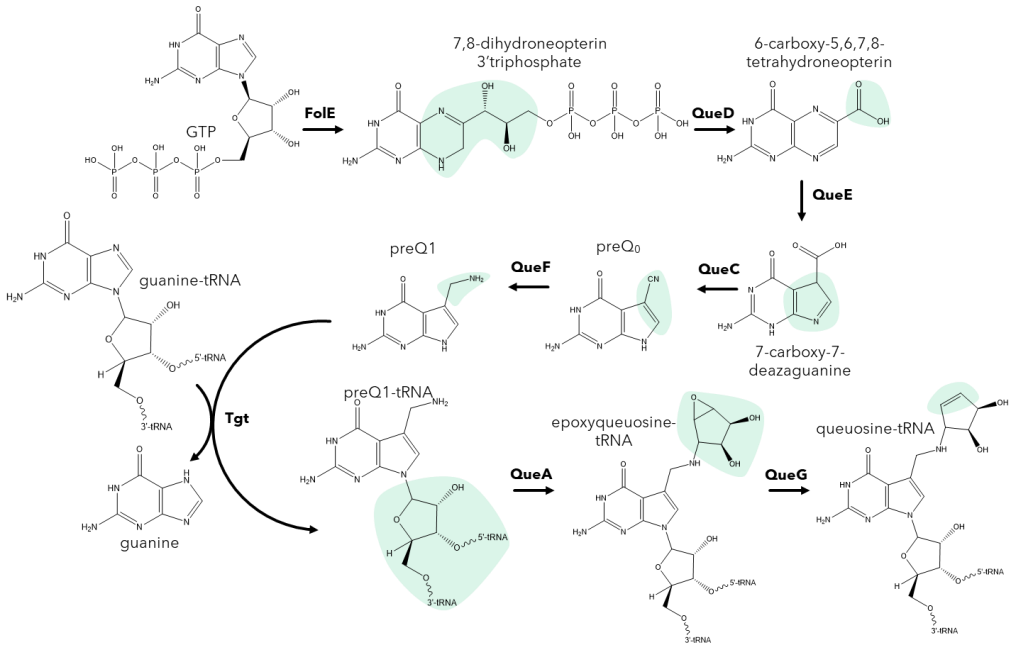


Figure 4.1: Biosynthesis pathway for queuosine-tRNA.[23]

Queuosine biosynthesis was first characterised in 2004 by Reader et al. with the discovery of a four-gene operon responsible for the production of intermediate preQ1 in *Bacillus subtilis*.[24] Further *in vitro* and *in vivo* (in *E. coli*) characterisation soon followed.[25–27] The control of queuosine metabolism has been linked to three classes of riboswitch responsive to queuosine-intermediate preQ1 (and immediately-upstream intermediate $preQ_0$), each differing in sequence, structure and mechanism.[28–31] The class I preQ1 riboswitches (preQ1-I) are some of the shortest known riboswitches (34 nucleotides required to sense ligand in *B. subtilis*).[32] Class II preQ1 riboswitches (preQ1-II) are described as being more specific to preQ1 induction, and have longer aptamers; McCown et al. (2014) state that the average aptamer length of preQ1-II riboswitches is 58 nucleotides.[30] Examples of both transcriptional and translationally-controlling preQ1 riboswitches have been described. The preQ1-I riboswitch from *B. subtilis* forms a pseudoknot structure upon binding of preQ1 to form a terminator hairpin and cease transcription, whilst another preQ1-I riboswitch from *Thermoanaerobacter tengcongensis* conceals the ribosome binding site upon preQ1 binding for translational control.[32,33]

Recently, metabolic engineers have found great value in riboswitches as an additional level of genetic control, aside from transcription factor systems.[34] The recent review by Hossain et al. (2020) gives a broad overview of how novel riboswitches have been generated to construct genetic biosensors to monitor production.[35] Over the past three decades, the process "systematic evolution of ligands by exponential enrichment" (SELEX) has enabled the evolution of aptamers specific to a wide variety of compounds enabling design of ligand-specific riboswitches.[36] For example, Jang et al. (2017) showed development of riboswitches to monitor *in vivo* production of naringenin, a flavonoid which can be used as a starting compound for synthesis of other flavonoids.[37] Riboswitches are also not limited to use in biosensors. Wang et al. (2019) also used a theophylline riboswitch in combination with a thiostrepton inducible promoter system and blue-light inducible split Cas9 (fused to fungal photo-receptor domains nMag and pMag) to combat the toxicity associated with expressing Cas9 in *Streptomyces*. Through this, they were able to increase both transformation efficiency and the success rate of genome editing.[38]

Upon examination of seven aminoglycoside clusters, we identified that queuosine biosynthesis genes appear to co-cluster within certain *Micromonospora*-linked aminoglycoside biosynthetic gene clusters. To show whether preQ1 may have an impact on aminoglycoside biosynthesis cluster gene expression, we studied intergenic regions of the Cluster 24 from *M.* sp. DEM32671 by *in silico* analysis and determined that the region upstream of putative queuosine biosynthesis genes *genA*, *genF*, and *genG* may contain a preQ1-linked riboswitch. We tested this region in *S. coelicolor* by coupling the putative riboswitch region to a codon-optimized mCherry reporter generated previously by Schlimpert al al. (2017),[39] but there was no apparent change in fluorescence upon induction of mCherry expression by presence of preQ1. Additionally, we tested alternative preQ1-responsive riboswitches, from *Mycobacterium abscessus* subsp. *abscessus* and *Lactobacillus rhamnosus* in *S. coelicolor*, with the aim of creating a toolbox of preQ1-responsive riboswitches for use in *Streptomyces* synthetic biology. Initial results showed a repressive

response using the *M. abscessus* subsp. *abscessus* riboswitch after induction with preQ1 in *S. coelicolor*. Finally, we investigated the impact *genA*, *genF*, and *genG* may have on biosynthesis of aminoglycoside by *M.* sp. DEM32671 by deleting these genes from the strain. We saw a slight increase in bioactivity from doing this, and believe this merits further investigation towards improving aminoglycoside yield from this gene cluster.

4.3. Results and Discussion

4.3.1. Queuosine biosynthesis genes are co-clustered in aminoglycoside biosynthetic gene clusters from *Micromonospora*

As of 2021, the gene cluster for gentamicin biosynthesis has largely been characterised. Recent publications (within the past five years) have shown routes to the terminal products gentamicin C2b and C1, and gentamicin B.[40–42] The last step of the pathway to be characterised was the role of GenB3 and GenB4 in biosynthesis of gentamicin C1a and C2a, previously understood to be carried out by GenP only.[41] Nevertheless, there remain several genes within the gentamicin gene cluster which have not been attributed to any biosynthetic function. Our previous work focused on improvement of product yield from a gene cluster homologous to gentamicin, Cluster 24 from *Micromonospora* sp. DEM32671. **Table 4.1** lists the uncharacterised genes in this cluster and the protein with highest percentage identity in the NCBI database. Interestingly, five of the nine identified genes have high similarity (>92% percentage identity) to queuosine biosynthesis genes, which merited further investigation. As the highest similarity was to *Micromonospora echinospora* DSM 43816 genes (which to the best of our knowledge have not been experimentally characterised, yet were assigned function), the predicted function of each query sequence was also investigated.

Table 4.1: List of proteins found in Cluster 24 homologous to those in *M. echinospora* DSM 43816 gentamicin gene cluster with currently experimentally-uncharacterised function. Conserved domains of interest were identified using BLASTp on the NCBI Conserved Domain Database.[43] Most similar proteins (measured by percentage identity) were identified through BLASTp on the NCBI database. Rows marked in blue are those implicated in biosynthesis of queuosine or queuosine intermediates.

Protein (*M.* sp. DEM32671)	Conserved domains of interest		Annotated protein with highest percentage identity			
	Labelled as	Accession No.	Labelled as	Species of origin	% identity	Accession No.
GenO	Tgt	pfam01702	queuine tRNA-ribosyltransferase family protein	*M. echinospora*	99.74%	AGB13894.1
GenD3	FabG	COG1028	putative gentamicin oxidoreductase	*M. echinospora*	99.70%	AGB13899.1
GenW	QueF-II	TIGR03139	preQ(1) synthase	*M. echinospora*	100%	WP_088981630.1
GenX	DUF664	pfam04978	DinB family protein	*M. echinospora*	97.06%	WP_088981624.1
GenU	WD40	COG2319	WD40 repeat domain-containing protein	*M. echinospora*	98.32%	SCE97073.1
GenY	KefB	COG0475	cation:proton antiporter	*M. echinospora*	97.81%	AGB13913.1
GenA	QueC	pfam06508	7-cyano-7-deazaguanine synthase	*M. echinospora*	98.22%	WP_088981619.1
GenF	QueE (Cx14CxxC type)	TIGR04508	7-carboxy-7-deazaguanine synthase	*M. echinospora*	97.67%	WP_088981618.1
GenG	QueD	COG0720	6-carboxytetrahydropterin synthase	*M. echinospora*	99.15%	WP_088981617.1

Comparing the putative queuosine biosynthesis genes from Cluster 24 to known sequences from *E. coli*, the sequence percentage identity ranges from 37.36% to 67.08% across all five protein sequences **(Table 4.2)**.

Table 4.2: Comparison of putative queuosine genes in Cluster 24 (*M.* sp. DEM32671) against known *E. coli* queuosine biosynthesis genes.

Protein in Cluster 24 (*M.* sp. DEM32671)	Protein in *E. coli*	% identity	Accession No. of *E. coli* sequence
GenW	NADPH-dependent 7-cyano-7-deazaguanine reductase QueF	42.97%	MZV02607.1
GenA	7-cyano-7-deazaguanine synthase QueC	60.91%	MRF41609.1
GenF	7-carboxy-7-deazaguanine synthase QueE	67.08%	WP_214293316.1
GenG	6-carboxytetrahydropterin synthase QueD	65.52%	EFF8507956.1
GenO	tRNA guanosine(34) transglycosylase Tgt	37.36%	TFQ26120.1

Queuosine biosynthesis genes appear to be co-localised with a sub-group of aminoglycoside BGCs. Homologues to the 7-cyano-7-deazaguanine synthase QueC from *B. subtilis* 168 can be found in the gentamicin producers *M. echinospora* and *Micromonospora pallida*, sagamicin producer *Micromonospora sagamiensis*, sisomicin producer *Micromonospora inyonensis,*[44] and in the fortimicin producer *Micromonospora olivasterospora* **(Figure 4.2)**.[45] Interestingly, queuosine biosynthesis genes only appear to be closely clustered in aminoglycoside BGCs from *Micromonospora* strains. They are also more apparent in BGCs that produce gentamicin or related 4,6-disubstituted 2-deoxystreptamine aminoglycosides, such as sisomicin and sagamicin. Though kanamycin also has a similar early biosynthesis pathway to that of gentamicin (and shares similar cluster architecture for those early biosynthesis genes *genS1*, *genC*, *genM2*, *genS2*/*kanB*, *kanA*, *kanE*, *kanD*) no queuosine biosynthesis genes are identified in this BGC.[45] This may be due to the kanamycin producer

not being part of the *Micromonospora* family, as kanamycin is produced by *Streptomyces kanamyceticus*. Other aminoglycoside gene clusters such as the neomycin, tobramycin, paromomycin, lividomycin BGCs expressed by *Streptomyces* spp. do not contain queuosine biosynthesis genes.[45–49] This is also the case for the *Bacillus*-derived butirosin cluster.[45] This could suggest that the co-clustering of queuosine biosynthesis genes is occurring early in the evolution of these bacteria.

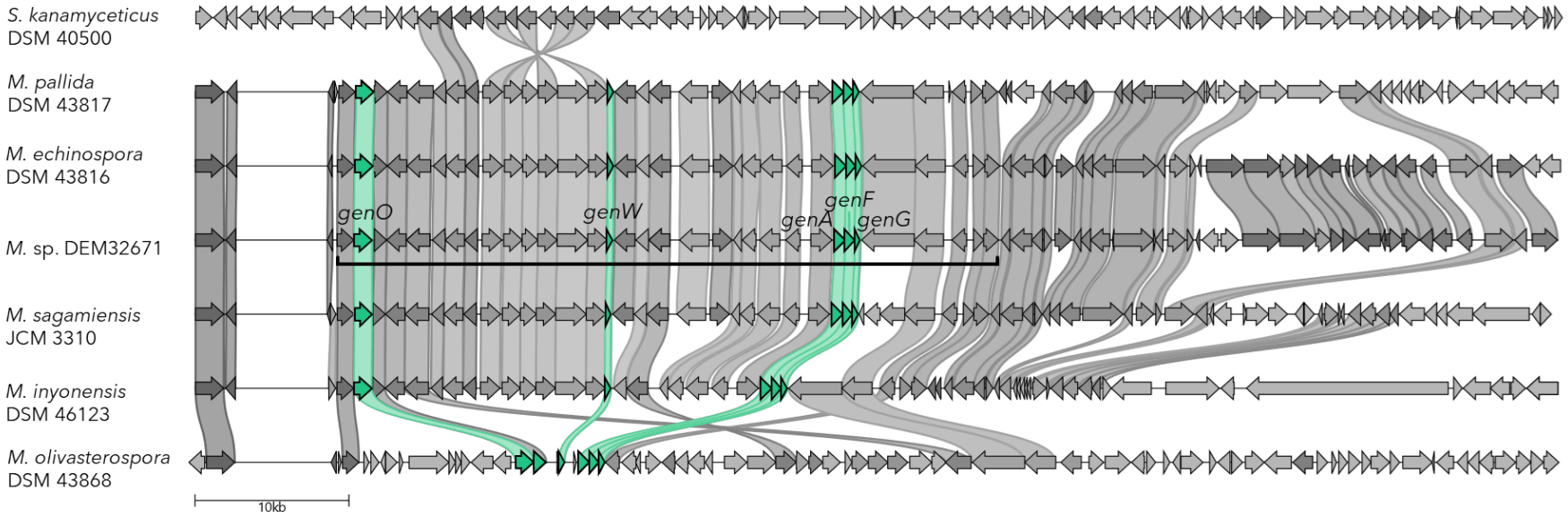


Figure 4.2: clinker diagram of sequence similarity between *M.* sp. DEM32671 Cluster 24 and similar aminoglycoside gene clusters, with genes implicated in queuosine biosynthesis marked in green.[50] The best link for each gene is visualised; similarity threshold for link drawing = 50%. Annotation on *M.* sp. DEM32671 marks core region of Cluster 24 from *genO* to *genN*.

The co-localisation of queuosine biosynthesis genes in these aminoglycoside biosynthetic gene clusters (BGCs) was interesting, especially considering the clusters for which this appears to be occurring. There are no clear aminoglycoside targets for the queuosine biosynthesis genes, and so it remains likely that these are acting to produce separate queuosine intermediates. The sisomicin and sagamicin gene clusters were likely to be modified from the gentamicin BGC at some point in evolution; they differ only by a few genes responsible for late-stage gentamicin C biosynthesis and have very similar chemical structures (differing only by a single side-chain on the purpurosamine ring, and double-bond introduction to the C4′ and C5′ atoms of the purpurosamine ring). It was recently revealed that sisomicin is an intermediate of gentamicin biosynthesis **(Figure 4.3)**.[41] Sagamicin is another name for gentamicin C2b, and *M. sagamiensis* remains capable of producing end-product gentamicins C1, C2 and C1a as minor products.[51] Therefore, the queuosine biosynthesis genes in these clusters could have been introduced at a single point in evolution and maintained as the clusters differentiated. The fortimicin BGC, however, remains structurally very different from the 4,6-disubstituted 2-deoxystreptamine aminoglycoside clusters, and still maintains homologues of all five queuosine biosynthesis genes maintained in the other clusters.

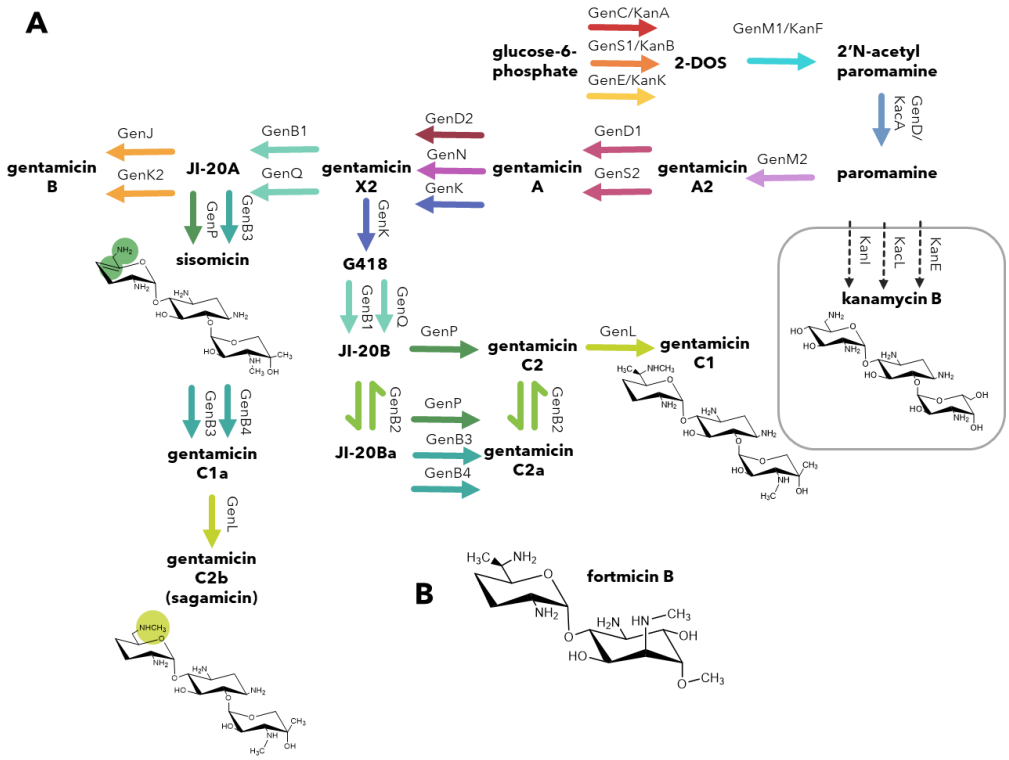


Figure 4.3: Structures and known pathways of biosynthesis for aminoglycosides mentioned in this work.[40–42,52–56] **A** – known biosynthesis from glucose-6-phosphate for 2-deoxystreptamine-derived aminoglycosides. Kanamycin B biosynthesis is possible from paromamine via the route shown, though other kanamycins are biosynthesised from 2-deoxystreptamine.[52,53] **B** – structure of pseudodisaccharide fortimicin B.

Queuosine improves translational fidelity, by being incorporated into the first position of certain tRNAs (asparagine, aspartic acid, histidine and tyrosine) and replacing similar nucleoside, guanosine.[22] A possible reason for queuosine biosynthesis genes being co-clustered with gentamicin biosynthesis genes is that the gentamicin cluster could have an increased percentage of asparagine, aspartic acid, histidine and tyrosine residues in comparison to the rest of the genome. This could merit a localised mechanism to ensure translation occurs accurately. Upon investigation of the queuosine biosynthesis pathway, it was determined that it was likely to be impossible to reach the end-point of queuosine being incorporated into the tRNA in these *Micromonospora* strains due to the pathway being incomplete, rendering the hypothesis that these genes were present to increase translational fidelity likely to be untrue. In

native cluster host *M.* sp. DEM32671, the biosynthetic machinery is present up until the incorporation of preQ1 to the tRNA, suggesting both preQ1 and immediate precursor $preQ_0$ are of most interest **(Figure 4.4)**. 7-deazapurine-based secondary metabolites have been described since 1956, with the discovery of combination antibiotic/antitumour compound toyocamycin from *Streptomyces toyocaensis*.[23,57] Xu et al. showed in 2014 that $preQ_0$ has cytotoxic activity in human cell lines and could possibly be used as an anti-cancer therapeutic.[58] It is possible that $preQ_0$ could also function as an antibiotic, and preQ1 as a bioactive molecule, though to the best of our knowledge experiments towards testing this have not yet been published.

Despite the pathway towards queuosine appearing incomplete, the percentage of asparagine, aspartic acid, histidine and tyrosine residues were calculated for each *M.* sp. DEM32671 Cluster 24 BGC protein to determine whether it is likely this function may be relevant **(Figure 4.4, Table S4.1)**. Those percentage values which are above average are marked by colour gradient from purple to yellow, yellow being the highest value in each dataset. By splitting the proteins into two subgroups (region 1: proteins predicted to be of biosynthetic relevance, region 2: proteins predicted to be of no biosynthetic relevance; marked in **Figure 4.4**) it is possible to calculate that the region with predicted biosynthetic relevance has a Asn + Asp + His + Tyr content of 13.391%, while the region without predicted biosynthetic relevance has a lower Asn + Asp + His + Tyr content of 12.026%. Additionally, the proteins attributed towards queuosine biosynthesis have a slightly higher Asn + Asp + His + Tyr content of 14.032% in comparison to those attributed to gentamicin/aminoglycoside biosynthesis (13.387%). Orf_5081, present in the region without predicted biosynthetic relevance and a clear outlier in this group, contains five poly-aspartic acid stretches interspaced with arginine at random intervals (5xAsp–Arg–7xAsp–Arg–4xAsp–Arg–9xAsp–Arg–3xAsp) at the C-terminus. This motif is present in the second-closest hit to this sequence when aligned with BLASTp (PepSY domain-containing protein, 89.77% identity to WP_088981612.1 from *Micromonospora echinospora*) but has been removed from the closest hit

(hypothetical protein, 92.62% identity to ARV75732.1 from *Micromonospora echinospora*). As this protein has not yet been assigned function, it is unclear the relevance of this stretch of amino acids. Removing this sequence reduces the Asn + Asp + His + Tyr content across the group to 11.706%, which is below the whole genome average of 11.981%.

