Understanding and exploiting MmyB-like transcriptional activators

By

Daniel Robertshaw

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Life Sciences

University of Warwick, Midlands Integrative Biosciences Training Partnership

May 2022

Table of Contents:

List of Figures

List of Tables

Acknowledgements

First of all the biggest thank you to my supervisor Dr Christophe Corre for not only giving me the opportunity to undertake this PhD but also for supporting me throughout these tough four and half years. I can't put into words how much I appreciate the guidance and patience he has shown me throughout this research. I'd like to thank my supervisor at Demuris Ltd, Dr Nick Allenby for the chance to have worked alongside him, unfortunately Covid decided to wasn't to be, which was a pity.

Some of this research would not have been possible without the help of Dr Lijiang Song. Without him, the LC-MS experiments wouldn't have happened and I'm extremely thankful to him for running my samples and providing the data that make up part of this thesis. My fellow PhD student Nazia Auckloo was also a huge help with the LC-MS and for providing me with some of the strains used my experiments.

I would like to extend my gratitude towards my supervisory panel, Professor Chris Dowson and Professor Elizabeth Wellington for their help and advice on my project.

To all the members of the Corre and Alberti groups both past and present, you were so welcoming throughout my time in the lab and gave me some great advice when needed. You all made it a memorable time of my life.

Finally the biggest thanks goes to my friends and family without who, I probably would never have been able finish.

The PhD was financially supported by the Midlands Integrative Biosciences Training Partnership and the UKRI Biotechnology and Biological Sciences Research Council.

Declaration and Inclusion of Material from a Prior Thesis

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been written by myself and has not been submitted in any previous application for any degree.

The work presented, including data generated and data analysis was carried out by the author except in the outlined scenarios below:

- The creation of the three *Streptomyces* mutants CO-121, CO-124 and CO-154 was completed by Nazia Auckloo
- The running of the LC-MS for data generation was performed by Dr Lijiang Song of the Mass Spectrometry Research Facility in the Chemistry department.

Abstract

One of the biggest ongoing challenges facing society and healthcare in the present is the prevalence of antimicrobial resistance. Novel sources of antibiotics are needed desperately to combat this crisis. A potential supply for this could spring from the MmyB-like transcriptional family, of which the methylenomycin cluster activator is a member of.

Three members of this group of activators were deregulated in *Streptomcyes coelicolor* M145 to investigate their prospective role in secondary metabolism that has been suggested through genome mining. These bacteria were RNA sequenced to study any effects on transcription and coupled with metabolic profile comparisons, which identified novel metabolites not seen before, alluding to some involvement of these genes in the secondary metabolism of *S. coelicolor.*

Bioinformatic analysis of the *Streptomyces coelicolor* MmyB-like activators was performed with support from the RNA sequencing to identify a potential conserved sequence within the promoters of this transcriptional family that may be the target of autoregulation.

Further members of this transcriptional family were identified in *Rhodococcus* species, a criminally understudied genus, with a high potential for secondary metabolism. Two of the newly discovered MmyB-like activators were overexpressed in *R. jostii* RHA1 and *R. erythropolis* PR4 and underwent metabolic comparison to look for secondary metabolites, although none were found.

This family of transcriptional activators needs to be studied further as the widespread presence of these genes across actinomycetes could be a catalyst for natural product discovery and the beginning of the fight back against antimicrobial resistance.

Abbreviations

AIDS	Acquired immunodeficiency syndrome
AMR	Antimicrobial resistance
antiSMASH	Antibiotics & Secondary Metabolite Analysis Shell
ATP	Adenosine triphosphate
BGC	Biosynthetic gene cluster
BPC	Base peak chromatogram
BLAST	Basic Local Alignment Search Tool
CDA	Calcium dependent antibiotic
CRISPR	Clustered regularly interspaced short palindromic repeats
DNA	Deoxyribonucleic acid
EIC	Extracted ion chromatogram
FIMO	Find Individual Motif Occurrences
HIV	Human immunodeficiency virus
HPLC	High performance liquid chromatography
LB	Lysogeny broth
LCMS	Liquid chromatography mass spectrometry
MEME	Motif Em for Motif Elicitation
MIBIG	Minimum Information about a Biosynthetic Gene Cluster
NCBI	National Centre for Biotechnology Information
NEB	New England Biolabs
NRPS	Non-ribosomal peptide synthetase
PCR	Polymerase chain reaction
PKS	Polyketide synthase
RiPP	Ribosomally synthesised and post-translationally modified peptides
RNA	Ribonucleic acid
rSAP	Recombinant shrimp alkaline phosphatase
SDR	Short chain dehydrogenase/reductase family
SFM	Soya flour mannitol

SHC	Supercoiling hypersensitive cluster
SMM	Supplemental Minimum Media
YEME	Yeast Extract Malt Extract