Investigating the pathway further, a secondary hypothesis was built: that perhaps these BGCs rely on queuosine intermediates for regulation of aminoglycoside biosynthesis. Of the aminoglycosides, complex regulation has only been described for the non-deoxystreptamine-based streptomycin. In this case, microbial hormone A-factor causes a signalling cascade leading to streptomycin production in *Streptomyces griseus*.[59–62] We identified that the final intermediate possible in *M.* sp. DEM32671 is preQ1, which is likely to be incorporated into the tRNA by action of GenO. However, we were unable to find a homologue to QueA or QueG/QueH, which carry out the final steps of modification through to queuosine-tRNA.[63,64]

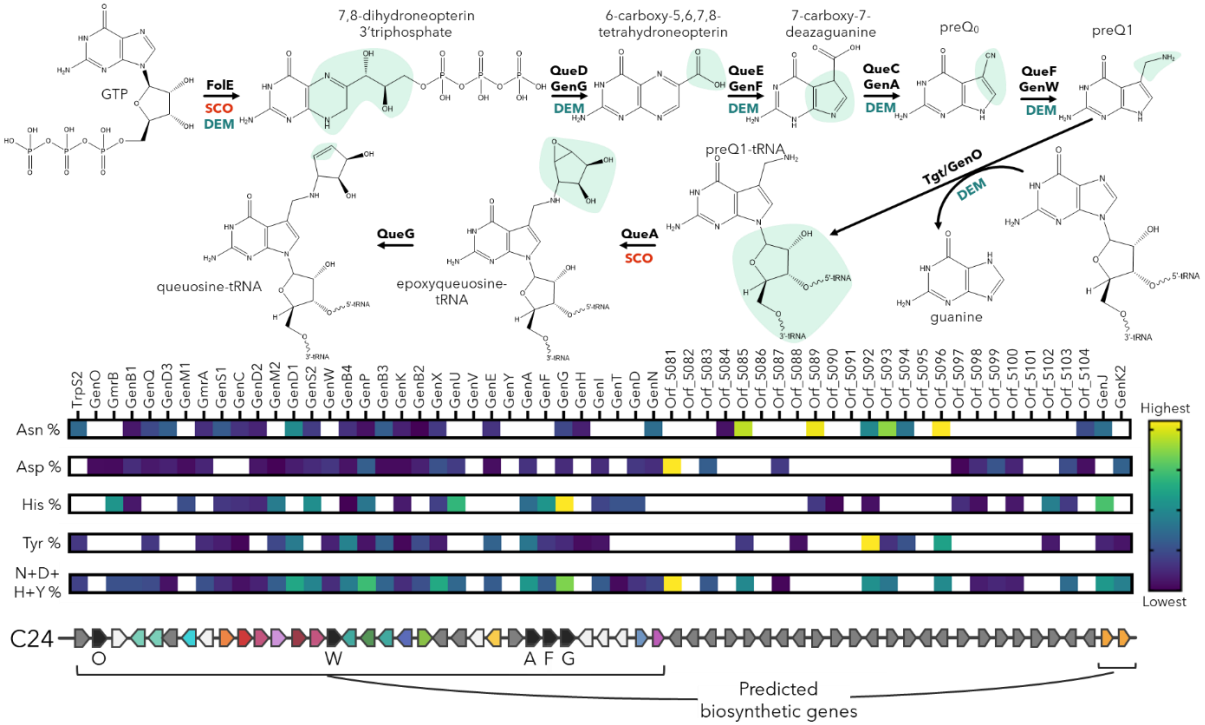


Figure 4.4: Pathway of queuosine-tRNA biosynthesis and heatmap of percentage of queuosine-linked residues in *M.* sp. DEM32671 Cluster 24 proteins. Where genes in the gentamicin BGC are homologous to queuosine biosynthesis genes, both names are given. Enzymes for which genes are present in the *S. coelicolor* A3(2) genome are marked in red 'SCO'; genes present in the *M.* sp. DEM32671 genome are marked in green 'DEM'. Each heatmap was calculated separately based on average percentage in entire M. sp. DEM32671 genome (average Asn % = 1.6768%; Asp % = 6.0987%; His % = 2.2463%; Tyr % = 1.9593%; Asn + Asp + His + Tyr % = 11.9811%) and coloured only if % of the residue(s) exceeds the average for that residue(s); colours range from dark purple (lowest value for that residue) to yellow (highest for that residue). Cluster schematic is colour-coded as follows: homologues to gentamicin biosynthetic genes are coloured; queuosine biosynthetic genes are black; gentamicin transport and resistance genes are white.

4.3.2. Investigating putative preQ1 riboswitches in *M.* sp. DEM32671 Cluster 24

It was next posited that one reason why biosynthesis of preQ1 is clustered in the aminoglycoside BGCs is that portions of these clusters may be under the regulation of preQ1 riboswitches. The only example of a preQ1 riboswitch in Actinobacteria in the Rfam database was in *Mycobacterium abscessus* subsp. *abscessus*. To the best of our knowledge, the experimental characterisation of this riboswitch has not been published, with a 2017 paper by van Vlack et al. remarking that mycobacteria appear to lack preQ1 biosynthetic genes entirely. In their work, they aimed to use preQ1 riboswitches in mycobacteria as they would remain orthogonal to native inducible systems.[65] A third riboswitch, from *Lactobacillus rhamnosus*, was chosen as it was confirmed to function in mycobacteria in the work by van Vlack et al.; additionally, the structure had previously been determined by X-ray crystallography.[31,65] As mycobacteria were the closest relative to *Streptomyces* that could be identified in the literature to have preQ1 riboswitches tested, it was deemed likelier that the *M. abscessus* would be able to function in *Streptomyces* than the *L. rhamnosus* riboswitch.

The 97 bp long sequence from *M. abscessus* was used to probe for similar sequences in *M.* sp. DEM32671 Cluster 24. The analysis identified a region of similarity upstream of *genA* (homologous to *queC*; **Figure 4.5**). This is to be expected in the context of these riboswitches, being most commonly found upstream of their own biosynthesis genes.[28] Aligning these two sequences with ClustalX showed some conservation of key bases described by Weinberg et al., (2007) though the majority of highly conserved regions were not present.[66] preQ1-II riboswitches are challenging to predict *in silico* due to their H-type pseudoknot structure.[67] In this type of structure, a stretch of basepairs is formed between the loop of an RNA hairpin and the single-stranded region.[68] ProbKnot was first used as a comparison of the *M. abscessus, L. rhamnosus* preQ1-II riboswitches and the putative riboswitch region from *M.* sp. DEM32671 Cluster 24. Whilst all three showed predicted pseudoknots, the *M.*

abscessus riboswitch showed greater similarity to the canonical pseudoknot of the consensus sequence, which was not clear in the *M.* sp. DEM32671 region under investigation. All four sequences maintained key preQ1 nucleotide binding positions shown by Liberman et al. (on the *L. rhamnosus* sequence, C30 and U41); [31] however, only the *M. abscessus* and *L. rhamnosus* riboswitches maintained the nucleotides highlighted by Kang et al. to form the rest of the preQ1 binding pocket.[67]

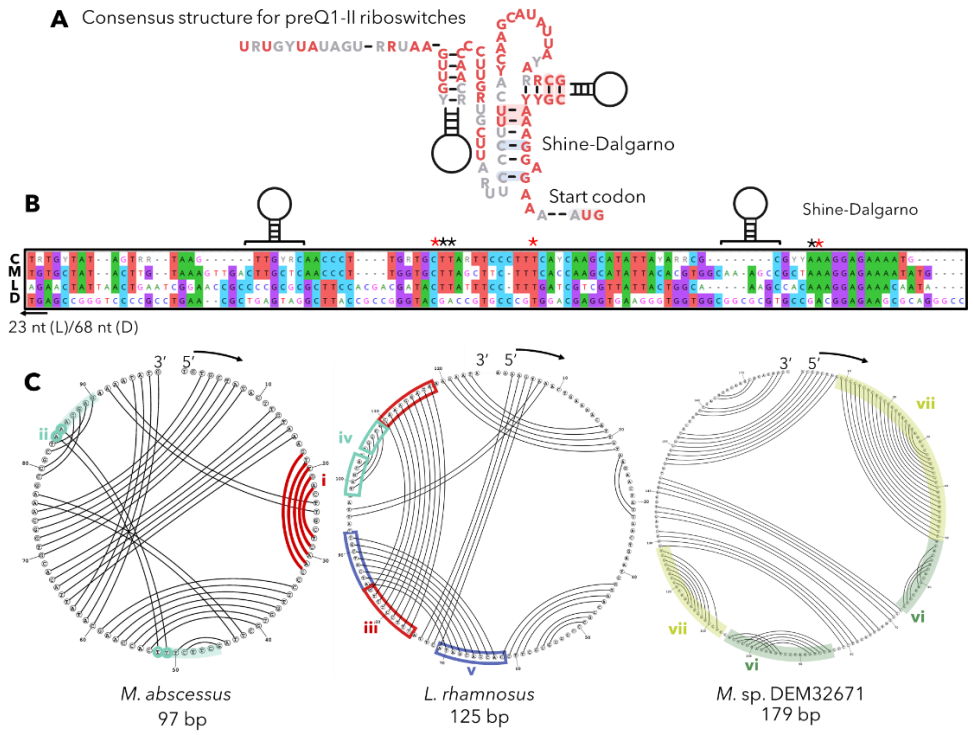


Figure 4.5: Comparison of the *M. abscessus* (accession number FSGV01000016.1), *L. rhamnosus* (accession number AEYM01000577.1 truncated as in van Vlack et al. (2017))[65] and *M.* sp. DEM32671 putative riboswitch DNA sequences against the consensus sequence DNA for preQ1-II riboswitches. A – consensus sequence for preQ1-II riboswitches, taken from McCown et al. (2014)[30] **B** – alignment of candidate riboswitch sequences to preQ1-II ribsowtch consensus sequence. C = consensus sequence, M = *M. abscessus* subsp. *abscessus* riboswitch sequence, L = *L. rhamnosus* riboswitch sequence, D = *M.* sp. DEM32671 candidate riboswitch sequence. Alignment was done using ClustalW; bases are coloured where >50% conserved.[30] Conserved bases found to be structurally relevant by Liberman et al. and Kang et al. are marked with asterisks (red for those conserved across all 4 sequences, black for those conserved across all except the *M.* sp. DEM32671 sequence.) [31,67] **C** – ProbKnot was used to estimate pseudoknots in *M. abscessus*, *L. rhamnosus* and *M.* sp. DEM32671 sequences (5 iterations, minimum helix length of 2); pseudoknots estimated based on cross-linking of sequence. i) first hairpin in *M. abscessus* riboswitch sequence; ii) proposed Shine-Dalgarno sequestration site is not fully identified; iii) P3 stem of *L. rhamnosus* riboswitch; iv) P4 stem of *L. rhamnosus* riboswitch; v) P2 stem of *L. rhamnosus* riboswitch;[67] vi) putative pseudoknot regions vii) putative hairpin regions.

The ProbKnot results align in some cases with the literature consensus folds. For example, the 'GTTGA' pairing to 'TCAAC' in the *M. abscessus* DNA

sequence matches with the first hairpin in the structure (Figure 4.5Ci), but the sequence most closely resembling the consensus sequence for the second hairpin does not show the expected linkage. The pseudoknot region '5′ – CCCUUU – 3′ pairing to '5′ – AAAGG(A)G – 3′ (or 5′ – TCTTTT – 3′ to 5′ – AAAGG(A)G – 3′ in the case of the *M. abscessus* DNA sequence) seems intact, but ProbKnot was not able to annotate a link longer than 2nt long ('AA') at position 84+85 to 'TT' at position 51+52 (Figure 4.5Cii). While the *L. rhamnosus* riboswitch sequence does not match the consensus as well as the *M. abscessus* sequence, the sequence which binds to sequester the Shine-Dalgarno sequence (P3 stem; 5′ – UUUCCUUUG – 3′; (Figure 4.5Ciii)) remains intact, as does the P4 stem (Figure 4.5Civ) and P2 stem (Figure 4.5Cv). The *L. rhamnosus* preQ1 riboswitch pseudoknot structure has been confirmed experimentally,[31] and so the success of ProbKnot here can be verified. Analysing the *M.* sp. DEM32671 sequence remains difficult: the sequence did not show clear runs aligning to the consensus, but ProbKnot did show some pseudoknots (Figure 4.5Cvi) and hairpins (Figure 4.5Cvii) indicating that there is secondary structure present. Without experimental characterisation, it remained challenging to predict whether this secondary structure represented a preQ1-responsive region.

A secondary tool, SPOT-RNA, was used to further investigate the folding of the sequences *in silico* **(Figure 4.6)**. SPOT-RNA functions to predict RNA secondary structure using deep learning and has been shown to significantly improve upon pseudoknot prediction than other similar tools.[69] The *M. abscessus* and *L. rhamnosus* riboswitches were included for comparison. Despite the *L. rhamnosus* riboswitch not matching the preQ1-II consensus as closely as the *M. abscessus* riboswitch, SPOT-RNA predicts more non-nested interactions (i.e., pseudoknot interactions) in the *L. rhamnosus* riboswitch. Interestingly, the two-dimensional secondary structure prediction appears very different between the three sequences tested, though the *M. abscessus* and *L. rhamnosus* sequences are more recognisable in comparison to the known consensus, and the secondary structure upstream of *genA* in Cluster 24 consists of longer hairpin sequences which do not appear to interact with each other. The base-

pair probability plots confirm the likeliness of these hairpins, according to the SPOT-RNA model.

SPOT-RNA was also used to check the upstream secondary structure of two other genes implicated in queuosine biosynthesis: *genO*, which is predicted to function to swap preQ1 with a guanine linked to a tRNA, and *genW*, which is predicted to catalyse the reaction between $preQ_0$ and preQ1 **(Figure 4.7)**. 100 bp of the sequence upstream of *genW* was taken for analysis, as the stop codon of *genS2* is only 7 bp upstream, whilst the region between *genO* and the stop codon of upstream gene *trpS2* (66 bp) was used. Each sequence had two stems and one loop predicted in the two-dimensional secondary structure, which does not match the single stem-loop seen in the shorter preQ1-I class. In addition, no sequence matches to the preQ1-I riboswitch consensus sequences described by McCown et al. were observed.[30] These were not selected for experimental characterisation due to i) the short length of the sequence upstream of *genW* which was not part of the coding sequence of the gene before it ii) *genO* only being identified as potentially biosynthetically relevant late in this work. As riboswitches are generally found in the 5′-UTR of coding sequences,[65,70] it remains highly unlikely that there is one present upstream of *genW*; however, the region upstream of *genO* may provide a further target for investigation in the future.

Pseudoknots remain challenging to predict *in silico*; they remain much more complex to predict than standard hairpins due to the requirement of testing the pairing of bases within a hairpin loop to bases outside it.[71] Experimentally, this could be determined with crystallisation after confirmation of function. With these *in silico* predictions, it appeared that some secondary structure was present upstream of *genA* in Cluster 24, but experimental characterisation was required to confirm whether it represented a region responsive to preQ1.

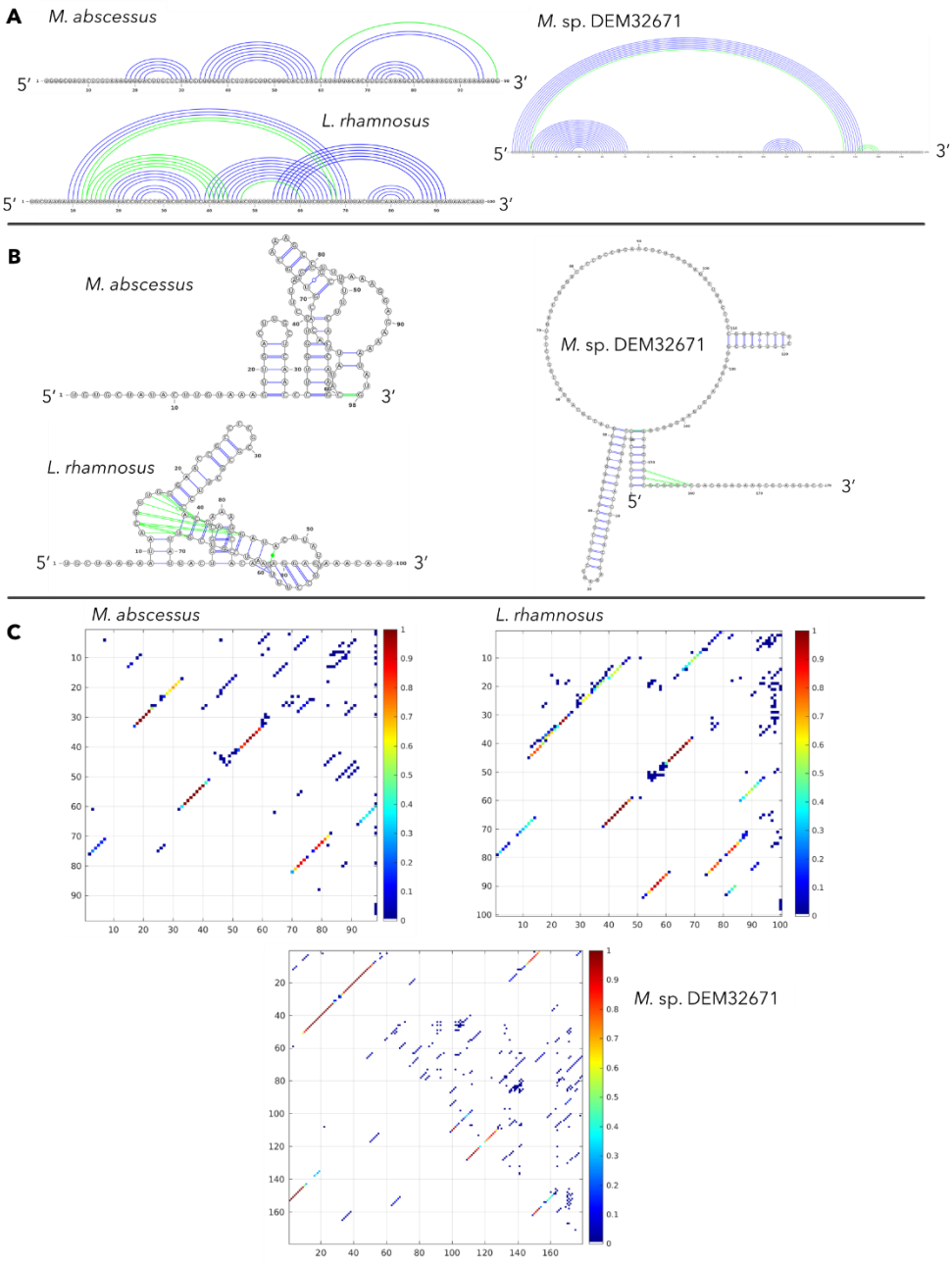


Figure 4.6: SPOT-RNA output using *M. abscessus* subsp. *abscessus*, *L. rhamnosus* preQ1-II riboswitch sequences and comparing to *M.* sp. DEM32671 *genA* upstream region. A – arc-representation of RNA secondary structure. Canonical base pairs are coloured in blue arcs and non-nested/non-canonical base pairs are in green. **B** – predicted secondary structure folding. Canonical base pairs are coloured in blue arcs and non-nested/non-canonical base pairs are in green. **C** - Base-pair probability plot (upper triangle uses ensemble of 5 models for prediction, lower triangle uses baseline SPOT-RNA model).

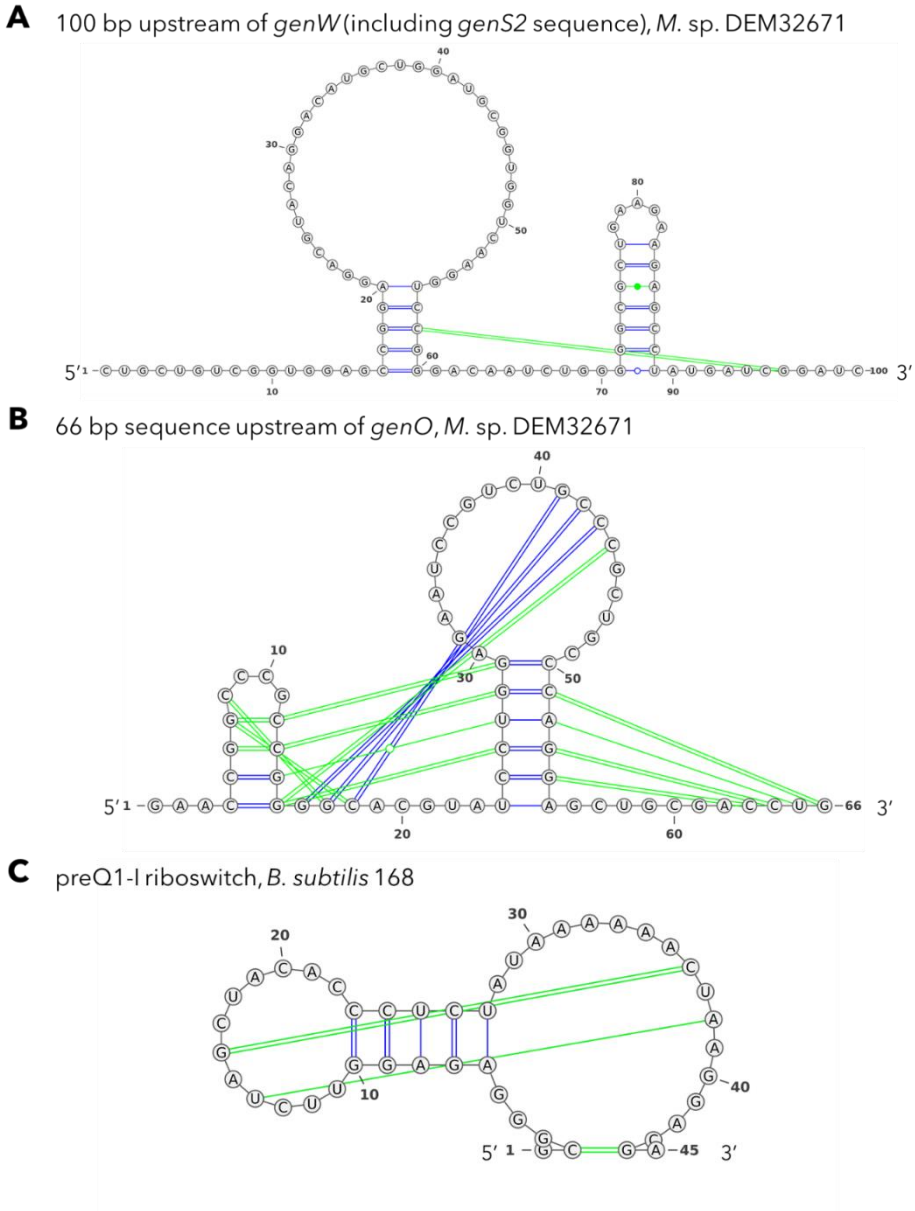


Figure 4.7: SPOT-RNA output of regions upstream of *genO* and *genW* with known preQ1-I riboswitch from *B. subtilis* 168 (AL009126.3) for comparison. Canonical base pairs are coloured in blue arcs and non-nested/non-canonical base pairs are in green. A – Predicted secondary structure of *genW* upstream sequence truncated to 100 bp. **B** – Predicted secondary structure of *genO* upstream region folding. **C** – Predicted secondary structure of known preQ1-I riboswitch from *B. subtilis* 168.[28]

4.3.3. The region upstream of *genA* does not show canonical preQ1 riboswitch activity

To test the response of the putative riboswitch region (identified in the *M.* sp. DEM32671 gentamicin BGC) to preQ1, plasmids based on pTE1332 were constructed. pTE1332 is a pSET152-based plasmid with mCherry codon-optimized for *Streptomyces* expression,[39] under the control of the medium-strength *ermEp1* promoter and a synthetic ribosome binding site described by Horbal et al.[72] The first constructs tested contained the riboswitch region truncated to 30 nt to test the strength of the Shine-Dalgarno sequence without interference from upstream secondary structure **(Table S4.3)**, inserted between the promoter and coding sequence via amplifying the whole plasmid with primer overhangs. After successful conjugation into *S. coelicolor* M1146 was confirmed by PCR, the cells were grown in liquid culture (smaller scale, 15 mL media) and fluorescence measured at three time points **(Figure 4.8)**. For normalisation of data for comparison, total fluorescence for each sample was divided by total protein of the same sample (as a measure of bacterial growth). Samples from positive control (*S. coelicolor* M1146 pTE1332) showed strong fluorescence peaking at 48 hours growth. The only putative truncated riboswitch to show any measurable fluorescence from mCherry was pTE1332-Mabs-trunc (truncated from *M. abscessus* preQ1-II riboswitch). This suggested that the only suitable Shine-Dalgarno sequence for expression was in this construct. The sequence aligns well with the *Streptomyces* consensus RBS (5′ – AAGGAGGU – 3′; pTE1332-Mabs-trunc RBS 5′ – AAGGAG**AA**– 3′; different bases underlined and bolded) which contributes to the good function across both strains.

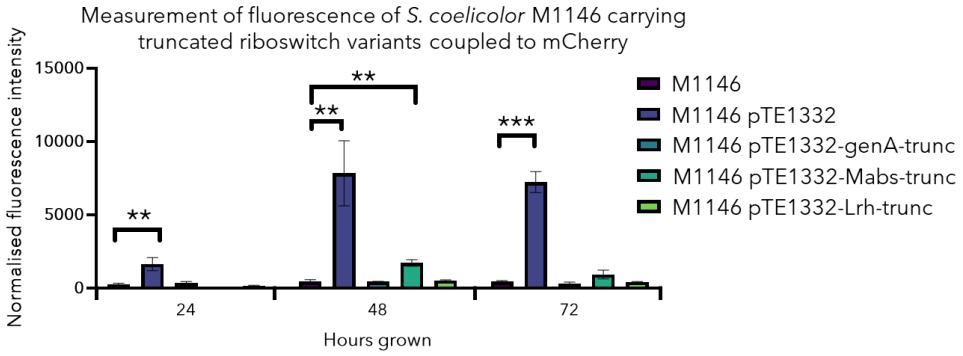


Figure 4.8: Measurement of normalised mCherry fluorescence under control of truncated riboswitch sequences (30 nt). M1146 represents negative control (*S. coelicolor* M1146, no plasmids), M1146 pTE1332 represents positive control (*S. coelicolor* M1146 pTE1332 with synthetic ribosome binding site between *ermEp1* and mCherry.[72] pTE1332-genA-trunc (*S. coelicolor* M1146 pTE1332-genA-trunc) includes truncated sequence from upstream of *genA* homologue from *M.* sp. DEM32671 upstream of mCherry; pTE1332-Mabs-trunc (*S. coelicolor* M1146 pTE1332-Mabs-trunc) includes truncated sequence from *M. abscessus* subsp. *abscessus* preQ1-II riboswitch upstream of mCherry; pTE1332-Lrh-trunc (*S. coelicolor* M1146 pTE1332-Lrh-trunc) includes truncated sequence from *L. rhamnosus* preQ1-II riboswitch upstream of mCherry. Statistical significance of experimental samples were compared against negative control M1146 using unpaired t-test. In comparison of fluorescence difference between samples * denotes statistical significance $p < 0.05$; ** denotes statistical significance $p < 0.001$; *** denotes statistical significance $p < 0.0001$.

After further investigation, it was determined that there were two errors likely in the design, arising from *genA* start codon misannotation from the genome sequence of *M.* sp. DEM32671 and reproduction of the *L. rhamnosus* riboswitch design established in the literature. Firstly, the annotated start codon for *genA* differs between the *M. echinospora* DSM 43816 genome sequence, the MiBIG gentamicin gene cluster (BGC0000696) and the *M.* sp. DEM32671 genome annotation. To determine which of these was the correct start codon, the Salis Lab Ribosome Binding Site Calculator was used as a prediction of Shine-Dalgarno sequence.[73] From this, it was determined that the start codon annotated in the *M. echinospora* DSM 43816 genome (ORF 2) was likely to be correct as it had the highest translation initiation rate. Additionally, ORF 1, annotated from the MiBIG gentamicin gene cluster (BGC0000696), yields a

protein 88 amino acids long with no significant sequence similarity to any sequences available via BLASTp; the start codons for ORF 2 and ORF 3 are in frame with each other. The translation initiation rate has been shown to change considerably with the sequence of the protein directly downstream,[74] and so the strength of each sequence (truncated to the 30 nt directly upstream of the predicted start codon) controlling translation of mCherry was also predicted. ORF 2 also provided the strongest predicted translation initiation rate of 63.85 (au), while the sequence upstream of the previously annotated start codon (ORF 3) showed a predicted translation initiation rate of only 0.81 (au; **Figure 4.9**).

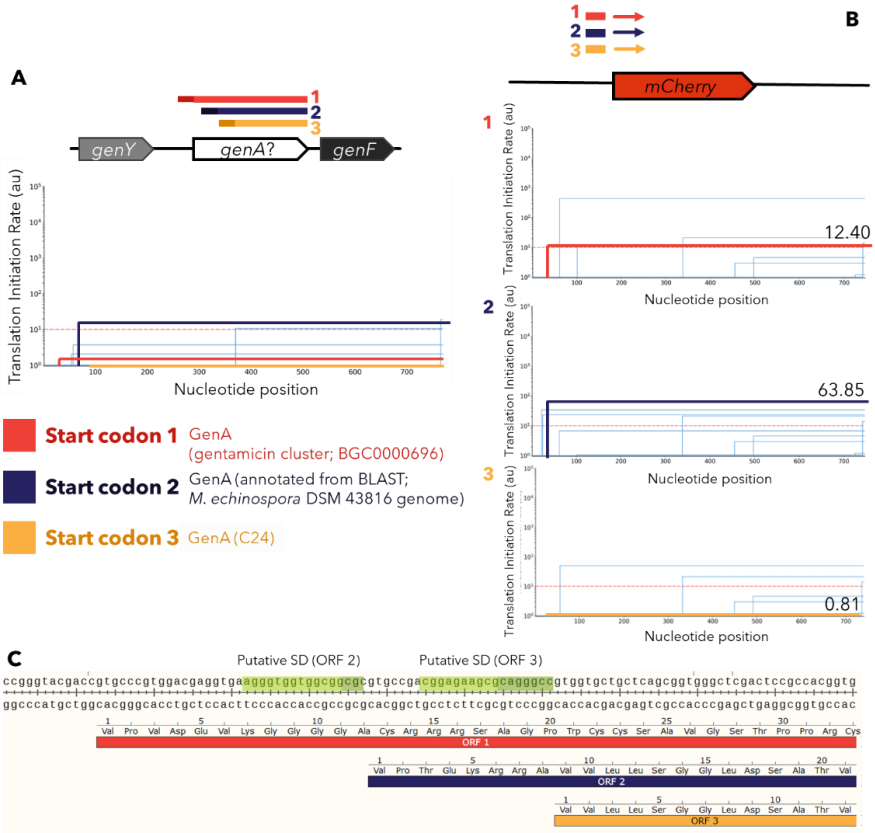


Figure 4.9: *In silico* estimation of *genA* start codon based on Salis Lab Ribosome Binding Site Calculator.[73] **A** – Sequence upstream of *genA* plus the entire *genA* sequence were input and the strongest translation initiation rate marked as putative ORF 2, matching the *M. echinospora* DSM 43816 genome sequence. **B** – 30 nt upstream of each candidate start codon was selected as candidate drivers of *genA* expression. *genA* was replaced by *mCherry* downstream of each 30 nt sequence to determine suitability in the context of the pTE1332-based test plasmids. **C** – genetic context of each ORF with putative Shine-Dalgarno sequences marked in green.

Interestingly, both old and new proposed *genA* RBS sequences differed from the *Streptomyces* consensus by three base-pair changes (5′ – AAGGAGGU – 3′; new proposed *genA* RBS 5′ – **AG**GG**T**GGU – 3′; old proposed *genA* RBS 5′ – A**C**GGAG**AA** – 3′; different bases underlined and bolded). The spacing of these sequences were different (8 nt between last proposed RBS nucleotide and start codon for new RBS, 10 nt last proposed RBS nucleotide and start codon for old RBS) which is likely to have contributed to the different results seen via strength prediction.

As changing the assigned start codon truncated the sequence upstream of *genA* to 154 bp instead of 179 bp, the *in silico* structure prediction was repeated **(Figure 4.10)**. There did not appear to be any significant differences in secondary structure formation with the reduction in sequence size, as the predicted secondary structure generally formed at the 5′ end of the sequence and this was unaffected by the reassignment of the start codon of *genA*. However, the sequence similarity to the confirmed riboswitch sequences appeared to have lessened slightly (45 bp shared by at least one other candidate, within consensus region, in comparison to a previous 56 bp).

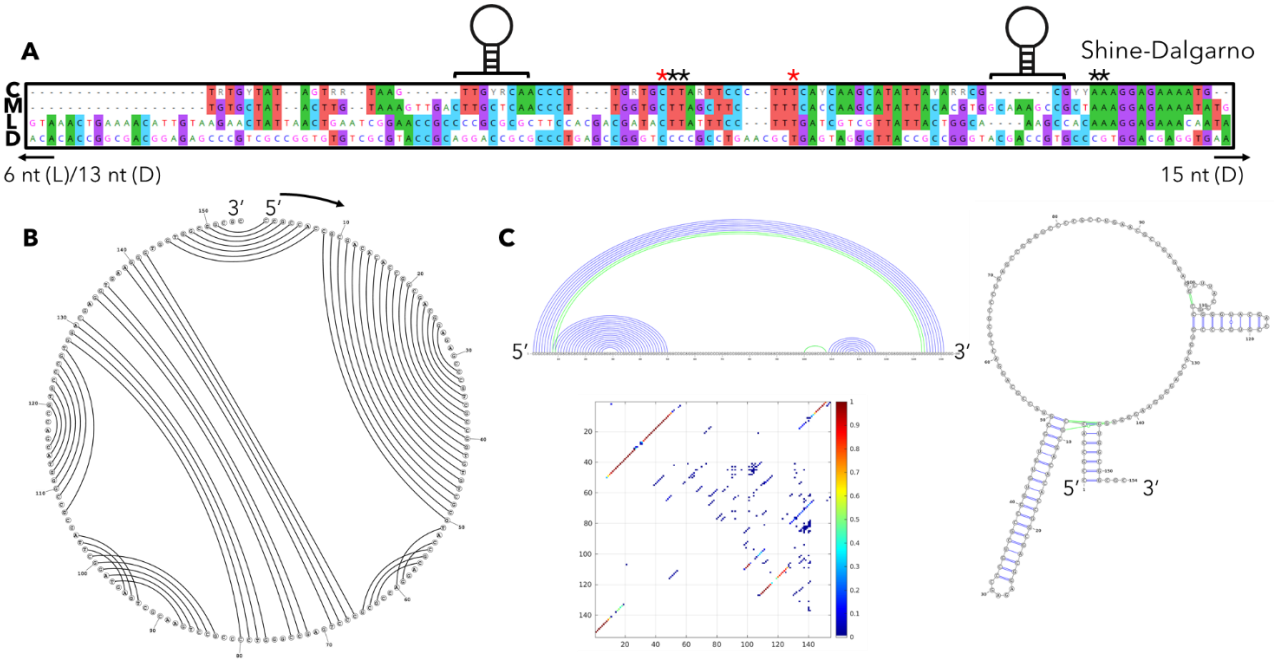


Figure 4.10: *In silico* structure prediction for new *genA* upstream sequence. A – Comparison of the *M. abscessus* (accession number FSGV01000016.1), *L. rhamnosus* (accession number AEYM01000577.1 truncated as in van Vlack et al.)[65] and new *M.* sp. DEM32671 putative riboswitch DNA sequences against the consensus sequence DNA for preQ1-II riboswitches. Alignment was done using ClustalW; consensus sequence was taken from the literature. Conserved bases found to be structurally relevant by Liberman et al. and Kang et al. are marked with asterisks (red for those conserved across all 4 sequences, black for those conserved across all except the *M.* sp. DEM32671 sequence.) [31,67] **B** – ProbKnot was used to estimate pseudoknots in new *M.* sp. DEM32671 sequences (5 iterations, minimum helix length of 2); pseudoknots estimated based on cross-linking of sequence. **C** – SPOT-RNA output of new *M.* sp. DEM32671 *genA* upstream region, showing arc-representation of RNA secondary structure (canonical base pairs in blue arcs and non-nested/non-canonical base pairs in green), predicted secondary structure folding (canonical base pairs in blue arcs and non-nested/non-canonical base pairs in green), base-pair probability plot (upper triangle uses ensemble of 5 models for prediction, lower triangle uses baseline SPOT-RNA model).

Secondly, an error had been made when designing the *L. rhamnosus* construct using the information provided by van Vlack et al.;[65] when the DNA sequence was altered, the correct Shine-Dalgarno sequence was apparent upon prediction with the Salis Lab RBS calculator **(Figure 4.11)**. Fixing the sequence improved predicted translation initiation rate by 343-fold.

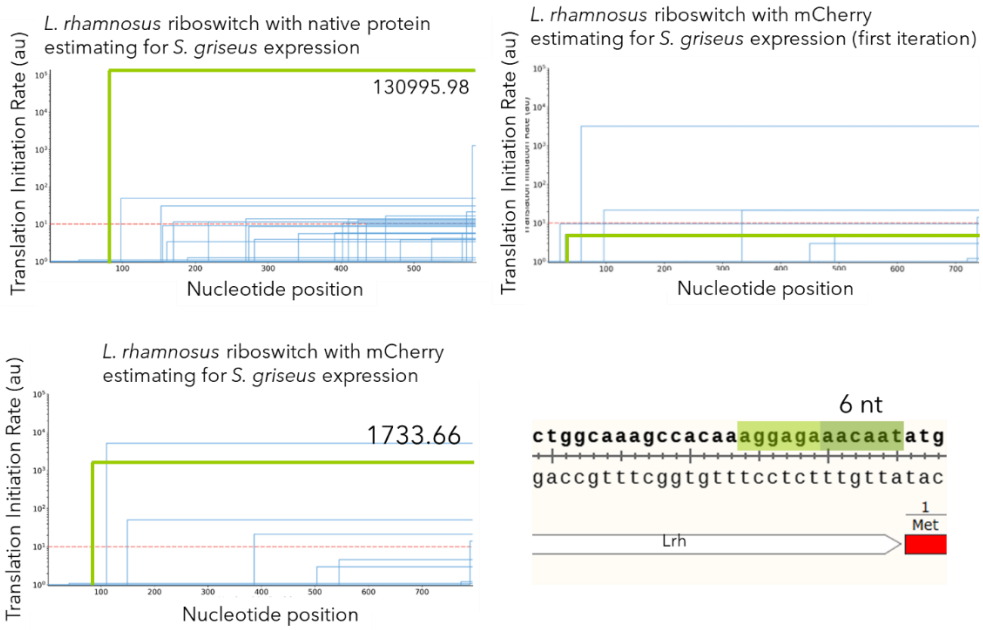


Figure 4.11: *In silico* estimation of *L. rhamnosus riboswitch* start codon based on the Salis Lab Ribosome Binding Site Calculator.[73] The translation initiation rate upstream of the native ORF in the *L. rhamnosus* genome was first determined, and once orientation was confirmed to be correct, the sequence was placed upstream of mCherry to check that translation would be likely to occur. A comparison was carried out with the reversed sequence which showed very low translation initiation rate in comparison (343.3-fold increase). The correct ribosome binding site was also able to be determined by eye based on a 5′ – AGGAGA – 3′ consensus sequence.

pTE1332-genA-full and pTE1332-Lrh-full were constructed using the new information from **Figure 4.9** and **Figure 4.11**, and introduced into *S. coelicolor* M1146. *S. coelicolor* M1146 pTE1332-Lrh-trunc was retained as a negative control, as the sequence between the promoter and coding sequence had been identified to be non-functional in earlier experiments **(Figure 4.8)**. The strains were induced at 0 hours, cultured (larger scale, 50 mL media), and the fluorescence measured at 48 hours **(Figure 4.12)**. In *S. coelicolor* M1146 carrying

pTE1332-genA-full, no significant decrease of fluorescence upon addition of preQ1 was observed. It was therefore determined that the *genA* upstream region was unlikely to represent a preQ1-responsive region. Addition of preQ1 appeared to increase fluorescence by 1.12-, 1.44- and 1.16-fold in the 1 μM, 10 μM and 100 μM preQ1 conditions, respectively, when comparing to the uninduced condition. An unpaired t-test was carried out and the change in fluorescence was determined not to be statistically significant (p values of 0.4306, 0.0819, 0.2599 in the 1 μM, 10 μM and 100 μM preQ1 conditions respectively comparing to the uninduced condition). Additionally, testing between the induced conditions did not yield changes which were statistically significant (p = 0.1762 when comparing induction between 1 μM and 10 μM preQ1 induction, 0.2241 when comparing induction between 10 μM and 100 μM preQ1 induction). The riboswitch from *L. rhamnosus* did not show any significant increase in fluorescence from the fluorescence of the negative control, aside from the condition with 100 μM preQ1 (p = 0.0103).

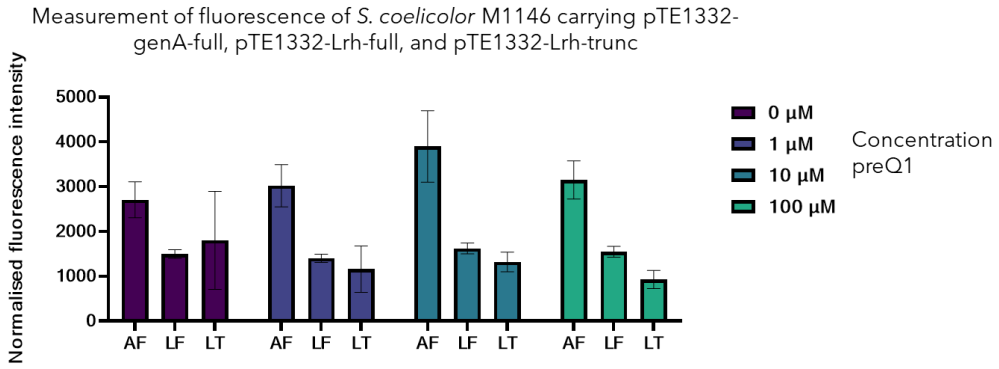


Figure 4.12: Measurement of normalised fluorescence using full sequences of *genA* upstream region (pTE1332-genA-full; shorthand AF), *L. rhamnosus* preQ1-II riboswitch (pTE1332-Lrh-full; shorthand LF) and negative control truncated *L. rhamnosus* preQ1-II riboswitch (previously shown to be non-functional; pTE1332-Lrh-trunc; shorthand LT) coupled to mCherry. *S. coelicolor* strains were induced at 0 hour with 1 μM, 10 μM, 100 μM preQ1 and fluorescence measured after 48 hours of growth in TSB medium.

Of the intergenic regions of Cluster 24, the intergenic region upstream of *genA* appeared the most likely to represent a riboswitch. This was due to a low

sequence similarity to known preQ1-II riboswitch consensus sequences.[30] In comparison, no other intergenic sequence showed any sequence match to the preQ1-I or preQ1-II riboswitch consensus sequences. We posited that if a preQ1-responsive region was identified upstream of biosynthesis-related genes, it may represent a region outside of the currently known consensus sequence of preQ1-II riboswitches that could be further investigated. Unfortunately, it appears the biosynthesis from 7,8-dihydroneopterin 3'triphosphate to $preQ_0$ in Cluster 24 is not regulated by presence of preQ1 as previously described,[30,75] because the change in fluorescence is not statistically significant **(Figure 4.12)**. It is likely that preQ1 addition does not have an effect on downstream expression of *genA*, *genF* and *genG*. However, the *genA* RBS does appear to be functional in *S. coelicolor*, though fluorescence measurements remain low and suggest it may be a weak RBS in this strain. A clearer picture of the cluster regulation may be established with RT-qPCR experiments determining the effect of increasing concentrations of preQ1 on gene expression. With the difficulties of *in silico* prediction of pseudoknot structures, this would be the next step in establishing the role of queuosine biosynthesis in Cluster 24, if it has any. A different approach was recently described by Balaratnam et al. on using preQ1-derived probes for identification of RNA-binding sites using ChemCLIP (Chemical Cross-Linking and Isolation by Pull-down).[76] In this technique, a purification tag is cross-linked to preQ1, which binds RNA-specifically. The RNA which has been bound by preQ1 can then be purified and the sequence identified through PCR amplification. While this method was used to identify preQ1-I aptamers in total human RNA, it could be applicable to identifying whether preQ1 binding sites exist within or outside of the Cluster 24 region in *M.* sp. DEM32671.

While the region upstream of *genA* did not appear to be responsive to preQ1 in these experiments, expression was still able to occur despite the complex secondary structure presence in this intergenic region. The predicted secondary structure using ProbKnot and SPOT-RNA both show a hairpin of length 42 bp, which we deemed likely to be a transcriptional terminator if not

a preQ1-responsive region. This, however, would have inhibited mCherry expression entirely. The HiTES (High Throughput Elicitor Screening) platform could be used to establish whether any other small molecules have an inducing effect on BGC regulation, not limited only to preQ1.[77]

Conversely, the riboswitch from *L. rhamnosus* did not appear functional even without the presence of inducer. This riboswitch was chosen due to good performance in Actinobacteria *Mycobacterium smegmatis* seen by van Vlack et al. (2017).[65] Additionally, the structure has previously been determined.[31] The preQ1-II riboswitch functions to generate an 'OFF' state upon addition of inducer, and so it is unlikely that regulation is tight enough to remain in an 'OFF' state when no inducer is present. Through experiments described later in this work to characterise the riboswitch from *M. abscessus* subsp. *abscessus*, we were able to determine that preQ1 can be taken up by *S. coelicolor* and cause an intracellular effect; therefore, the dysfunction with the *L. rhamnosus* riboswitch candidate is likely a sequence-specific issue. After further investigating this, it appeared there was some discrepancy with the sequence sourced from van Vlack et al. (2017; 83 bp), and the sequence deposited on the Rfam database (Accession: AEYM01000577.1; 125 bp). This is likely to impact on the folding of this sequence, despite the sequence yielding the main secondary structure being present and unaltered (P2/P3/P4 stem). Both testing this full 125 bp sequence and truncating this sequence again to test the Shine-Dalgarno sequence without secondary structure would be the likeliest path forward to establishing conditions for this riboswitch to be functional in *S. coelicolor*.

4.3.4. The preQ1-II riboswitch from *M. abscessus* subsp. *abscessus* functions in *S. coelicolor*

As the truncated *M. abscessus* riboswitch showed fluorescence prior to the other candidates, the full *M. abscessus* riboswitch was next tested **(Figure 4.13, Figure S4.4**, **Table S4.3;** larger scale, 50 mL media). The strains in this experiment were induced at two timepoints: 0 h **(Figure 4.13)** and 24 h (**Figure S4.4**). As before, the pTE1332-Lrh-trunc from the first test was retained as a negative control for this experiment. From these results, induction with preQ1 appears to have a dampening effect on fluorescence but is not titratable like in some other inducible systems (for example, tetracycline inducible transcription factors). This goes against previous results in mycobacteria, where preQ1 induction gave a titratable response.[65] At 48 hours, after inducing the cultures at 0 hours, it is possible to see approximately a 36.73% decrease from 0 μM to 1 μM preQ1, and a 9.72% decrease from 1 μM to 10 μM preQ1. In comparison, after inducing at 24 hours, a 27.71% decrease from 0 μM to 1 μM preQ1, and a 9.49% decrease from 1 μM to 10 μM preQ1 is observed. From this, induction at 0h appears to provide a stronger dampening effect on fluorescence after 48h growth than induction at 24h (comparison of 0h/24h induction with 1 μM preQ1 at 48h total growth, $p = 0.0026$; comparison of 0h/24h induction with 10 μM preQ1 at 48h total growth, $p = 0.0034$). Fluorescence of the uninduced strain drops by 72h, to be comparable to the flasks 'suppressed' by preQ1. As expected, strains carrying pTE1332-Lrh-trunc do not fluoresce, nor do they show a response to preQ1.

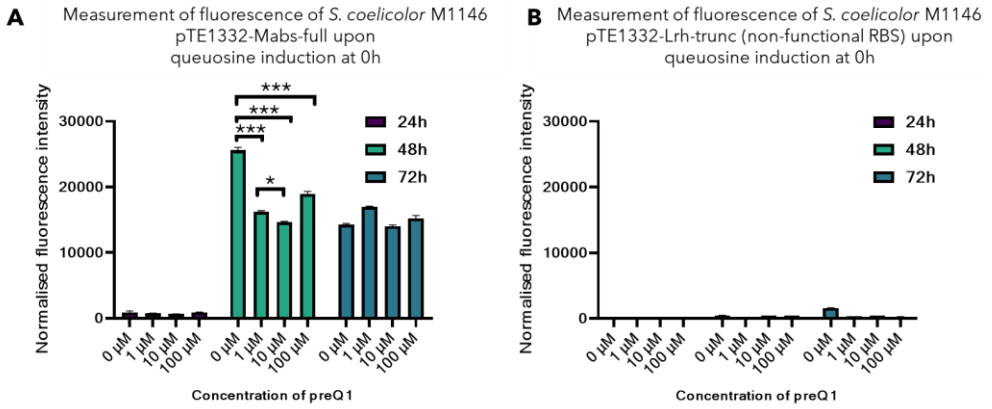


Figure 4.13: Normalised fluorescence of *S. coelicolor* M1146 pTE1332-Mabs-full and negative control pTE1332-Lrh-trunc under preQ1 induction of three different concentrations. Induction with preQ1 was done at 0 h growth. A – Normalised fluorescence of *S. coelicolor* M1146 pTE1332-Mabs-full induced with preQ1; **B** – Normalised fluorescence of *S. coelicolor* M1146 pTE1332-Lrh-trunc (negative control) induced with preQ1. Statistically significant changes between uninduced and induced conditions are plotted: * denotes statistical significance $p < 0.05$; ** denotes statistical significance $p < 0.001$; *** denotes statistical significance $p < 0.0001$.

To ensure addition of preQ1 did not have a deleterious effect on the strains, the growth of each (measured by total protein; mg/mL) was also plotted **(Figure 4.14)**. While comparing the growth of induced samples to the uninduced samples often yielded slight differences between the two, no major increase or decrease could be seen between inducing at 0 hours and 24 hours of growth, nor when preQ1 concentration was increased within these subgroups. It remained clear that preQ1 did not have generally negative impact on growth. In some cases, samples with preQ1 appeared to grow slightly better than without, though it remains unlikely that this nucleoside has any huge cellular benefit for *Streptomyces*. Between the strains carrying pTE1332-Mabs-full and pTE1332-Lrh-trunc, there appears to be a difference in growth: strains carrying pTE1332-Lrh-trunc in general have an increased biomass in comparison to strains carrying pTE1332-Mabs-full after 24h. Though unlikely (as it is only expressed as a single copy, as the vector is integrated into the *S. coelicolor* genome), the expression of mCherry may be

slightly detrimental to the growth of the strain. Comparison with a positive control pTE1332-Mabs-trunc (which should produce a fluorescent signal at all times regardless of presence of preQ1, and would maintain the same strength of ribosome binding site) would allow for determination of possible toxicity in this case.

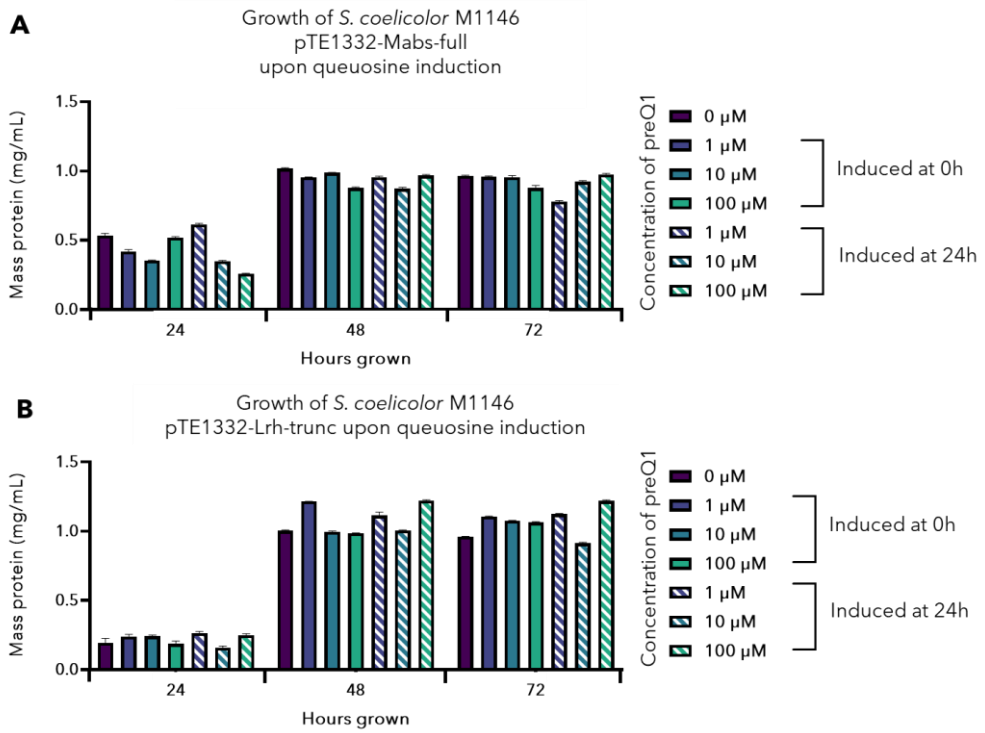


Figure 4.14: Growth of *S. coelicolor* M1146 pTE1332-Mabs-full and negative control pTE1332-Lrh-trunc under different induction conditions. Induction with three different concentrations of preQ1 (1 µM, 10 µM, 100 µM) at two different timepoints (0h, 24h) was tested. Samples induced at 24 hr are marked with diagonal lines. **A** – Growth of *S. coelicolor* M1146 pTE1332-Mabs-full induced with preQ1; **B** - Growth of *S. coelicolor* M1146 pTE1332-Lrh-trunc (negative control) induced with preQ1.

The riboswitch from *M. abscessus* subsp. *abscessus* appears to be functional and gives the expected reduction in fluorescence highlighted by previous studies for this class of riboswitches.[65,75] However, this still requires further fine-tuning to be useful for total repression/control of downstream gene expression. The strength of the promoter coupled to the riboswitch is an important factor

which will impact greatly on riboswitch performance, as this provides another control on the levels of transcription. In this work, *ermEp1* was selected as a driver of transcription of mCherry due to being a medium-strength promoter used in *Streptomyces* routinely. By starting with a medium level of expression, we hoped to be able to tune the promoter strength upstream of the *M. abscessus* subsp. *abscessus* preQI-II riboswitch by replacing with one of many stronger or weaker promoters available, dependent on these first results.[78] It may be that a greater repression of fluorescence could be seen if a weaker promoter was tested: if a lower baseline of gene expression is there to begin with, total repression may be possible upon addition of preQ1. Additionally, recent experiments in our laboratory with alternative reporter genes have suggested that the β-glucuronidase product of *gusA* outperforms mCherry in similar sensor measurements (Erik Hanko, personal communication, unpublished). The product of *gusA* is a well-established fluorimetric/colorimetric reporter for *Streptomyces* when used with specific substrates.[79] It is sensitive, with a low background in uninduced conditions due to the high specificity of the enzyme action. Moreover, measurement of riboswitch-mediated dampening of expression may be challenging in this work due to mCherry's long half-life. Early transcription could contribute to lingering fluorescence even when repression is occurring, and so use of an alternative reporter such as *gusA* may provide a better option.[80]

Once thoroughly optimised, this riboswitch could be used for the design of genetic circuits. *S. coelicolor* currently does not have well characterised and applicable genetic circuits for fine-tuning of gene cluster expression. Providing more ligand-inducible systems may make *S. coelicolor* a more attractive heterologous host for BGC expression.[81] The preQ1 class of riboswitches is well-suited to *S. coelicolor*; as it most likely lacks preQ1 biosynthesis genes, there is no intracellular production of preQ1 which could impact on function. There is one other option for riboswitches used and reported in this organism, the theophylline riboswitch.[82] This functions to have an 'ON state' with ligand

addition, and so establishing a riboswitch with the opposite function (e.g., an 'OFF state' with ligand addition) would complement this well.

4.3.5. Deletion of key queuosine biosynthesis genes has a positive impact on antibiotic production from *Streptomyces* strains carrying Cluster 24

Finally, we were interested in the impact that removal of queuosine biosynthesis genes could have on gentamicin production. As previously described, Cluster 24 from M. sp. DEM32671 is predicted to produce an aminoglycoside, gentamicin.[44] Homologues for *genA*, *genF*, and *genG* were chosen for knock-out as these were the first genes in the pathway which were specific to queuosine biosynthesis. *folE* (GTP cyclohydrolase 1), carrying out the first step in the pathway, has been shown to additionally be involved in biosynthesis of the pterin moiety of folic acid in both *Streptomyces tubercidicus* and *E. coli*. As folates are essential for cell growth, it was therefore deemed a risky target which may impact heavily on the strain's metabolic fitness.[83,84] *genA*, *genF*, and *genG* mediate the steps from 7,8-dihydroneopterin 3'triphosphate to $preQ_0$, and so the pathway would be disrupted after the first conversion of GTP to 7,8-dihydroneopterin 3'triphosphate **(Figure 4.4)**. Secondly, as these genes were colocalised, a single modification with Cas9 could achieve knock-out of all three genes simultaneously. By choosing to eliminate all but one of the queuosine pathway precursors, we hoped to remove the potential regulatory effect on the gene cluster to see the largest possible difference in antibiotic production phenotype.

A CRISPR-Cas9 plasmid based on pCM4.4 was constructed to include homology arms removing these three genes (C24 locations: 40,611 bp → 41,610 bp, 43,088 bp → 44,087 bp; **Figure S4.5**). Correct knock-out was determined by colony PCR and confirmed with Sanger sequencing **(Figure S4.6)**. Two *Streptomyces* strains carrying C24 had *genA*, *genF*, and *genG* knocked out: i) *S.*

coelicolor M1146 C24 2R, which we previously generated through random mutagenesis to have increased tolerance to gentamicin, and ii) *S. coelicolor* M1146 C24 2R + constitutive promoter cassette, which is identical to *S. coelicolor* M1146 C24 2R except with an *ermEp1-fd-sp44* promoter cassette inserted between *gmrA* and *genS1* homologues.[44,85] In solid and liquid culture, both *ΔgenA ΔgenF ΔgenG* knock-out strains (*S. coelicolor* M1146 C24 2R *ΔgenAΔgenFΔgenG*, *S. coelicolor* M1146 C24 2R + promoter cassette *ΔgenAΔgenFΔgenG*) retained the same growth phenotype as the parental strains, typical of *S. coelicolor* (proper spore pigment formation on solid medium and dispersed growth in liquid medium).

After carrying out liquid cultures in TSB/R2YE production medium and antimicrobial bioassay of culture extracts, it appeared that there was some positive impact on bioactivity upon knockout of *genA*, *genF*, and *genG* **(Figure 4.15)**. One replicate of the *S. coelicolor* M1146 C24 2R *ΔgenAΔgenFΔgenG* and two replicates of the *S. coelicolor* M1146 C24 2R + promoter cassette *ΔgenAΔgenFΔgenG* strains appeared to show bioactivity, while none of the parental strains nor the negative control showed visible antibiotic activity. We previously reported that while we were able to see a yield improvement upon insertion of the promoter cassette to the strain, production appeared to be less reproducible than we would like.[85] The variability seen here again shows the challenges of heterologous expression, particularly when dealing with low production titres where variability between culture samples may be more obvious. As the bioactivity did not appear to be hugely increased across all three replicates in comparison to the parental strain, it cannot be stated that this is a major effect on aminoglycoside production. However, it does suggest that perhaps these genes are important in some capacity, and the evidence for this could be bolstered with further experiments. We were unable to carry out LC-MS experiments to determine whether the product profile (in terms of gentamicin congener production) was changed upon deletion of *genA*, *genF*, and *genG*. However, this would be invaluable towards understanding the function of the queuosine biosynthesis genes in the context of the Cluster 24,

and perhaps in the wider context of aminoglycoside clusters from *Micromonospora* as a whole.

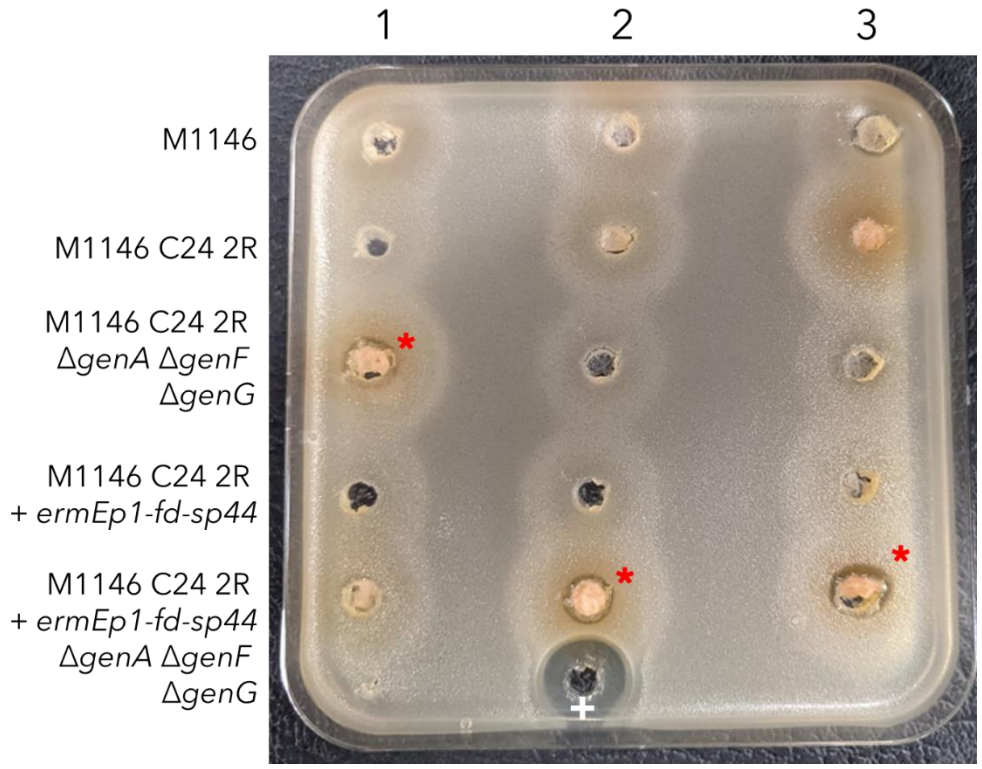


Figure 4.15: Antimicrobial bioassay of Δ*genA* Δ*genF* Δ*genG* knock-out strains. 50 µL of 25X concentrated culture supernatant was added to each well of an LBA + *B. subtilis* plate. + = 5 µg gentamicin sulfate. Samples with visible bioactivity are marked with asterisk (*).

Additionally, it remains unclear how well the queuosine biosynthesis genes are expressed, their role in the heterologous host, or if their expression may be linked to certain conditions such as nutrient depletion or at certain stages in the growth cycle. Understanding when these genes are expressed (if at all) could give a key insight into the role of this co-clustering in aminoglycoside biosynthesis. Temporal expression is common in bacteria, and the rRNA operons in *S. coelicolor* follow this; in Kim et al.'s work (2007), 3 of the 6 rRNA operons were shown to be expressed early in the growth cycle, while the other

3 were shown to express later in the growth cycle when the growth cycle moves towards sporulation.[86]

A further expansion of the work towards carrying out RT-qPCR experiments on the newly created queuosine biosynthesis gene mutant strain to identify the effects on gene cluster expression could be of additional interest. An experiment where both the Δ*genA* Δ*genF* Δ*genG* knock-out strains and parental strains are cultured in the presence of preQ1 would be of interest, as this would bypass the original biosynthesis pathway. From this, it should become clearer what the effect on Cluster 24 gene expression is, and therefore how queuosine biosynthesis may be associated functionally with aminoglycoside biosynthesis.

4.4. Conclusions

In this work we investigated the relationship between queuosine biosynthesis genes and aminoglycoside gene cluster Cluster 24 from *M.* sp. DEM32671. We identified that queuosine biosynthesis genes appears to co-localise within certain aminoglycoside gene clusters from the genus *Micromonospora*. While Cluster 24 appeared to have a higher percentage of queuosine-associated amino acids present than the mean percentage of these amino acids in the rest of the genome, the pathway for the biosynthesis of queuosine production remains likely to be incomplete in *M.* sp. DEM32671. Our secondary hypothesis was that queuosine precursors may have a role in gene cluster regulation, through possible presence of preQ1 riboswitch(es) in Cluster 24. While *in silico* analysis suggested the presence of secondary structure upstream of genes with similarity to preQ1-controlled mRNA sequences, this region was shown to be unresponsive to preQ1 through fluorescence assay in *S. coelicolor*. We additionally tested two alternative candidate preQ1-II riboswitches in *S. coelicolor*. The riboswitch from *M. abscessus* subsp. *abscessus* appeared to be functional but requires further fine-tuning for use in applications such as genetic circuits. However, if the function is further confirmed, it would provide an orthogonal repressive riboswitch, which the heterologous host, *S.*

coelicolor, is currently lacking. Finally, we investigated whether deletion of key queuosine biosynthesis genes would have an impact on production of antibiotic from *Streptomyces* strains carrying aminoglycoside cluster Cluster 24. We identified that knockout of three queuosine biosynthesis genes was associated with an increase in bioactivity from the strains carrying Cluster 24, though as production titres are low there remains huge variability observed between culture samples. As we were unable to determine the product output of the knockout strains, it remains unclear whether these genes have any impact on aminoglycoside biosynthesis aside from this small difference in yield (for example, in the ratios of congeners produced).

Overall, the biological relevance/selective advantage – if any – of queuosine biosynthesis being so closely linked to aminoglycoside biosynthesis remains unknown. Further experiments investigating the transcription of queuosine biosynthesis genes, or the transcription of aminoglycoside biosynthesis genes from Cluster 24 in response to preQ1 induction could shed light on whether this nucleobase is linked to gentamicin BGC regulation, and further aid in designing improved production strategies for aminoglycosides in heterologous hosts.

4.5. Materials and Methods

4.5.1. Strains and plasmids

All bacterial strains used and generated in this work are listed in **Table 4.3.**

Table 4.3: Bacterial strains and plasmids used in this work.

Bacterial strain	Genotype	Use	Reference
Escherichia coli NEB5α	*fhuA2 (argF-lacZ)U169 phoA glnV44 80 (lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	Cloning	[87]
E. coli ET12567 pUZ8002	*Fdam13::Tn9 dcm6 hsdM hsdR zjj202::Tn10 recF143 galK2 galT22 ara14 lacY1 xyl5 leuB6 thi1 tonA31 rpsL136 hisG4 tsx78 mtl1 glnV44* pUZ8002	Conjugation of plasmids into *Streptomyces*	[88]
Bacillus subtilis 168	*trpC2*	Indicator strain for bioassays	[89,90]
Streptomyces coelicolor M1146	*S. coelicolor* M145 *Δact Δred Δcda Δcpk*	Expression of Cluster 24-derived plasmids	[91]
Streptomyces coelicolor M1146 C24 2R	*S. coelicolor* M145 *Δact Δred Δcda Δcpk* C24 (92 mutations, detailed in **Chapter 2**)	Modification of Cluster 24	[44]
Streptomyces coelicolor M1146 C24 2R + *ermEp1-fd-sp44*	*S. coelicolor* M145 *Δact Δred Δcda Δcpk* C24(*ermEp1-fd-sp44*)	Modification of Cluster 24	This work

Plasmid	Backbone	Insert	Reference
pTE1332	pSET152 (replicative, Apra^R)	*fd-ermEp1*-rbs-mCherry-*fd*	Erik Hanko (unpublished)
pTE1332-genA-full	pSET152 (replicative, Apra^R)	*fd-ermEp1*-region upstream of *genA* (C24)-mCherry-*fd*	This work
pTE1332-genA-trunc	pSET152 (replicative, Apra^R)	*fd-ermEp1*-truncated region upstream of *genA* (C24)-mCherry-fd	This work
pTE1332-Mabs-full	pSET152 (replicative, Apra^R)	*fd-ermEp1*-full *M. abscessus* preQ1-II riboswitch-mCherry-*fd*	This work
pTE1332-Mabs-trunc	pSET152 (replicative, Apra^R)	*fd-ermEp1*-truncated *M. abscessus* preQ1-II riboswitch-mCherry-*fd*	This work
pTE1332-Lrh-full	pSET152 (replicative, Apra^R)	*fd-ermEp1*-full *L. rhamnosus* preQ1-II riboswitch-mCherry-*fd*	This work
pTE1332-Lrh-trunc	pSET152 (replicative, Apra^R)	*fd-ermEp1*-truncated *L. rhamnosus* preQ1-II riboswitch-mCherry-*fd*	This work
pCM4.4	pCRISPomyces 2.0 (replicative, Apra^R)	UNS2-*fd-ermE**-spCas9-*fd*-gapdhp-gRNA-tracr-ori-UNS6	[92]

pCM4.4Δ*genAgenFgenG*	pCM4.4 (replicative, Apra^R)	UNS6-5′ homologous arm-3′ homologous arm-UNS7	This work

4.5.2. Media and cultivation conditions

For generation of spore suspensions, *S. coelicolor* was grown on SFM agar (2% agar, 2% mannitol, 2% soy flour) with 50 μg/mL apramycin.

For *S. coelicolor* liquid cultures, either 15 mL of media was used in a 100 mL flask or 50 mL of media was used in a 250 mL flask (indicated per experiment). 10 mM diameter coil springs were inserted to ensure dispersed growth. Cultures were grown in tryptone soya broth (TSB; 1.7% pancreatic digest of casein, 0.3% enzymatic digest of soya bean, 0.5% sodium chloride, 0.25% K_2HPO_4, 0.5% glucose; obtained as pre-mixed powder from Oxoid cat. no. CM0129) for 48 hours at 180 rpm and 30°C. Liquid cultures were grown without antibiotics. preQ1 dihydrochloride (Sigma-Aldrich; cat. no. SML0807) was added at a concentration of 1, 10, or 100 μM where investigating riboswitch-based repression of fluorescence. For testing the impact of deletion of *genA*, *genF*, and *genG*, production cultures were inoculated 1/50 from TSB seed cultures grown as described above (without preQ1) addition. Production medium was TSB/R2YE medium, prepared with a 1:1 ratio of unautoclaved TSB and previously autoclaved R2YE Media A (10.3% sucrose, 1% glucose, 1.12% $MgCl_2.6H_2O$, 0.025% K_2SO_4, 0.01% Difco casamino acids, 0.5% Difco yeast extract).

E. coli NEB5α, *E. coli* ET12567 pUZ8002 and *B. subtilis* were cultivated in Luria-Bertani Miller (LB) broth (Formedium; 1% NaCl, 1% tryptone, 0.5% yeast extract) at 37°C and 180 rpm. Luria-Bertani Miller (LB) agar (Formedium; as above but with 1.5% agar) was used as a solid medium.

4.5.3. Constructing heatmaps to estimate regions enriched with specific amino acids

The percentage of specific amino acids in each ORF of the *M.* sp. DEM32671 genome was calculated separately in a Microsoft Excel Workbook, as well as the percentage of summed Asp+Asn+His+Tyr. The mean of each amino acid was taken as the average across the entire genome as a point of comparison for Cluster 24 genes. GraphPad Prism v9 for Windows (GraphPad Software, www.graphpad.com) was used to generate separate heatmaps for each query amino acid.

4.5.4. Prediction of RNA secondary structure

ProbKnot (5 iterations and a minimum helix length of 2) and SPOT-RNA were used to investigate secondary structures in riboswitch candidates.[69,93] Clustal X was used to align query sequences to preQ1-II consensus sequence, which was taken from Weinberg et al. (2007).[66,94]

4.5.5. Construction of pTE1332 variants

pTE1332, consisting of pSET152 with an insert containing codon-optimized mCherry (codon-optimization carried out by Schlimpert et al., sequence kindly gifted by Susan Schlimpert)[39] under control of the ermEp1 promoter, was kindly gifted for this work by Erik Hanko. For replacement of the pTE1332 ribosome binding site with riboswitch candidates, PCR primers were designed with overhangs to generate a linearised plasmid, with the riboswitch region present on the 5′-end of the linear DNA fragment **(Table S4.2;** Table S4.3). The amplification was done using PrimeSTAR Max polymerase (Takara; cat. no. R045A) with the following cycle: initial denaturation 98°C for 2 min; amplification (35X cycles) 98°C for 10s, 55°C for 15s, 72°C for 45s. PCR products were checked via agarose gel electrophoresis before purification with the QIAQuick PCR Purification Kit (Qiagen; cat. no. 28104), treating with T4

Polynucleotide Kinase (NEB; cat. no. M0201S) according to the manufacturer's instructions, before ligation for 18 hours (T4 DNA Ligase, NEB; cat. no. M0202S) and transformation into chemically competent *E. coli* NEB® 5α (NEB; cat. no. C2987H). Successful construction was verified through colony PCR with OneTaq® DNA polymerase (NEB; cat. no. M0482S) and the following cycle: initial denaturation 94°C for 2 min; amplification (35X cycles) 94°C for 30 s, 55°C for 30 s, 68°C for 1.5 min; further extension 68°C for 5 min. Colony PCR was checked via agarose gel electrophoresis for presence of correct product, before confirmation with Sanger sequencing.

4.5.6. Construction of pCM4.4ΔgenAFG

Protospacer was inserted into 'empty' pCM4.4,75to create pCM4.4ΔgenAgenFgenG-p using annealed oligos (genAFGpfw 5′ – **ACGC**GTTGGCCATCGTCTCGTAGG – 3′, genAFGprev 5′ – **AAAC**CCTACGAGACGATGGCCAAC –3′) with overhangs compatible with Golden gate assembly. Golden gate was performed with BbsI (NEB) and T4 DNA Ligase (NEB) with the following cycle: 10X cycles of 37°C and 16°C alternating (10 min each), then for heat inactivation of the ligase 50°C (5 min), then for heat inactivation of BbsI 65°C (20 min). After transformation using chemically competent *E. coli* NEB® 5α (NEB, cat. no. C2987H) following standard manufacturer's protocol, successful insertion (and loss of *lacZ* gene) was determined by blue-white screening and further Sanger sequencing.

Amplification of fragments for HiFi assembly (NEB; cat. no. E5520) to insert the homology arms for repair template (creating pCM4.4ΔgenAFG; **Figure S4.5**) was carried out using PrimeSTAR Max DNA polymerase (Takara; cat. no. R045A) with the following cycle: initial denaturation 98°C for 2 min; amplification (35X cycles) 98°C for 10s, 55°C for 15s, 72°C for 45s. Assembly was carried out as a 3-part reaction: Fragment 1 (*aac3(IV)-pSG5 rep-oriT-traJ*-UNS2) was amplified with pCM4.4 HA site UNS7 fw and dLancI-II 1 rev; Fragment 2 (UNS2-*fd-ermE**-spCas9-*fd*-gapdhp-protospacer-gRNA-tracr-*ori*-

UNS6) was amplified with Cas9 UNS2 fw and pCM4.4 HA site UNS6 rev. Fragments 1 and 2 used pCM4.4Δ*genAgenFgen*G-p as template. Fragment 3, consisting of the left and right homology arms flanked by UNS6 and UNS7 sequences,[95] was ordered as a synthetic DNA fragment from Thermo Fisher Scientific (GeneArt Gene Synthesis). To generate more DNA, Fragment 3 was amplified with UNS6F and UNS7R primers with PrimeSTAR Max DNA polymerase (Takara; cat. no. R045A) and the following cycle: initial denaturation 98°C for 2 min; amplification (35X cycles) 98°C for 10s, 55°C for 15s, 72°C for 20s. (see Primer Sequences in **Table S4.2)**. Agarose gel electrophoresis was used to confirm correct amplification **(Figure S4.6A)**. Fragments were treated with DpnI (NEB) for 18 hours (37°C) prior to assembly. Assembly was carried out according to the manufacturer's protocol with NEBuilder® HiFi DNA Assembly Master Mix (NEB, cat. no. E5520S) and transformed into chemically-competent *E. coli* NEB® 5α (NEB, cat. no. C2987H) following standard manufacturer's protocol. After purification of plasmid using a QIAprep Spin Miniprep Kit (Qiagen, cat. no. 27106X4), correct construction was verified by Sanger sequencing **(Figure S4.6D**).

4.5.7. Conjugation of plasmids to *S. coelicolor*

Competent cells of *E. coli* ET12567 pUZ8002 were obtained using a standard calcium chloride protocol and transformation carried out with standard heat-shock method.[96] Conjugation from *E. coli* ET12567 pUZ8002 carrying riboswitch variant plasmids or CRISPR plasmids into *S. coelicolor* was carried out as described by Kieser et al. in Practical Streptomyces Genetics.[97] Screening for proper introduction of plasmid was done using by Terra polymerase (Takara; cat. no. 639270) on collected single colonies (initial denaturation 98°C for 2 min; amplification (35X cycles) 98°C for 10s, 68°C for 3 min) using primers detailed in **Table S4.2**.

4.5.8. Measurement of mCherry fluorescence in *S. coelicolor*

1 mL samples were collected from *S. coelicolor* grown in TSB and the supernatant collected by centrifugation before addition of 0.5 volume lysis buffer (100 mM Tris pH 7.0, 1 mM EDTA, 150 mM NaCl). Entire mixture was lysed in lysing matrix E (MP-Bio, 116914050-CF) in FastPrep-24 lysis system (3 cycles of 20 seconds) before centrifugation for 10 minutes. Fluorescence measurement was taken with 150 μL of protein supernatant in black, flat-bottomed Corning plates (cat. no. 3916) and the following conditions in CLARIOstar plate reader (BMG Labtech): Excitation 574 nM, Emission 617 nM, gain 1500. Protein was quantified by carrying out a BSA standard curve (0.14 mg/mL to 0.00875 mg/mL decreasing by half each time) measured at 590 nM in ClarioSTAR plate reader, with protein supernatant being diluted from 1:10 to 1:30 in lysis buffer to ensure it was in the linear range.

4.5.9. Preparation of liquid samples for bioassay

Entire cultures (cells in growth media), frozen at -20°C, were thawed and centrifuged at 7000 *x g* for 5 minutes. An aliquot of culture supernatant was taken into a 50 mL Corning tube before snap-freezing in liquid nitrogen and further freezing at -80°C for one hour. Samples were freeze dried for 22 hours and then resuspended in sterile ddH_2O to the concentration stated for each experiment.

4.5.10. Antibiotic bioassay

Bioassay was carried out against indicator strain *B. subtilis* (grown in LB-Miller broth, at 37°C, at 180 rpm, for 16 hours before subculturing again in the same medium; 1/100 dilution for 3 hours). 50 mL agar inoculated with 50 μL of subcultured *B. subtilis* in 10ml of LB media (OD_{600} = 0.6) was used for each 120 mm square bioassay plate. To analyse levels of antibiotic production in samples from liquid culture, a 13 mM diameter hole was cut into the indicator

plate and 50 μL of concentrated culture extract was added to each. The plate was incubated for 16 hours at 37°C before photographing.

4.6. Acknowledgements

We thank Michael Vockenhuber for his assistance with direction on riboswitch class prediction and Erik Hanko for supplying plasmid pTE1332 and for assistance with fluorescence measurement in *Streptomyces*.

1. I. B. Rogozin *et al.*, "Congruent evolution of different classes of non-coding DNA in prokaryotic genomes," *Nucleic Acids Research*, vol. 30, no. 19, p. 4264, 2002, doi: 10.1093/NAR/GKF549.

2. L. F. Abu-Qatouseh *et al.*, "Identification of differentially expressed small non-protein-coding RNAs in *Staphylococcus aureus* displaying both the normal and the small-colony variant phenotype," *J Mol Med (Berl)*, vol. 88, no. 6, pp. 565–575, 2010, doi: 10.1007/S00109-010-0597-2.

3. C. Pichon and B. Felden, "Small RNA genes expressed from *Staphylococcus aureus* genomic and pathogenicity islands with specific expression among pathogenic strains," *PNAS*, vol. 102, no. 40, pp. 14249–14254, 2005, doi: 10.1073/PNAS.0503838102.

4. R. Barrangou *et al.*, "CRISPR provides acquired resistance against viruses in prokaryotes," *Science (1979)*, vol. 315, no. 5819, pp. 1709–1712, 2007, doi: 10.1126/SCIENCE.1138140/

5. A. Zhang, S. Altuvia, A. Tiwari, L. Argaman, R. Hengge-Aronis, and G. Storz, "The OxyS regulatory RNA represses *rpoS* translation and binds the Hfq (HF-I) protein.," *The EMBO Journal*, vol. 17, no. 20, p. 6061, 1998, doi: 10.1093/EMBOJ/17.20.6061.

6. A. Muffler, D. Fischer, and R. Hengge-Aronis, "The RNA-binding protein HF-I, known as a host factor for phage Qbeta RNA replication, is essential for *rpoS* translation in *Escherichia coli*," *Genes Dev*, vol. 10, no. 9, pp. 1143–1151, 1996, doi: 10.1101/GAD.10.9.1143.

7. D. D. Sledjeski, C. Whitman, and A. Zhang, "Hfq Is Necessary for Regulation by the Untranslated RNA DsrA," *Journal of Bacteriology*, vol. 183, no. 6, p. 1997, 2001, doi: 10.1128/JB.183.6.1997-2005.2001.

8. R. R. Breaker, "Riboswitches and the RNA World," *Cold Spring Harbor Perspectives in Biology*, vol. 4, no. 2, pp. 3566–3567, 2012, doi: 10.1101/CSHPERSPECT.A003566.

9. A. Serganov and E. Nudler, "A Decade of Riboswitches," *Cell*, vol. 152, no. 0, p. 17, 2013, doi: 10.1016/J.CELL.2012.12.024.

10. K. Papenfort, Y. Sun, M. Miyakoshi, C. K. Vanderpool, and J. Vogel, "Small RNA-mediated activation of sugar phosphatase mRNA regulates glucose homeostasis," *Cell*, vol. 153, no. 2, pp. 426–437, 2013, doi: 10.1016/J.CELL.2013.03.003.

11. M. S. Guo, T. B. Updegrove, E. B. Gogol, S. A. Shabalina, C. A. Gross, and G. Storz, "MicL, a new σE-dependent sRNA, combats envelope stress by repressing synthesis of Lpp, the major outer membrane lipoprotein," *Genes Dev*, vol. 28, no. 14, pp. 1620–1634, 2014, doi: 10.1101/GAD.243485.114.

12. A. Serganov and D. J. Patel, "Amino acid recognition and gene regulation by riboswitches," *Biochim Biophys Acta*, vol. 1789, no. 0, p. 592, 2009, doi: 10.1016/J.BBAGRM.2009.07.002.

13. P. J. McCown, K. A. Corbino, S. Stav, M. E. Sherlock, and R. R. Breaker, "Riboswitch diversity and distribution," *RNA*, vol. 23, no. 7, pp. 995–1011, 2017, doi: 10.1261/RNA.061234.117.

14. A. G. Vitreschak, D. A. Rodionov, A. A. Mironov, and M. S. Gelfand, "Regulation of the vitamin B12 metabolism and transport in bacteria by a conserved RNA structural element," *RNA*, vol. 9, no. 9, p. 1084, 2003, doi: 10.1261/RNA.5710303.

15. M. T. Croft, M. Moulin, M. E. Webb, and A. G. Smith, "Thiamine biosynthesis in algae is regulated by riboswitches," *PNAS*, vol. 104, no. 52, pp. 20770–20775, 2007, doi: 10.1073/PNAS.0705786105.

16. S. Stav, R. M. Atilho, G. Mirihana Arachchilage, G. Nguyen, G. Higgs, and R. R. Breaker, "Genome-wide discovery of structured noncoding RNAs in bacteria," *BMC Microbiology*, vol. 19, no. 1, pp. 1–18, 2019, doi: 10.1186/S12866-019-1433-7

17. R. C. Morris and M. S. Elliott, "Queuosine Modification of tRNA: A Case for Convergent Evolution," *Molecular Genetics and Metabolism*, vol. 74, no. 1–2, pp. 147–159, 2001, doi: 10.1006/MGME.2001.3216.

18. H. Kasai, Y. Kuchino, K. Nihei, and S. Nishimura, "Distribution of the modified nucleoside Q and its derivatives in animal and plant transfer RNA's," *Nucleic Acids Res*, vol. 2, no. 10, pp. 1931–1940, 1975, doi: 10.1093/NAR/2.10.1931.

19. J. R. Katze, B. Basile, and J. A. Mccloskey, "Queuine, a modified base incorporated posttranscriptionally into eukaryotic transfer RNA: wide distribution in nature," *Science*, vol. 216, no. 4541, pp. 55–56, 1982, doi: 10.1126/SCIENCE.7063869.

20. R. K. Slany and H. Kersten, "Genes, enzymes and coenzymes of queuosine biosynthesis in procaryotes," *Biochimie*, vol. 76, no. 12, pp. 1178–1182, 1994, doi: 10.1016/0300-9084(94)90047-7.

21. K. A. P. Payne *et al.*, "Epoxyqueuosine Reductase Structure Suggests a Mechanism for Cobalamin-dependent tRNA Modification," *The Journal of Biological Chemistry*, vol. 290, no. 46, p. 27572, 2015, doi: 10.1074/JBC.M115.685693.

22. D. Iwata-Reuyl, "An embarrassment of riches: the enzymology of RNA modification," *Current Opinion in Chemical Biology*, vol. 12, no. 2, pp. 126–133, 2008, doi: 10.1016/J.CBPA.2008.01.041.

23. R. M. McCarty and V. Bandarian, "Biosynthesis of pyrrolopyrimidines," *Bioorganic Chemistry*, vol. 43, pp. 15–25, 2012, doi: 10.1016/J.BIOORG.2012.01.001.

24. J. S. Reader, D. Metzgar, P. Schimmel, and V. de Crécy-Lagard, "Identification of four genes necessary for biosynthesis of the modified nucleoside queuosine," *The Journal of Biological Chemistry*, vol. 279, no. 8, pp. 6280–6285, 2004, doi: 10.1074/JBC.M310858200.

25. R. Gaur and U. Varshney, "Genetic analysis identifies a function for the *queC* (*ybaX*) gene product at an initial step in the queuosine biosynthetic pathway in *Escherichia coli*," *J Bacteriol*, vol. 187, no. 20, pp. 6893–6901, 2005, doi: 10.1128/JB.187.20.6893-6901.2005.

26. R. M. McCarty, Á. Somogyi, and V. Bandarian, "Escherichia coli QueD is a 6-carboxy-5,6,7,8-tetrahydropterin synthase," *Biochemistry*, vol. 48, no. 11, pp. 2301–2303, 2009, doi: 10.1021/BI9001437.

27. R. M. McCarty, Á. Somogyi, G. Lin, N. E. Jacobsen, and V. Bandarian, "The deazapurine biosynthetic pathway revealed: in vitro enzymatic synthesis of PreQ(0) from guanosine 5′-triphosphate in four steps," *Biochemistry*, vol. 48, no. 18, pp. 3847–3852, 2009, doi: 10.1021/BI900400E.

28. A. Roth *et al.*, "A riboswitch selective for the queuosine precursor preQ1 contains an unusually small aptamer domain," *Nature Structural & Molecular Biology 2007 14:4*, vol. 14, no. 4, pp. 308–317, 2007, doi: 10.1038/nsmb1224.

29. M. M. Meyer, A. Roth, S. M. Chervin, G. A. Garcia, and R. R. Breaker, "Confirmation of a second natural preQ1 aptamer class in *Streptococcaceae* bacteria," *RNA*, vol. 14, no. 4, pp. 685–695, 2008, doi: 10.1261/RNA.937308.

30. P. J. McCown, J. J. Liang, Z. Weinberg, and R. R. Breaker, "Structural, functional, and taxonomic diversity of three preQ1 riboswitch classes," *Chemistry & Biology*, vol. 21, no. 7, pp. 880–889, 2014, doi: 10.1016/J.CHEMBIOL.2014.05.015.

31. J. A. Liberman, M. Salim, J. Krucinska, and J. E. Wedekind, "Structure of a class II preQ1 riboswitch reveals ligand recognition by a new fold," *Nature Chemical Biology*, vol. 9, no. 6, pp. 353–355, 2013, doi: 10.1038/NCHEMBIO.1231.

32. D. J. Klein, T. E. Edwards, and A. R. Ferré-D'Amaré, "Cocrystal structure of a class I preQ1 riboswitch reveals a pseudoknot recognizing an essential hypermodified nucleobase," *Nature Structural & Molecular Biology*, vol. 16, no. 3, pp. 343–344, 2009, doi: 10.1038/NSMB.1563.

33. J. L. Jenkins, J. Krucinska, R. M. McCarty, V. Bandarian, and J. E. Wedekind, "Comparison of a PreQ1 riboswitch aptamer in metabolite-bound and free states with implications for gene regulation," *Journal of Biological Chemistry*, vol. 286, no. 28, pp. 24626–24637, 2011, doi: 10.1074/JBC.M111.230375.

34. L. Horbal and A. Luzhetskyy, "Dual control system – A novel scaffolding architecture of an inducible regulatory device for the precise regulation of gene expression," *Metabolic Engineering*, vol. 37, pp. 11–23, 2016, doi: 10.1016/J.YMBEN.2016.03.008.

35. G. S. Hossain, M. Saini, R. Miyake, H. Ling, and M. W. Chang, "Genetic Biosensor Design for Natural Product Biosynthesis in Microorganisms," *Trends in Biotechnology*, vol. 38, no. 7, pp. 797–810, 2020, doi: 10.1016/J.TIBTECH.2020.03.013.

36. C. Zhu, G. Yang, M. Ghulam, L. Li, and F. Qu, "Evolution of multi-functional capillary electrophoresis for high-efficiency selection of aptamers," *Biotechnology Advances*, vol. 37, no. 8, 2019, doi: 10.1016/J.BIOTECHADV.2019.107432.

37. S. Jang *et al.*, "Development of Artificial Riboswitches for Monitoring of Naringenin in Vivo," *ACS Synthetic Biology*, vol. 6, no. 11, pp. 2077–2085, 2017, doi: 10.1021/ACSSYNBIO.7B00128/

38. K. Wang *et al.*, "Multi-Layer Controls of Cas9 Activity Coupled With ATP Synthase Over-Expression for Efficient Genome Editing in *Streptomyces*," *Frontiers in Bioengineering and Biotechnology*, vol. 7, p. 304, 2019, doi: 10.3389/fbioe.2019.00304.

39. S. Schlimpert *et al.*, "Two dynamin-like proteins stabilize FtsZ rings during Streptomyces sporulation," *PNAS*, vol. 114, no. 30, pp. E6176–E6183, 2017, doi: 10.1073/PNAS.1704612114/

40. Y. H. Ban *et al.*, "Complete reconstitution of the diverse pathways of gentamicin B biosynthesis," *Nature Chemical Biology*, vol. 15, pp. 295–303, 2019, doi: 10.1038/s41589-018-0203-4.

41. S. Li *et al.*, "Mechanistic Insights into Dideoxygenation in Gentamicin Biosynthesis," *ACS Catalysis*, vol. 11, no. 19, pp. 12274–12283, 2021, doi: doi.org/10.1021/acscatal.1c03508.

42. S. Li *et al.*, "Methyltransferases of gentamicin biosynthesis," *PNAS*, vol. 115, no. 6, pp. 1340–1345, 2018, doi: 10.1073/pnas.1711603115.

43. M. Yang, M. K. Derbyshire, R. A. Yamashita, and A. Marchler-Bauer, "NCBI's Conserved Domain Database and Tools for Protein Domain Analysis," *Current Protocols in Bioinformatics*, vol. 69, no. 1, p. e90, 2020, doi: 10.1002/CPBI.90.

44. K. V. Baker et al., "Chapter 2 - Establishing M. sp. DEM32671 as a producer of aminoglycosides through heterologous expression in Streptomyces spp.," PhD Thesis Chapter, University of Manchester, Manchester, 2022.

45. F. Kudo and T. Eguchi, "Biosynthetic genes for aminoglycoside antibiotics," *The Journal of Antibiotics*, vol. 62, no. 9, pp. 471–481, 2009, doi: 10.1038/ja.2009.76.

46. R. P. Frohardt, T. H. Haskell, J. Ehrlich, and M. Penner Knudsen, "Antibiotic and methods for obtaining same," 1959

47. K. F. Koch, F. A. Davis, and J. A. Rhoades, "Nebramycin: separation of the complex and identification of factors 4, 5, and 5'," *J Antibiot (Tokyo)*, vol. 26, no. 12, pp. 745–751, 1973, doi: 10.7164/ANTIBIOTICS.26.745.

48. T. Oda, T. Mori, H. Itō, T. Kunieda, and K. Munakata, "Studies on new antibiotic lividomycins I. Taxonomic studies on the lividomycin-producing strain *Streptomyces lividus* nov. sp. ," *The Journal of Antibiotics*, vol. 24, no. 6, pp. 333–338, Jun. 1971, doi: 10.7164/ANTIBIOTICS.24.333.

49. S. A. Waksman and H. A. Lechevalier, "Neomycin, a New Antibiotic Active against Streptomycin-Resistant Bacteria, including Tuberculosis Organisms," *Science*, vol. 109, no. 2830, pp. 305–307, 1949, doi: 10.1126/SCIENCE.109.2830.305.

50. C. L. M. Gilchrist and Y.-H. Chooi, "clinker & clustermap.js: automatic generation of gene cluster comparison figures," *Bioinformatics*, vol. 37, no. 16, pp. 2473–2475, 2021, doi: 10.1093/BIOINFORMATICS/BTAB007.

51. Hamao. Umezawa and I. R. Hooper, "Aminoglycoside Antibiotics," 1982.

52. J. Zhang, P.-H. Fan, G.-M. Lin, W.-C. Chang, and H. Liu, "Recent Progress in Unusual Carbohydrate-Containing Natural Products Biosynthesis," in *Comprehensive Natural Products*, Third Edition., vol. 2, H.-W. (Ben) Liu and T. P. Begley, Eds. Elsevier Science, 2020, pp. 336–392.

53. J. W. Park *et al.*, "Discovery of parallel pathways of kanamycin biosynthesis allows antibiotic manipulation," *Nature Chemical Biology*, vol. 7, no. 11, pp. 843–852, 2011, doi: 10.1038/nchembio.671.

54. J. W. Park *et al.*, "Genetic dissection of the biosynthetic route to gentamicin A2 by heterologous expression of its minimal gene set.," *PNAS*, vol. 105, no. 24, pp. 8399–404, 2008, doi: 10.1073/pnas.0803164105.

55. C. Huang *et al.*, "Delineating the Biosynthesis of Gentamicin X2, the Common Precursor of the Gentamicin C Antibiotic Complex," *Chemistry & Biology*, vol. 22, no. 2, pp. 251–261, 2015, doi: 10.1016/j.chembiol.2014.12.012.

56. J. Guo *et al.*, "Specificity and Promiscuity at the Branch Point in Gentamicin Biosynthesis," *Chemistry & Biology*, vol. 21, no. 5, pp. 608–618, 2014, doi: 10.1016/j.chembiol.2014.03.005.

57. H. Nishimura, K. Katagiri, K. Sato, M. Mayama, N. Shimaoksa, "Toyocamycin, a new anti-candida antibiotics," *J Antibiot (Tokyo)*, vol. 9, no. 2, pp. 60-2, 1956, doi: 10.11554/antibioticsa.9.2_60.

58. D. Xu *et al.*, "PreQ0 Base, an Unusual Metabolite with Anti-cancer Activity from *Streptomyces qinglanensis* 172205," *Anti-Cancer Agents in Medicinal Chemistry*, vol. 15, no. 3, pp. 285–290, Oct. 2015, doi: 10.2174/1871520614666141027144653.

59. O. Hara and T. Beppu, "Mutants blocked in streptomycin production in *Streptomyces griseus* - the role of A-factor.," *J Antibiot (Tokyo)*, vol. 35, no. 3, pp. 349–58, 1982.

60. J. Distler, K. Mansouri, G. Mayer, M. Stockmann, and W. Piepersberg, "Streptomycin biosynthesis and its regulation in Streptomycetes," *Gene*, vol. 115, no. 1–2, pp. 105–111, Jun. 1992, doi: 10.1016/0378-1119(92)90547-3.

61. S. Horinouchi and T. Beppu, "Hormonal control by A-factor of morphological development and secondary metabolism in Streptomyces," *Proceedings of the Japan Academy. Series B, Physical and Biological Sciences*, vol. 83, no. 9–10, p. 277, 2007, doi: 10.2183/PJAB/83.277.

62. S. Horinouchi, "Mining and Polishing of the Treasure Trove in the Bacterial Genus *Streptomyces*," *Bioscience, Biotechnology, and Biochemistry*, vol. 71, no. 2, pp. 283–299, 2007, doi: 10.1271/BBB.60627.

63. Q. Li *et al.*, "Epoxyqueuosine Reductase QueH in the Biosynthetic Pathway to tRNA Queuosine Is a Unique Metalloenzyme," *Biochemistry*, vol. 60, no. 42, pp. 3152–3161, Oct. 2021, doi: 10.1021/ACS.BIOCHEM.1C00164.

64. Z. D. Miles, R. M. McCarty, G. Molnar, and V. Bandarian, "Discovery of epoxyqueuosine (oQ) reductase reveals parallels between halorespiration and tRNA modification," *PNAS*, vol. 108, no. 18, pp. 7368–7372, 2011, doi: 10.1073/PNAS.1018636108.

65. E. R. van Vlack, S. Topp, and J. C. Seeliger, "Characterization of Engineered PreQ1 Riboswitches for Inducible Gene Regulation in Mycobacteria," *Journal of Bacteriology*, vol. 199, no. 6, 2017, doi: 10.1128/JB.00656-16.

66. Z. Weinberg *et al.*, "Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline," *Nucleic Acids Research*, vol. 35, no. 14, p. 4809, 2007, doi: 10.1093/NAR/GKM487.

67. M. Kang, C. D. Eichhorn, and J. Feigon, "Structural determinants for ligand capture by a class II preQ1 riboswitch," *PNAS*, vol. 111, no. 6, pp. E663–E671, 2014, doi: 10.1073/PNAS.1400126111/

68. S. Cao and S. J. Chen, "Predicting structures and stabilities for H-type pseudoknots with interhelix loops," *RNA*, vol. 15, no. 4, p. 696, 2009, doi: 10.1261/RNA.1429009.

69. J. Singh, J. Hanson, K. Paliwal, and Y. Zhou, "RNA secondary structure prediction using an ensemble of two-dimensional deep neural networks and transfer learning," *Nature Communications*, vol. 10, no. 1, pp. 1–13, Nov. 2019, doi: 10.1038/s41467-019-13395-9.

70. A. D. Garst, A. L. Edwards, and R. T. Batey, "Riboswitches: Structures and Mechanisms," *Cold Spring Harbor Perspectives in Biology*, vol. 3, no. 6, pp. 1–13, 2011, doi: 10.1101/CSHPERSPECT.A003533.

71. J. Sperschneider, A. Datta, and M. J. Wise, "Predicting pseudoknotted structures across two RNA sequences," *Bioinformatics*, vol. 28, no. 23, p. 3058, 2012, doi: 10.1093/BIOINFORMATICS/BTS575.

72. L. Horbal, T. Siegl, and A. Luzhetskyy, "A set of synthetic versatile genetic control elements for the efficient expression of genes in Actinobacteria," *Scientific Reports*, vol. 8, p. 491, 2018, doi: 10.1038/s41598-017-18846-1.

73. H. M. Salis, E. A. Mirsky, and C. A. Voigt, "Automated design of synthetic ribosome binding sites to control protein expression," *Nature Biotechnology 2009 27:10*, vol. 27, no. 10, pp. 946–950, 2009, doi: 10.1038/nbt.1568.

74. T. Tuller and H. Zur, "Multiple roles of the coding sequence 5′ end in gene expression regulation," *Nucleic Acids Research*, vol. 43, no. 1, pp. 13–28, 2015, doi: 10.1093/NAR/GKU1313.

75. M. C. Wu *et al.*, "Rational Re-engineering of a Transcriptional Silencing PreQ1 Riboswitch," *Journal of the American Chemical Society*, vol. 137, no. 28, pp. 9015–9021, 2015, doi: 10.1021/JACS.5B03405.

76. S. Balaratnam *et al.*, "A chemical probe based on the PreQ 1 metabolite enables transcriptome-wide mapping of binding sites," *Nature Communications*, vol. 12, no. 1, 2021, doi: 10.1038/S41467-021-25973-X.

77. F. Xu *et al.*, "A Genetics-Free Method for High-Throughput Discovery of Cryptic Microbial Metabolites," *Nature Chemical Biology*, vol. 15, no. 2, p. 161, 2019, doi: 10.1038/S41589-018-0193-2.

78. M. Myronovskyi and A. Luzhetskyy, "Native and engineered promoters in natural product discovery," *Natural Product Reports*, vol. 33, no. 8, pp. 1006–19, 2016, doi: 10.1039/c6np00002a.

79. M. Myronovskyi, E. Welle, V. Fedorenko, and A. Luzhetskyy, "β-Glucuronidase as a Sensitive and Versatile Reporter in Actinomycetes," *Applied and Environmental Microbiology*, vol. 77, no. 15, p. 5370, 2011, doi: 10.1128/AEM.00434-11.

80. N. C. Shaner, R. E. Campbell, P. A. Steinbach, B. N. G. Giepmans, A. E. Palmer, and R. Y. Tsien, "Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein," *Nature Biotechnology 2004 22:12*, vol. 22, no. 12, pp. 1567–1572, 2004, doi: 10.1038/nbt1037.

81. J. A. N. Brophy and C. A. Voigt, "Principles of genetic circuit design," *Nature Methods*, vol. 11, no. 5. Nature Publishing Group, pp. 508–520, 2014. doi: 10.1038/nmeth.2926.

82. M. M. Rudolph, M. P. Vockenhuber, and B. Suess, "Synthetic riboswitches for the conditional control of gene expression in Streptomyces coelicolor," *Microbiology (N Y)*, vol. 159, no. Pt 7, pp. 1416–1422, 2013, doi: 10.1099/MIC.0.067322-0.

83. A. Bermingham and J. P. Derrick, "The folic acid biosynthesis pathway in bacteria: Evaluation of potential for antibacterial drug discovery," *BioEssays*, vol. 24, no. 7, pp. 637–648, 2002, doi: 10.1002/BIES.10114.

84. J. C. Yoo, J. M. Han, O. H. Ko, and H. J. Bang, "Purification and characterization of GTP cyclohydrolase I from Streptomyces tubercidicus, a producer of tubercidin," *Arch Pharm Res*, vol. 21, no. 6, pp. 692–697, 1998, doi: 10.1007/BF02976759.

85. K. V. Baker *et al.*, "Chapter 3 - Towards improved selective gentamicin congener production in Streptomyces spp.," PhD Thesis Chapter, University of Manchester, Manchester, 2022.

86. H. L. Kim *et al.*, "Heterogeneous rRNAs are differentially expressed during the morphological development of Streptomyces coelicolor," *FEMS Microbiology Letters*, vol. 275, no. 1, pp. 146–152, 2007, doi: 10.1111/J.1574-6968.2007.00872.X.

87. B. P. Anton and E. A. Raleigh, "Complete Genome Sequence of NEB 5-alpha, a Derivative of Escherichia coli K-12 DH5α.," *Genome Announcements*, vol. 4, no. 6, 2016, doi: 10.1128/genomeA.01245-16.

88. D. J. MacNeil, K. M. Gewain, C. L. Ruby, G. Dezeny, P. H. Gibbons, and T. MacNeil, "Analysis of Streptomyces avermitilis genes required for avermectin biosynthesis utilizing a novel integration vector," *Gene*, vol. 111, no. 1, pp. 61–68, 1992, doi: 10.1016/0378-1119(92)90603-M.

89. P. R. Burkholder and N. H. Giles, "Induced biochemical mutations in *Bacillus subtilis*.," *Am J Bot*, vol. 34, no. 6, pp. 345–8, Jun. 1947.

90. F. Kunst *et al.*, "The complete genome sequence of the gram-positive bacterium Bacillus subtilis.," *Nature*, vol. 390, no. 6657, pp. 249–256, 1997, doi: 10.1038/36786.

91. J. P. Gomez-Escribano and M. J. Bibb, "Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters," *Microbial Biotechnology*, vol. 4, no. 2, pp. 207–215, 2011, doi: 10.1111/j.1751-7915.2010.00219.x.

92. S. Ye, B. Enghiad, H. Zhao, and E. Takano, "Fine-tuning the regulation of Cas9 expression levels for efficient CRISPR-Cas9 mediated recombination in *Streptomyces*," *Journal of Industrial Microbiology & Biotechnology*, vol. 47, no. 4–5, pp. 413–423, 2020, doi: 10.1007/s10295-020-02277-5.

93. S. Bellaousov and D. H. Mathews, "ProbKnot: fast prediction of RNA secondary structure including pseudoknots," *RNA*, vol. 16, no. 10, pp. 1870–1880, 2010, doi: 10.1261/RNA.2125310.

94. M. A. Larkin *et al.*, "Clustal W and Clustal X version 2.0," *Bioinformatics*, vol. 23, no. 21, pp. 2947–2948, 2007, doi: 10.1093/BIOINFORMATICS/BTM404.

95. J. P. Torella, F. Lienert, C. R. Boehm, J. H. Chen, J. C. Way, and P. A. Silver, "Unique nucleotide sequence-guided assembly of repetitive DNA parts for synthetic biology applications," *Nature Protocols*, vol. 9, no. 9, pp. 2075–2089, 2014, doi: 10.1038/nprot.2014.145.

96. J. Sambrook and D. W. Russell, *Molecular cloning: a laboratory manual.* Cold Spring Harbour Laboratory, 2001.

97. T. Kieser, M. J. Bibb, M. J. Buttner, K. F. Chater, and D. A. Hopwood, "Practical *Streptomyces* Genetics," *John Innes Foundation,* p. 529, 2000. doi: 10.4016/28481.01.

Table S4.1: Percentage of each amino acid with 'GUN' anticodon in Cluster 24 genes, presented alternatively as Figure 4.4.

Gene Name	His %	Asp %	Asn %	Tyr %	His+Asp+Asn+Tyr %
trpS2	2.22%	5.56%	3.06%	2.78%	13.61%
genO	2.07%	6.48%	1.04%	1.81%	11.40%
gmrB	4.18%	6.46%	1.52%	1.90%	14.07%
genB1	2.48%	7.71%	1.93%	1.93%	14.05%
genQ	1.59%	6.94%	2.58%	2.78%	13.89%
genD3	1.82%	7.27%	2.91%	0.36%	12.36%
genM1	3.13%	6.51%	0.24%	1.45%	11.33%
gmrA	1.85%	7.41%	2.22%	2.59%	14.07%
genS1	2.84%	4.98%	2.84%	2.37%	13.03%
genC	2.77%	5.29%	2.27%	2.02%	12.34%
genD2	2.67%	6.82%	2.08%	1.19%	12.76%
genM2	3.79%	6.16%	1.42%	3.08%	14.45%
genD1	2.12%	7.13%	3.64%	4.10%	17.00%
genS2	4.07%	7.66%	2.39%	1.67%	15.79%
genW	1.63%	8.13%	0.81%	2.44%	13.01%
genB4	2.25%	6.74%	2.25%	4.27%	15.51%
genP	3.70%	10.00%	1.85%	2.22%	17.78%
genB3	1.78%	6.44%	2.89%	3.56%	14.67%
genK	2.51%	6.43%	2.04%	2.04%	13.01%
genB2	1.93%	7.25%	1.69%	3.14%	14.01%
genX	2.94%	9.41%	2.35%	2.35%	17.06%
genU	4.70%	8.39%	0.67%	1.01%	14.77%
genV	0.00%	2.08%	1.62%	1.85%	5.56%
genE	1.96%	6.54%	2.29%	3.59%	14.38%
genY	1.55%	2.87%	0.66%	1.10%	6.18%
genA	3.56%	8.00%	0.89%	4.00%	16.44%
genF	4.19%	6.05%	1.40%	2.79%	14.42%
genG	5.98%	7.69%	2.56%	2.56%	18.80%
genH	1.49%	5.27%	2.02%	2.11%	10.89%
genI	3.00%	7.74%	1.42%	2.21%	14.38%
genT	3.21%	5.77%	1.60%	1.60%	12.18%
genD	3.20%	8.22%	0.46%	1.37%	13.24%
genN	1.25%	7.79%	3.12%	1.56%	13.71%
orf_5081	0.55%	19.34%	0.55%	0.00%	20.44%
orf_5082	1.14%	5.49%	0.92%	0.00%	7.55%
orf_5083	2.16%	9.96%	0.87%	1.30%	14.29%
orf_5084	0.62%	4.97%	1.86%	0.62%	8.07%
orf_5085	1.99%	5.68%	5.40%	2.84%	15.91%
orf_5086	0.00%	5.84%	0.00%	1.30%	7.14%
orf_5087	1.40%	8.37%	0.47%	1.86%	12.09%
orf_5088	1.92%	5.04%	1.44%	2.16%	10.55%
orf_5089	2.86%	0.00%	5.71%	0.00%	8.57%
orf_5090	2.30%	5.65%	1.04%	1.04%	10.02%
orf_5091	0.00%	5.00%	0.00%	0.00%	5.00%
orf_5092	2.40%	2.99%	3.59%	7.19%	16.17%
orf_5093	1.88%	5.65%	5.11%	2.69%	15.32%
orf_5094	0.82%	3.28%	3.28%	3.28%	10.66%

orf_5095	0.00%	5.13%	0.00%	1.28%	6.41%
orf_5096	0.87%	4.36%	5.81%	4.94%	15.99%
orf_5097	2.72%	6.27%	0.54%	0.82%	10.35%
orf_5098	2.38%	7.75%	1.00%	1.50%	12.63%
orf_5099	1.42%	9.54%	0.77%	1.16%	12.89%
orf_5100	2.48%	7.43%	1.24%	1.55%	12.69%
orf_5101	1.09%	2.55%	1.09%	1.82%	6.55%
orf_5102	3.83%	1.92%	1.53%	2.30%	9.58%
orf_5103	2.87%	9.18%	0.96%	1.91%	14.91%
orf_5104	0.00%	6.58%	2.63%	1.32%	10.53%
genJ	4.91%	5.96%	3.16%	2.46%	16.49%
genK2	2.23%	10.19%	0.96%	2.23%	15.61%

Table S4.2: PCR primers used in this study

Primer Name	Primer Sequence 5′ – 3′	Product length (bp)	Function
pTE1332-Mabsfw	ACACGTGGCAAAGCCGCTAAAGGAGAAAATATGGTCTCCAAGGGCGAGGAG	6557	Generation of pTE1332-Mabs-trunc(truncated *M. abscessus* riboswitch)
pTE1332ermEp1rev	CGCTGGATCCTACCAACCGGC		
pET1332ermEp1Mabsrev	AATATGCTTGGTGAAAGAAGCTAAGCACCAAGGGTTGAGCAAGTCAACTTTACAAGTATAGCACACGCTGGATCCTACCAACCGGC	6622	Generation of pTE1332-Mabs-full (full *M. abscessus* riboswitch); using pTE1332-Mabsfw as forward primer
pTE1332-Lrhfw	GCCACAAAGGAGAAACAATATGGCCAAGACAATTGCGGATCATGGTCTCCAAGGGCGAGGAG	6557	Generation of pTE1332-Lrh-trunc (truncated *L. rhamnosus* riboswitch) using pTE1332ermEp1rev as reverse primer
pTE1332-Lrh2fw	TTTGATCGTCGTTATTACTGGCAAAGCCACAAAGGAGAAACAATATGGTCTCCAAGGGCGAGGAG	6610	6610 generation of pTE1332-Lrh-full (full *L. rhamnosus* riboswitch)
pTE1332-Lrh2rev	GGAAATAAGTATCGTCGTGGAAGCGCGCGGGGCGGTTCCCGCTGGATCCTACCAACCGGC		
pTE1332-Lrhtrunc2fw	TTACTGGCAAAGCCACAAAGGAGAAACAATATGGTCTCCAAGGGCGAGGAG	6557	Redesigned primer for generation of pTE1332-Lrh-trunc (truncated *L. rhamnosus* riboswitch) using pTE1332ermEp1rev as reverse primer
pTE1332-genAnp-fw	GGGTGGTGGCGGCGCGTGCCGACGGAGAAGCGCAGGGCCATGGTCTCCAAGGGCGAGGAG	6653	Primers for use in nested primer PCR to amplify pTE1332 with *genA* upstream region in place of RBS. First rxn = pTE1332-genAnp-rev, pTE1332-genAnp-fw; second rxn uses template from first with pTE1332-genAnp-rev, pTE1332-genAnp-2fw.
pTE1332-genAnp-rev	GGGCGCGGTCCTGCGGTACGCGACACACCGGCGACGGGCTCTCCGTCGCCGGTGTGTCGCGGTGGCGGCCGCTGGATCCTACCAACCGGC		
pTE1332-genAnp-2fw	TGAGCCGGGTCCCCGCCTGAACGCTGAGTAGGCTTACCGCCGGGTACGACCGTGCCCGTGGACGAGGTGAAGGGTGGTGGCGGCGCGTGC	6713	

pTE1332genAfw	CCCCGCCTGAACGCTGAGTAGGCTTACCGCCGGGTACGACCGTGCCCGTGGACGAGGGTGAAGGGTGGTGGCGGCGCATGGTCTCCAAGGGCGAGGAG	6557	Redesigned primers to generate pTE1332 variant with correct *genA* upstream region upstream of mCherry. First reaction = pTE1332genAfw + pTE1332ermEp1rev; Second reaction = pTE1332genAfw2 + pTE1332ermEp1rev; Third reaction = pTE1332genAfw3 + pTE1332ermEp1rev
pTE1332genAfw2	GACGGAGAGCCCGTCGCCGGTGTGTCGCGTACCGCAGGACCGCGCCCTGAGCCGGGTCCCCGCCTGAACGCTGAGTA	6661	
pTE1332genAfw3	GCCGCCACCGCGACACACCGGCGACGGAGAGCCCGTCG	6683	
RBswitch_genA_colfw	GACGGAGAAGCGCAGGG	980	Colony PCR to detect first pTE1332-genA-trunc plasmid (to be used with EHseq002)
RBswitch_Mabs_colfw	GGCAAAGCCGCTAAAGGAGAAAAT	980	Colony PCR to detect pTE1332-Mabs-trunc/pTE1332-Mabs-full plasmids (to be used with EHseq002)
RBswitch_Lrh_colfw	GGCCAAGACAATTGCGGATC	980	Colony PCR to detect first pTE1332-Lrh-trunc plasmid (to be used with EHseq002)
genAcolfw2	GTGGACGAGGGTGAAGGGTG	792	Colony PCR to detect second pTE1332-genA-full plasmid (to be used with EHseq002)
Lrhcolfw2	TTACTGGCAAAGCCACAAAGGAG	792	Colony PCR to detect second pTE1332-Lrh-full plasmid (to be used with EHseq002)
EHseq002	GTGTGGAATTGTGAGCGGA	792 – 980	Colony PCR to detect correct pTE1332-based plasmid assembly, sequencing primer
EHseq001	TTTCCCAGTCACGACGTTG	-	Sequencing primer
pCM4.4 HA site UNS7 fw	CAAGACGCTGGCTCTGACATTTCCGCTACTGAACTACTCGACGCTCAGTGGAACGAAAAC	5041	Amplification of Fragment 1 for assembly of pCM4.4 pCM4.4ΔgenAFG
dLancI-II 1 rev	GCTTGGATTCTGCGTTTGTTTCCGTCTACGAACTCCCAGCGGGACGTGCTTGGCAATCA		
Cas9 UNS2 fw	GCTGGGAGTTCGTAGACGGAAACAAACGCAGAATCCAAGCCATGCGCTCCATCAAGAA	5743	Amplification of Fragment 2 for assembly of pCM4.4 pCM4.4ΔgenAFG
pCM4.4 HA site UNS6 rev	GTATGTGACCGTAGAGTATTCTTAGGTGGCAGCGAACGAGCAGACCCCGTAGAAAAGA		
UNS6F	CTCGTTCGCTGCCACCTAAGAATACTCTACGGTCACATAC	2080	Amplification of Fragment 3 for assembly of pCM4.4 pCM4.4ΔgenAFG
UNS7R	CGAGTAGTTCAGTAGCGGAAATGTCAGAGCCAGCGTCTT		
C24_AFGdelfw	GCTGGAAGACGTGATGGCCTGGAT	2160	Colony PCR to detect successful knockout of *genA*, *genF*, *genG* from Cluster 24 (3635 bp if not mutated, 2160 bp if successful knock-out occurred)
C24_AFGdelrev	GTTCGACAAGCAGGGCAACG		

Table S4.3: Sequences of riboswitch regions inserted to each test plasmid.

Plasmid name	Sequence of riboswitch region 5′ – 3′
pTE1332-genA-full	GCCGCCACCGCGACACACCGGCGACGGAGAGCCCGTCGC CGGTGTGTCGCGTACCGCAGGACCGCGCCCTGAGCCGGG TCCCCGCCTGAACGCTGAGTAGGCTTACCGCCGGGTACG ACCGTGCCCGTGGACGAGGTGAAGGGTGGTGGCGGCGC
pTE1332-genA-trunc	CGGCGCGTGCCGACGGAGAAGCGCAGGGCC
pTE1332-Mabs-full	TGTGCTATACTTGTAAAGTTGACTTGCTCAACCCTTGGT GCTTAGCTTCTTTCACCAAGCATATTACACGTGGCAAAG CCGCTAAAGGAGAAAAT
pTE1332-Mabs-trunc	ACACGTGGCAAAGCCGCTAAAGGAGAAAAT
pTE1332-Lrh-full	GGAACCGCCCCGCGCGCTTCCACGACGATACTTATTTCC TTTGATCGTCGTTATTACTGGCAAAGCCACAAAGGAGAA ACAAT
pTE1332-Lrh-trunc (non-functional)	GAAACAATATGGCCAAGACAATTGCGGATC

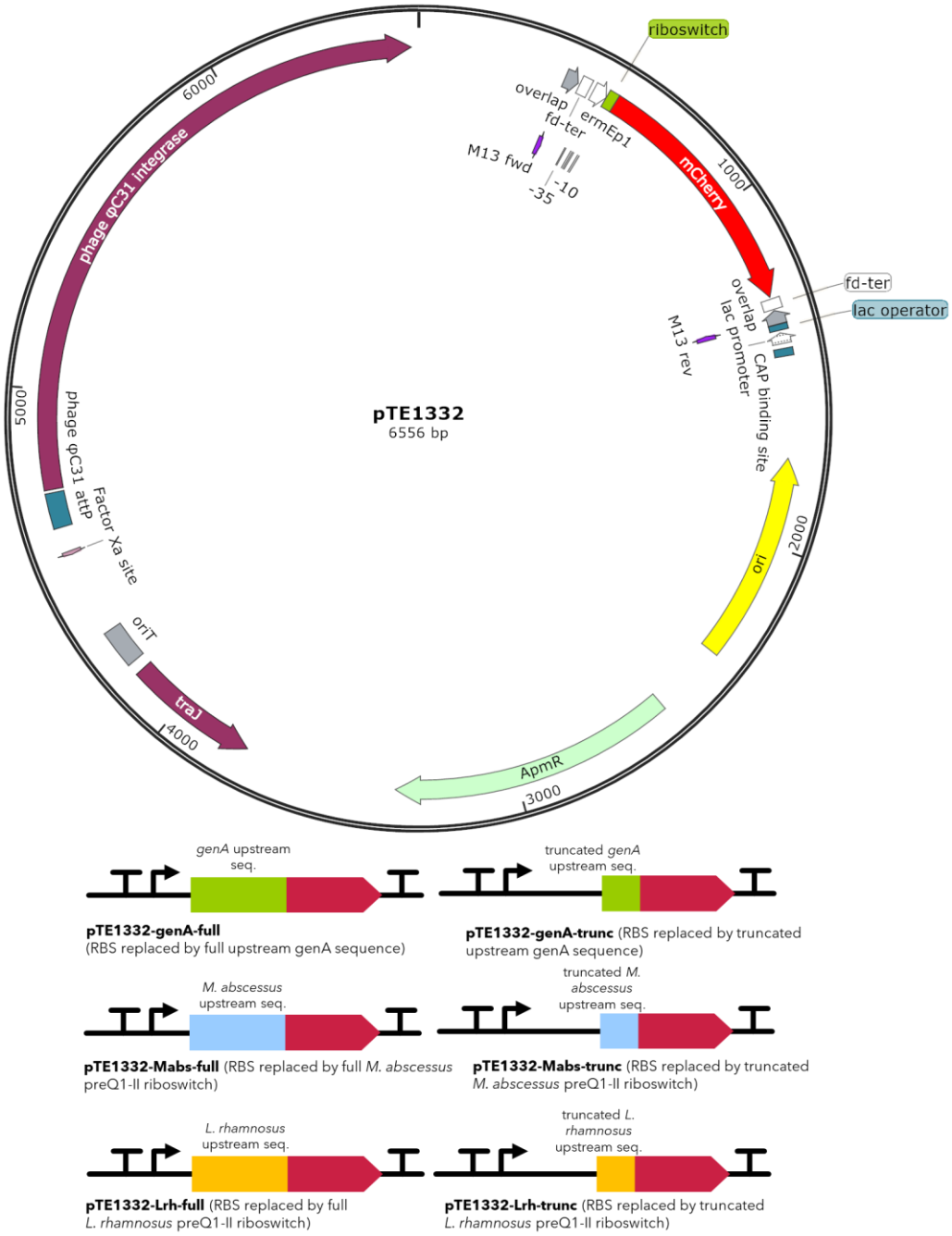
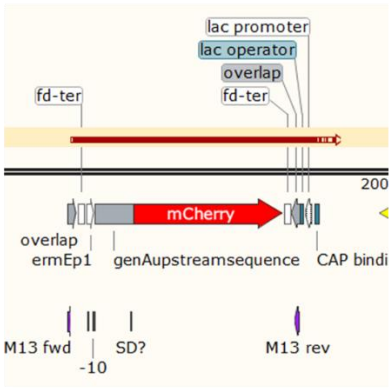
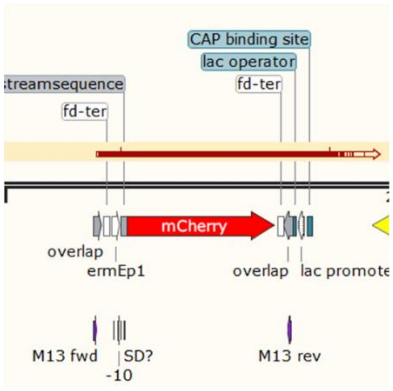


Figure S4.1: Plasmid map of pTE1332 and synthetic biology open language (SBOL) schematics of each construct. Riboswitch insert region is marked in green on pTE1332 plasmid map.

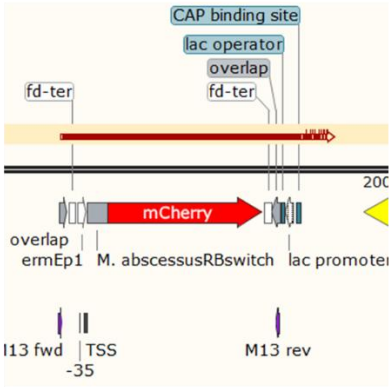
pTE1332-genA-full
(RBS replaced by full upstream *genA* sequence)



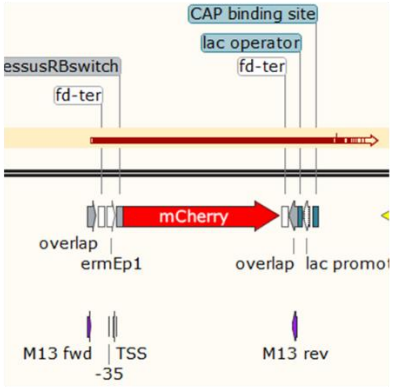
pTE1332-genA-trunc
(RBS replaced by truncated upstream *genA* sequence)



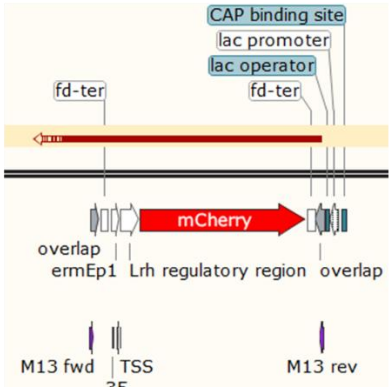
pTE1332-Mabs-full
(RBS replaced by full *M. abscessus* preQ1-II riboswitch)



pTE1332-Mabs-trunc
(RBS replaced by truncated *M. abscessus* preQ1-II riboswitch)



pTE1332-Lrh-full
(RBS replaced by full *L. rhamnosus* preQ1-II riboswitch)



pTE1332-Lrh-trunc
(RBS replaced by truncated *L. rhamnosus* preQ1-II riboswitch)

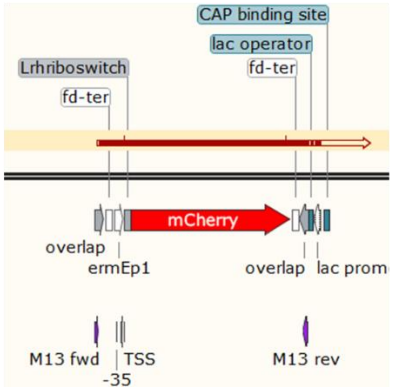


Figure S4.2: Positive Sanger sequencing results for all six riboswitch variant plasmids tested in this work. For all plasmids except pTE1332-Lrh-full, primer EHseq001 was used. For pTE1332-Lrh-full, EHseq002 was used.

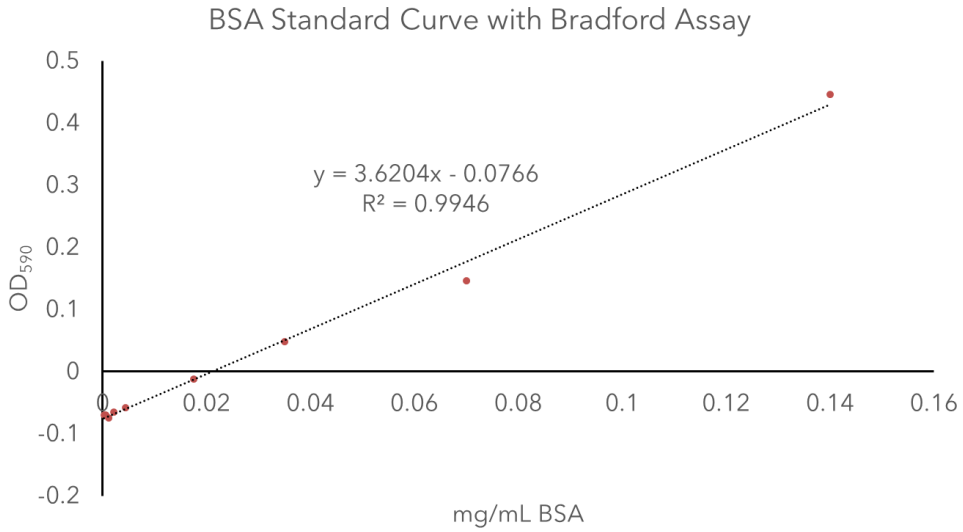


Figure S4.3: Bovine serum albumin standard curve for normalisation of fluorescence measurements.

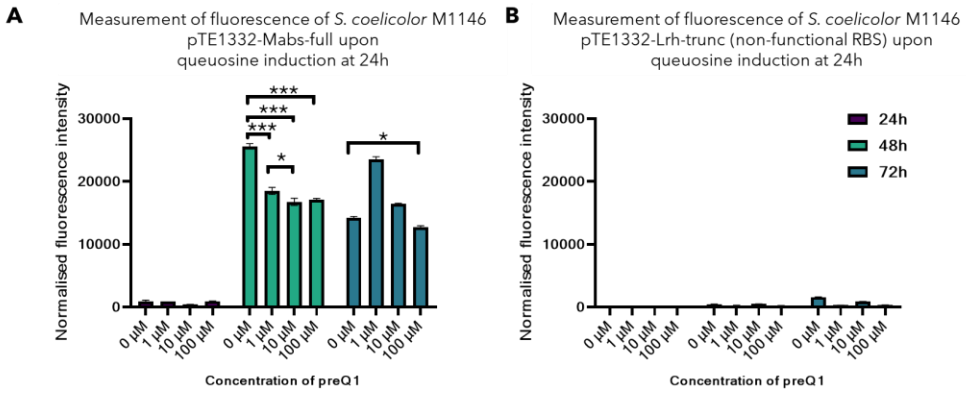


Figure S4.4: Normalised fluorescence of *S. coelicolor* M1146 pTE1332-Mabs-full and negative control pTE1332-Lrh-trunc under preQ1 induction of three different concentrations. Induction with preQ1 was done at 24h growth. A – Normalised fluorescence of *S. coelicolor* M1146 pTE1332-Mabs-full induced with preQ1; **B** – Normalised fluorescence of *S. coelicolor* M1146 pTE1332-Lrh-trunc (negative control) induced with preQ1. * denotes statistical significance $p < 0.05$; ** denotes statistical significance $p < 0.001$; *** denotes statistical significance $p < 0.0001$.

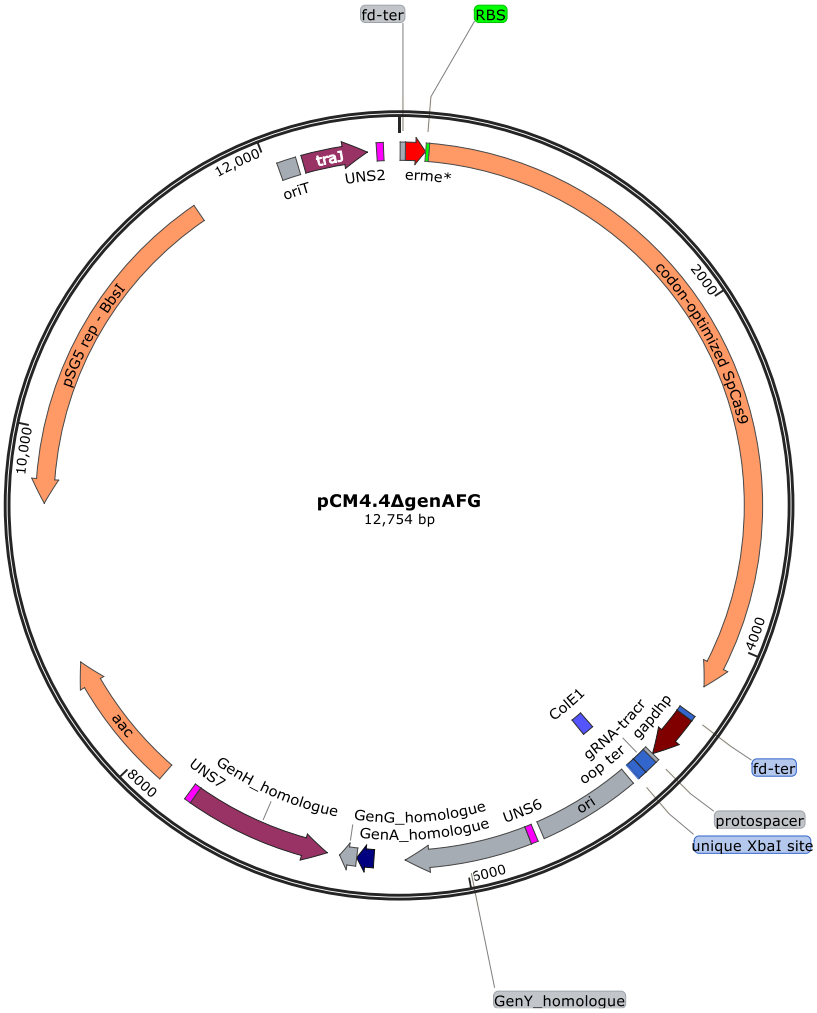


Figure S4.5: Plasmid map of pCM4.4ΔgenAFG.

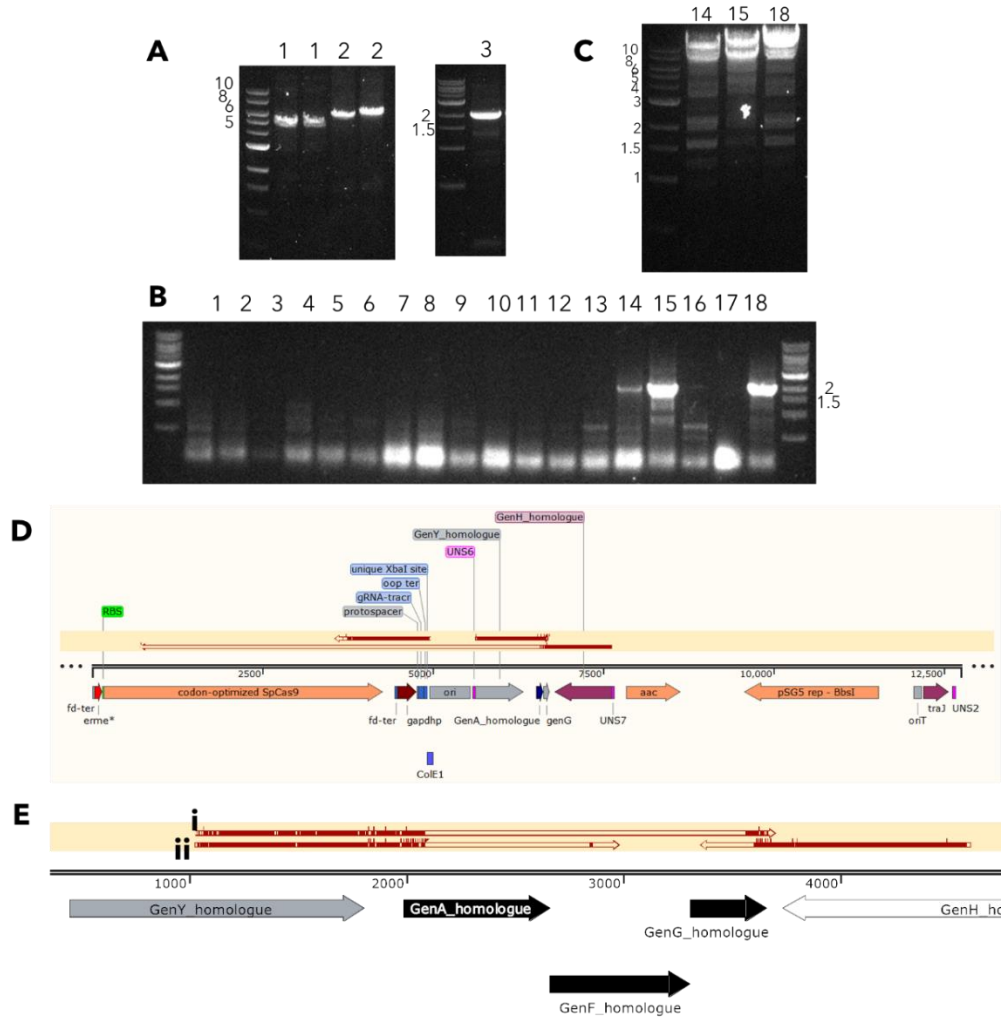


Figure S4.6: Construction of pCM4.4ΔgenAFG. A – PCR amplification of fragments for assembly (1 – amplified with pCM4.4 HA site UNS7 fw and dLancI-II 1 rev, expected band size = 5041 bp; 2 – amplified with Cas9 UNS2 fw and pCM4.4 HA site UNS6 rev, expected band size = 5743 bp; 3 – amplified with UNS6F and UNS7R, expected band size = 2080 bp. **B** – *E. coli* colony PCR to screen for possible correct assemblies, using UNS6F/UNS7R primers screening for a product of 2080 bp. **C** – PstI restriction digest of possible correct assemblies; expected band sizes = 6443 bp, 2202 bp, 1558 bp, 1212 bp, 920 bp, 419 bp. **D** – Sanger sequencing of protospacer and homology arm regions to check correct assembly. **E** – Sanger sequencing of i) *S. coelicolor* M1146 C24 2R *ΔgenA ΔgenF ΔgenG* ii) *S. coelicolor* M1146 C24 2R + promoter cassette *ΔgenA ΔgenF ΔgenG* after curing of pCM4.4ΔgenAFG, to confirm correct deletion.

5. Conclusions and Future Perspectives

5.1. Summary of work

The work presented in this thesis uses multiple approaches towards heterologous expression of a novel aminoglycoside cluster from *Micromonospora* sp. DEM32671 in *Streptomyces* spp. Our initial *in silico* predictions suggested that the cluster product would be gentamicin, an aminoglycoside which has been linked to nephro- and ototoxic side effects when used clinically.[1,2] Gentamicin is administered as a mixture of five different congeners, the proportions of which vary hugely in each batch.[3] Different congeners have additionally been shown to convey different levels of nephro- and ototoxicty.[4–6] Therefore, the tailored production of specific gentamicin congeners is of huge value to revitalise this compound as a therapeutic.

In Chapter 2, we first describe the analysis of novel strain *M.* sp. DEM32671. We carried out *gyrB*-based phylogenetic tree construction, digital DNA-DNA hybridisation amongst nearest neighbours and orthologous average nucleotide identity amongst nearest neighbours to place the strain in the genus *Micromonospora*. We believe this may represent a novel species based on the diversity in cluster products between this strain and its closest relative *Micromonospora echinospora* (the industrial gentamicin producer), but further phenotypic and metabolic characterisation experiments would more firmly establish this. We identified Cluster 24 as the gene cluster of interest for our future work after carrying out antiSMASH analysis of the genome sequence. Cluster 24 was predicted to be an aminoglycoside cluster, matching the existing secondary metabolite production from the native strain. After selecting three *Streptomyces* strains for testing of the gene cluster we identified low levels of antibiotic production in two of these, *S. lividans* TK23 C24 and *S. coelicolor* M1152 C24. As these strains were not fit for further use (due to being less amenable to CRISPR-Cas9 modification and having lower growth fitness respectively) we decided to further improve our ideal host, *S. coelicolor* M1146 C24. This strain, carrying the intact unmodified Cluster 24 sequence, had not previously

shown bioactivity n in early antimicrobial bioassay tests. We identified that the parts of the cluster were being lost from the strain upon repeat passaging, even with the presence of thiostrepton for cosmid maintenance. Hypothesising that the cluster product may be toxic to the cells, and at the same time establishing that this strain, *S. coelicolor* M1146, was indeed sensitive to gentamicin, we generated a cohort of resistant 'superhost' strains through iterative exposure to increasing concentrations of gentamicin sulfate. After carrying out genome sequencing of two of these strains, we identified that a large array of mutations had taken place within the strains which could not be easily explained to be causative of this resistant phenotype. One strain (*S. coelicolor* M1146 C24 2R) also showed an increased antimicrobial activity (measured through antimicrobial bioassays) in comparison to the parental strain; when we carried out LC-MS analysis of this stain, we were unable to identify end-point gentamicin congeners. Instead, we identified several precursors for 4,6-disubstituted 2-deoxystreptamine aminoglycoside biosynthesis, as well as a peak of m/z 502.2246 with predicted chemical formula $C_{18}H_{35}N_3O_{13}$ (m/z 502.2248), which was not seen in the negative control and may represent an underseen or novel aminoglycoside.

Chapter 3 describes our early work towards production of selected gentamicin congeners from Cluster 24, using a two-pronged approach. We aimed to test both CRISPR-Cas9 gene-knockouts and total cluster rebuilding to produce gentamicin A2 and gentamicin C1a as sole fermentation products. Gentamicin A2 represents the first pseudo-trisaccharide of the pathway,[7] and we originally aimed to use this as a feedstock for further side-chain modification. Gentamicin C1a was identified by Ishikawa et al. (2019) as being the least ototoxic gentamicin C congener of those they tested,[6] and therefore was prioritised from the other congeners available. Our first tests with a basic *S. coelicolor* M1152 C24 strain (*Δact Δred Δcpk Δcda rpoB*[S433L]; otherwise unmodified from *S. coelicolor* M145 aside from insertion of C24)[8] to knock out the genes responsible for downstream gentamicin C congener and gentamicin B biosynthesis. This strain showed barely detectable antibiotic activity on antimicrobial

bioassay; however, the yield was unsuitable for further characterisation via LC-MS as it was too low for detection (based on testing of standards of known concentration, described in Chapter 2). We next tested gentamicin A2 production using two reconstructed pathways with minimal gene sets characterised *in vitro* by Park et al. (2008).[7] The first plasmid had gene expression controlled using a tetracycline-inducible system.[9] As we saw no production from strains carrying this construct, we further constructed a variant with the terminal genes of each operon *genE* and *genM2* polyhistidine-tagged. We were unable to detect expression of these genes via western blot. To improve upon gene expression and therefore production of gentamicin A2, we constructed two further minimal gentamicin A2 pathway plasmid variants with each gene in its own transcriptional unit. Again, antimicrobial bioassay of these strains showed no antimicrobial activity. Through searching the literature, we determined that this congener did not show antibiotic activity against *Bacillus pumilus*,[10] making it likely that it also will not show any activity against our indicator strain, *Bacillus subtilis* 168. As we were unable to characterise the output from the knockout strains due to low yield, we instead tested three different minimal gentamicin A2 production constructs in combination with a plasmid carrying the required biosynthetic machinery for gentamicin C1a production. Through this, we were able to identify that the lack of host resistance to the heterologously-produced product is a clear bottleneck in this heterologous production strategy. A secondary aim of Chapter 3 was towards further improving antibiotic yield, and we utilised our over-producing 'superhost' strain for this. We designed a bi-directional promoter cassette to insert into Cluster 24 using CRISPR-Cas9, aiming to increase the expression levels of 12 of 34 cluster ORFs that were identified as homologous to gentamicin pathway genes with characterised function. From an antimicrobial bioassay, the yield of antibiotic active against *B. subtilis* appeared to be increased in the strain with inserted promoter cassette. Additionally, the predicted product $C_{18}H_{35}N_3O_{13}$ was again detected, this time being produced by the strain carrying the promoter cassette.

Finally, in Chapter 4, we identified the co-clustering of queuosine biosynthesis genes in certain aminoglycoside gene clusters; namely, the 4,6-disubstituted

deoxystreptamine aminoglycosides from *Micromonospora*, with the pseudodisaccharide aminoglycoside fortimicin as an outlier in the group.[11] Queuosine is a hyper-modified guanosine-based nucleobase which functions to improve translational fidelity at anticodons with sequence 'GUN'.[12] While the region of the Cluster 24 attributed to aminoglycoside biosynthesis was shown to have a higher proportion of 'GUN' amino acids than the region hypothesised to be 'non-biosynthetic', we determined that the queuosine biosynthesis pathway was likely to be non-intact even when the gene cluster was heterologously expressed in *S. coelicolor*. Our secondary hypothesis was that queuosine precursor preQ1 may have a regulatory effect on the gene cluster, as preQ1 has been linked to three classes of riboswitches which cause a dampening of gene expression upon ligand binding.[13] Through *in silico* sequence comparisons we found that the region upstream of *genA*, a putative 7-cyano-7-deazaguanine synthase, had some sequence similarity to known preQ1-II riboswitches but did not match the canonical H-type pseudoknot fold we would expect from riboswitches of this class.[13] We tested the preQ1 induction ability of a panel of three sequences in *S. coelicolor* (known preQ1-II riboswitches from *Mycobacterium abscessus* and *Lactobacillus rhamnosus*, as well as the region upstream of *M.* sp. DEM32671 Cluster 24 *genA*) by coupling them to a fluorescent reporter under control of a medium-strength constitutive promoter. While the *genA* putative preQ1-responsive region did not appear to be induced by presence of preQ1, the *M. abscessus* subsp. *abscessus* preQ1-II riboswitch showed a maximal dampening of expression by 36.73% from 0 μM to 1 μM preQ1, and a 9.72% decrease from 1 μM to 10 μM preQ1. This riboswitch candidate could be of great use in basic *Streptomyces* genetic circuits upon further characterisation.

5.1.1. Final conclusions

This work details the first steps towards heterologous production of an aminoglycoside from Cluster 24 of *M.* sp. DEM32671 in *Streptomyces coelicolor*. Overall, the work aimed to tackle the problems associated with heterologous expression of biosynthetic gene clusters in a variety of ways (**Figure 5.1**); namely the

alteration of media conditions, generation of a characterised panel of 'superhost' strains, cluster engineering via insertion of promoters and gene knock-outs for targeted congener production, and entire cluster rebuilding for establishment of a minimal genetic pathway towards high-value congener biosynthesis. The heterologous expression in this work undoubtedly showed numerous challenges, particularly as i) Cluster 24 had a degree of toxicity to the host strain, which had to be mitigated before production could be established ii) low titres ensured that any variability in production could appeared to abolish it in some cases. These difficulties meant that the biosynthesis of specific congeners was not able to be fully realised. Coupling use of the 'superhost' strains with the pre-built CRISPR plasmids for gene knockouts should, in the future, provide a robust platform for production of select congeners of gentamicin. We additionally aimed towards investigating putative regulatory regions of the cluster. The region of interest identified *in silico* upstream of *genA*, did not appear to contain a preQ1-linked regulatory region. However, we were able to characterise a preQ1-II riboswitch from *Mycobacterium abscessus* in *S. coelicolor*. This will undoubtedly be of use in the genetic engineering toolkit for this strain pending further characterisation.

Regarding the general heterologous expression approach going forward, there are clearly challenges to overcome. There are also numerous other factors to consider before undertaking this approach: amongst those which have been described are the location of pathway insertion into the chromosome,[14] the rearrangement of genetic elements to best suit production in the new host,[15] and the potential for toxicity from either pathway intermediates or final product.[16,17] Even still, across published work, a low number of instances of improvement in product titre from transfer from native to heterologous host have been described.[18] However, there are undoubtedly benefits to heterologous expression, particularly in decreasing culture time (in comparison to slow-growing producers) and in having a characterised and predictable host strain. Additionally, as the need for new antibiotics becomes greater as time passes, one of our greatest resources remains the uncultured microbes which may produce novel

bioactive scaffolds.[19,20] After using metagenomic approaches to capture DNA from the soil, a heterologous host is essential for production of the resultant natural products. For example, the discovery of the calcium-dependent antibiotics the malacidins in 2018 would not have been possible without use of a heterologous host. Despite the challenges, in some cases the heterologous expression approach is necessary for success. Nevertheless, it should be carefully utilised (and highly tailored) to ensure that it is the most suitable option for generation of the product of interest.

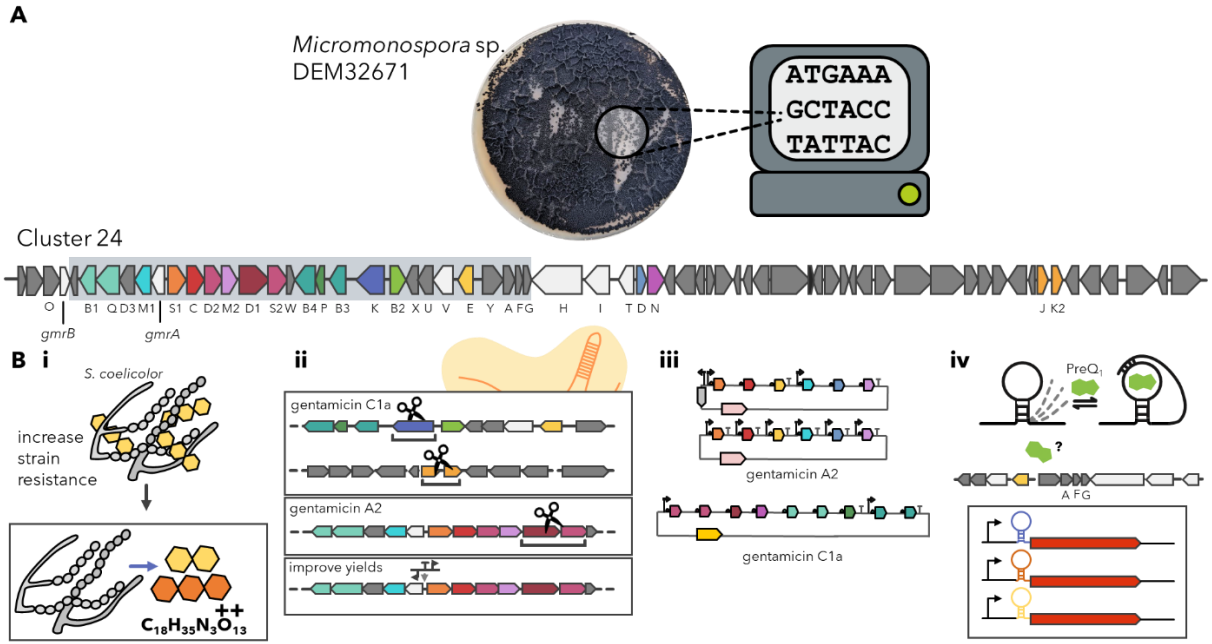


Figure 5.1: Summary of the work detailed in this thesis. A – Genome of *Micromonospora* sp. DEM32671 was sequenced, and Cluster 24 predicted to produce an aminoglycoside, most likely gentamicin. **B** – i) After initial difficulties with antibiotic production, exposure of *Streptomyces coelicolor* M1146 to increasing concentrations of gentamicin led to creation of 'superhost' strains, one of which was capable of an increased antibiotic production phenotype. ii) CRISPR-Cas9 was used for the knockout of *genK*, *genJ*, *genK2* for the selective production of gentamicin C1a, a congener purported to be less ototoxic;[6] the knock-out of *genS2* and *genD1*, for the selective production of gentamicin A2, a congener planned to be used as a starter feedstock for generation of designer hybrid antimicrobials; insertion of a promoter cassette in the core region of the gene cluster, for further improved yield. iii) Minimal gene sets for gentamicin A2 and gentamicin C1a were constructed and tested in *S. coelicolor* M1146. iv) Finally, the presence of a putative riboswitch in the cluster was tested through fluorescence assay, after identifying the co-clustering of queuosine biosynthesis genes within 4,6-disubstituted 2-deoxystreptamine aminoglycoside gene clusters from *Micromonospora*. We identified a riboswitch candidate from *M. abscessus* which may be of further use in *S. coelicolor*, dependent upon further testing.

The work detailed within this thesis shows the beginnings towards aminoglycoside production in heterologous host *S. coelicolor*. The most immediate future work from this is the combination of the 'superhost' strains with the designs for selective congener production, to establish strains with increased production yields of select aminoglycoside congeners. This would allow for full characterisation. Use of the native host to establish optimised analytical conditions for the cluster product would also be immediately valuable.

Next, this work could be further built upon with RT-qPCR experiments, including but not limited to: i) establishing the expression of cluster genes linked to aminoglycoside biosynthesis, in varying media conditions ii) use of primary expression data collected towards a rational selection of replacement promoters more suited for the *S. coelicolor* heterologous host, and further screening new engineered strains to check the changes made give desired effects; iii) establishing the expression profile of genes not anticipated to be involved in aminoglycoside biosynthesis, for example in the putative regulatory regions we identified as part of Chapter 4. This information would allow us to establish the next steps forward, using those strains where low antibiotic production inhibited further characterisation as examples where gene expression may be a bottleneck for antibiotic production.

The value in using a heterologous host is for easier production of specific congeners, for generating new hybrid antimicrobials and for having more flexibility in yield improvement strategies. As some *Micromonospora* strains have been shown to be genetically tractable (and CRISPR systems have been developed towards the modification of these strains),[21,22] use of a *Micromonospora* strain as a heterologous host may be more successful than a *Streptomyces* strain. While long culture times would inhibit use of the native producer in this work, sufficient growth of *M. echinospora* for gentamicin production has been reported in approximately one week,[23] similar to the required culture times for *Streptomyces* carrying Cluster 24.

However, if *Streptomyces* were continued to be used, our described pipeline should yield the selective congener once the described combination of knock-outs in the 'superhost'

strains is achieved. Since the decision to focus on gentamicin C1a as a less cytotoxic congener was made, further reports have suggested that gentamicin C2b convey less ototoxic effects.[5] The pathway we constructed in this work would require one extra gene, *genL*, for the biosynthesis of gentamicin C2b, which would be a simple additional modification.[10,24] This could be achieved via either supplementation of the Δ*genKgenJgenK2* strain with *genL*, or by incorporating *genL* into the minimal gentamicin C1a pathway plasmid pSG2. From our preliminary results presented in this thesis, supplementation of the Δ*genKgenJgenK2* strain appears likelier to yield success.

Once a feedstock of starter compound gentamicin A2 can be obtained, strategies could be developed towards a combinatorial aminoglycoside pathway, whereby the latter stages of the pathway are fulfilled by genes from other aminoglycosides. While this approach has been fully explored in the case of gentamicin–kanamycin hybrids,[25] to the best of our knowledge this has not been investigated using alternative aminoglycosides. 4,6-disubstituted 2-deoxystreptamine tobramycin or 4,5-disubstituted 2-deoxystreptamine butirosin may provide interesting starter points for this endeavour, due to their shared precursors with gentamicin.[26]

In terms of gentamicin/aminoglycoside yield, there remains room for improvement. The native host, *M. echinospora*, has been shown to produce gentamicin on the grams per liter scale, and our strains have been roughly estimated to produce micrograms per liter of aminoglycoside even with our early improvement strategies. After investigating the gene expression and improving promoter strength, another more purposeful strategy may be to investigate directed evolution of key gentamicin cluster genes. Glycosyltransferases *genM1* and *genM2* have been highlighted as being possibly rate-limiting in the pathway; therefore, if these could be engineered to have increased reaction kinetics, this may push yield up substantially. Wu et al tested the effect of pathway homologues in the native gentamicin producer by replacement of *genM1* by *kanM1*, an enzyme from the kanamycin biosynthesis pathway carrying out the same transfer of UDP-GlcNAc to 2-DOS. From doing this, they saw a yield improvement of gentamicin C1a of ~1.4-fold.[27] A similar approach could be coupled with the proposed generation of hybrid antimicrobials, to test a huge number of pathway variations. As we

are further improving on large and complex gene assemblies as time passes, use of a robotics platform would be ideal for generation of these hybrid pathways.

We anticipate that the panel of resistant strains we generated in this work may provide a useful platform for other heterologous expression experiments. A recombination directionality factor could be used with a serine integrase for the removal of the gene cluster from the strains shown to have the best antibiotic production phenotype.[28] This would then allow for the insertion of any gene cluster of interest. It could be of merit to determine whether these strains aid in heterologous production of other aminoglycosides or could be more broadly applied to heterologous production of other secondary metabolites of different classes. While we were not able to investigate any of the other gene clusters within *M.* sp. DEM32671 in this thesis, these 'superhost' *S. coelicolor* strains may be useful for further natural product production from clusters originating from this strain. The recent method proposed by Enghiad et al using Cas12a-assisted targeted cloning would allow for a more rapid testing pipeline which could yield other products of interest from *M.* sp DEM32671.[29]

Finally, further characterisation of the *M. abscessus* preQ1-II riboswitch is required before it can be fully utilised. Our preliminary results show a maximal dampening of expression of approximately ~37%, which does not currently allow for a total switch 'OFF' upon ligand induction. While the effect on gene expression will be dependent on the amino acid sequence downstream,[30] testing of different combinations of weaker promoters with the existing constructs could allow for the balancing required for a total switch on and off of gene expression. We hope that upon further tweaking, this candidate riboswitch could be utilised in situ for temporary switch-off of certain key genes, or alternatively for construction of basic genetic circuits in *S. coelicolor*.

These paths forward should establish a production platform for aminoglycosides, and – if yields can be further improved – this could represent a more rapid route towards production of target congeners than relying on the native *Micromonospora* production host.

1. S. J. McWilliam, D. J. Antoine, R. L. Smyth, and M. Pirmohamed, "Aminoglycoside-induced nephrotoxicity in children.," *Pediatric Nephrology*, vol. 32, no. 11, pp. 2015–2025, 2017, doi: 10.1007/s00467-016-3533-z.

2. E. Selimoglu, "Aminoglycoside-Induced Ototoxicity," *Current Pharmaceutical Design*, vol. 13, no. 1, pp. 119–126, 2007, doi: 10.2174/138161207779313731.

3. L. O. White, A. Lovering, and D. S. Reeves, "Variations in Gentamicin C1, C1a, C2, and C2a Content of Some Preparations of Gentamicin Sulphate Used Clinically as Determined by High-Performance Liquid Chromatography," *Therapeutic Drug Monitoring*, vol. 5, no. 1, pp. 123–126, 1983, doi: 10.1097/00007691-198303000-00014.

4. M. Kobayashi, M. Sone, M. Umemura, T. Nabeshima, T. Nakashima, and S. Hellström, "Comparisons of cochleotoxicity among three gentamicin compounds following intratympanic application," *Acta Oto-Laryngologica*, vol. 128, no. 3, pp. 245–249, 2008, doi: 10.1080/00016480701558948.

5. M. E. O'Sullivan *et al.*, "Dissociating antibacterial from ototoxic effects of gentamicin C-subtypes," *PNAS*, vol. 117, no. 51, pp. 32423–32432, 2020, doi: 10.1073/pnas.2013065117.

6. M. Ishikawa *et al.*, "Lower ototoxicity and absence of hidden hearing loss point to gentamicin C1a and apramycin as promising antibiotics for clinical use," *Scientific Reports*, vol. 9, no. 1, p. 2410, 2019, doi: 10.1038/s41598-019-38634-3.

7. J. W. Park *et al.*, "Genetic dissection of the biosynthetic route to gentamicin A2 by heterologous expression of its minimal gene set.," *PNAS*, vol. 105, no. 24, pp. 8399–404, 2008, doi: 10.1073/pnas.0803164105.

8. J. P. Gomez-Escribano and M. J. Bibb, "Engineering *Streptomyces coelicolor* for heterologous expression of secondary metabolite gene clusters," *Microbial Biotechnology*, vol. 4, no. 2, pp. 207–215, 2011, doi: 10.1111/j.1751-7915.2010.00219.x.

9. A. Rodríguez-García, P. Combes, R. Pérez-Redondo, M. C. A. Smith, and M. C. M. Smith, "Natural and synthetic tetracycline-inducible promoters for use in the antibiotic-producing bacteria *Streptomyces*," *Nucleic Acids Research*, vol. 33, no. 9, pp. e87–e87, 2005, doi: 10.1093/NAR/GNI086.

10. S. Li *et al.*, "Methyltransferases of gentamicin biosynthesis," *PNAS A*, vol. 115, no. 6, pp. 1340–1345, 2018, doi: 10.1073/pnas.1711603115.

11. F. Kudo and T. Eguchi, "Biosynthetic genes for aminoglycoside antibiotics," *The Journal of Antibiotics*, vol. 62, no. 9, pp. 471–481, 2009, doi: 10.1038/ja.2009.76.

12. D. Iwata-Reuyl, "An embarrassment of riches: the enzymology of RNA modification," *Current Opinion in Chemical Biology*, vol. 12, no. 2, pp. 126–133, 2008, doi: 10.1016/J.CBPA.2008.01.041.

13. P. J. McCown, J. J. Liang, Z. Weinberg, and R. R. Breaker, "Structural, functional, and taxonomic diversity of three preQ1 riboswitch classes," *Chemistry & Biology*, vol. 21, no. 7, pp. 880–889, 2014, doi: 10.1016/J.CHEMBIOL.2014.05.015.

14. B. Bilyk, L. Horbal, and A. Luzhetskyy, "Chromosomal position effect influences the heterologous expression of genes and biosynthetic gene clusters in Streptomyces albus J1074," *Microbial Cell Factories*, vol. 16, no. 1, 2017, doi: 10.1186/S12934-016-0619-Z.

15. M. J. Smanski, H. Zhou, J. Claesen, B. Shen, M. A. Fischbach, and C. A. Voigt, "Synthetic biology to access and expand nature's chemical diversity," *Nature Reviews Microbiology*, vol. 14, 2016, doi: 10.1038/nrmicro.2015.24.

16. M. J. Dunlop *et al.*, "Engineering microbial biofuel tolerance and export using efflux pumps," *Molecular Systems Biology*, vol. 7, no. 1, p. 487, 2011, doi: 10.1038/MSB.2011.21.

17. A. M. Miguez, M. P. McNerney, and M. P. Styczynski, "Metabolomics analysis of the toxic effects of the production of lycopene and its precursors," *Frontiers in Microbiology*, vol. 9, no. May, p. 760, 2018, doi: 10.3389/fmicb.2018.00760.

18. H.-J. Nah, H.-R. Pyeon, S.-H. Kang, S.-S. Choi, and E.-S. Kim, "Cloning and Heterologous Expression of a Large-sized Natural Product Biosynthetic Gene Cluster in Streptomyces Species," *Frontiers in Microbiology*, vol. 8, no. Mar, p. 394, 2017, doi: 10.3389/fmicb.2017.00394.

19. P. D. Schloss and J. Handelsman, "Status of the microbial census.," *Microbiology and Molecular Biology Reviews*, vol. 68, no. 4, pp. 686–91, 2004, doi: 10.1128/MMBR.68.4.686-691.2004.

20. M. S. Rappé and S. J. Giovannoni, "The Uncultured Microbial Majority," *Annual Review of Microbiology*, vol. 57, no. 1, pp. 369–394, 2003, doi: 10.1146/annurev.micro.57.030502.090759.

21. Y. Chang *et al.*, "Overproduction of gentamicin B in industrial strain *Micromonospora echinospora* CCTCC M 2018898 by cloning of the missing genes *genR* and *genS*," *Metabolic Engineering Communications*, vol. 9, p. e00096, 2019, doi: 10.1016/j.mec.2019.e00096.

22. Z. Wei, X. Shi, R. Lian, W. Wang, W. Hong, and S. Guo, "Exclusive Production of Gentamicin C1a from *Micromonospora purpurea* by Metabolic Engineering," *Antibiotics*, vol. 8, no. 4, 2019, doi: 10.3390/ANTIBIOTICS8040267.

23. X. Ni, Z. Sun, H. Zhang, H. He, Z. Ji, and H. Xia, "Genetic engineering combined with random mutagenesis to enhance G418 production in *Micromonospora echinospora*," *Journal of Industrial Microbiology and Biotechnology*, vol. 41, no. 9, pp. 1383–1390, 2014, doi: 10.1007/s10295-014-1479-3.

24. S. Li *et al.*, "Mechanistic Insights into Dideoxygenation in Gentamicin Biosynthesis," *ACS Catalysis*, vol. 11, no. 19, pp. 12274–12283, 2021, doi: doi.org/10.1021/acscatal.1c03508.

25. Y. Sun, X. Jian, and Z. Deng, "Preparation and applications of a series of novel aminoglycoside compounds, and construction method of high-yield engineering strain for producing novel aminoglycoside compounds," Chinese Patent CN106432382B, Apr. 01, 2019

26. S. R. Park, J. W. Park, Y. H. Ban, J. K. Sohng, and Y. J. Yoon, "2-Deoxystreptamine-containing aminoglycoside antibiotics: Recent advances in the characterization and manipulation of their biosynthetic pathways," *Natural Product Reports*, vol. 30, no. 1, pp. 11–20, 2013, doi: 10.1039/C2NP20092A.

27. Z. Wu, W. Gao, S. Zhou, Z. Wen, X. Ni, and H. Xia, "Improving gentamicin B and gentamicin C1a production by engineering the glycosyltransferases that transfer primary metabolites into secondary metabolites biosynthesis," *Microbiological Research*, vol. 203, pp. 40–46, 2017, doi: 10.1016/J.MICRES.2017.06.006.

28. F. J. Olorunniji, A. L. McPherson, S. J. Rosser, M. C. M. Smith, S. D. Colloms, and W. Marshall Stark, "Control of serine integrase recombination directionality by fusion with the directionality factor," *Nucleic Acids Research*, vol. 45, no. 14, pp. 8635–8645, 2017, doi: 10.1093/NAR/GKX567.

29. B. Enghiad *et al.*, "Cas12a-assisted precise targeted cloning using in vivo Cre-lox recombination," *Nature Communications*, vol. 12, no. 1, pp. 1–11, 2021, doi: 10.1038/s41467-021-21275-4.

30. T. Tuller and H. Zur, "Multiple roles of the coding sequence 5′ end in gene expression regulation," *Nucleic Acids Research*, vol. 43, no. 1, pp. 13–28, 2015, doi: 10.1093/NAR/GKU1313.