1. Introduction

1.1 Antimicrobials

The advent of modern day antimicrobial agents is arguably the greatest discovery in medicine and revolutionised the field. However, despite the successes these drugs have had, they have now developed into one of the biggest continuing challenges that healthcare faces[2].
The increasing prevalence of antimicrobial resistance (AMR) in bacteria, coupled with the dwindling discovery rate of new compounds since the 'golden age' means that untreatable infections are becoming much more widespread. An estimated 1.2 million people died in 2019 due to the direct result of AMR, almost as many fatalities attributed to HIV/AIDS and malaria combined[1]. The cost of the rise in AMR is not only taken against people's health but is predicted to soon cost the world economy billions of dollars a year[75].

Although the overuse and mishandling of antimicrobials is mainly responsible for the rise of AMR[2], there is no easy fix in sight. The over prescription of antibiotics to treat illness that does not warrant it and then the subsequent misuse by patients not fully completing their treatment has resulted in bacteria being exposed to low enough levels of the antibiotics to allow development of resistance mechanisms[76]. The classical treatments that were developed decades ago from natural sources are no longer as effective and although combinatorial treatment has helped, greater numbers of bacteria are acquiring resistance to multiple antibiotics[2]. This ease for bacteria to develop resistance due to being able evolve much quicker and some species being able to crosstalk with other bacteria to incorporate advantageous genetic material that contains resistance, means that current treatments are becoming obsolete[76].

The most common avenue at the moment by large pharmaceutical companies looking to develop new antibiotics is through small molecule synthetic chemical libraries being screened for activity against pathogens and promising leads being developed[78]. This approach while successful in improving current antibiotics, is predicated on the knowledge from those current drugs so a brand new antibiotic compound class that contains a unique backbone or

mechanism of action is unlikely[79]. Many of the libraries used are a result of chemical modification to existing natural product scaffolds, which while improving the efficacy, it still has the inherent same mode of action as the original natural product[80].

Along with better application of existing drugs and education as to their use, new classes of antimicrobials will be vital in the combatting the crisis being posed by multi-drug resistance micro-organisms. The big question is where will these will come from.

1.2 Natural Products

Natural products are compounds that are naturally found from environmental sources such as plants or microorganisms themselves. Natural products have always been a major source of antimicrobial drugs with many of those commercially available originating as natural products. The most famous example of this is the first modern antibiotic, penicillin, discovered by Alexander Fleming[3]. In the last couple of decades the emphasis on natural product programmes in pharmaceutical companies has slowly dwindled due to high cost and the uncertainty of investment returns and has largely taken a back seat to synthetic chemical libraries and high throughput screening due to their speed and relative success rate in producing drug leads[4]. However, natural compounds can be incredibly complex molecules due to millions of years of evolution which are difficult to access through synthetic chemistry. This complexity is the reason why natural products have never been fully discarded as novel antimicrobials and continue to be investigated[81].

1.2.1 Traditional Natural Product Discovery

Traditional bioactivity and culture based approaches in natural product discovery unfortunately have a high likelihood of discovering the precursors of many already developed antibiotics as well as previously encountered antibiotics[5]. Many of the ‘golden age’ antibiotics found used the technique of stressing the bacteria, by limiting different nutrients, in order to force the

production of antimicrobials that may be present in the genome. Once bioactivity was observed, any metabolites responsible for the activity would be isolated and identified. This is quite a time intensive method linked to the difficulties in creating the correct conditions for the bacteria to produce any antimicrobials[81].

There is also always the potential of redundant discovery. Many antimicrobials are found in multiple bacterial species across the environmental spectrum as the metabolites are successful and get passed throughout the spectrum of microorganisms. This is one of the reasons many companies gravitated towards synthetic libraries and combinatorial synthesis[6,78,79]. However, in the last couple of decades methodologies and tools have been developed along with the increased power of next-generation sequencing that has meant interest in natural products has been reignited with particular focus on the genome mining approach of natural product discovery[82].

1.2.2 Genome Mining Discovery

It has been known for many years that in prokaryotes, genes that are involved in the same metabolic process are closely grouped together[7]. These are known as biosynthetic gene clusters (BGCs). Many of the natural products shown to have antimicrobial properties are encoded in BGCs that contain all the genes necessary for the production of the molecule including; the regulatory proteins, the machinery that builds the active molecule, the resistance mechanism to protect the bacteria and the export system[86]. These BGCs can be put into different classifications mainly depending on the end product, the synthesis pipeline or the molecule's function[87]. The main two classes of BGCs that are of interest are the non-ribosomal peptide synthetases (NRPS) and the polyketide synthases (PKS). This is due to the historical evidence that many of the broadest range antibiotics that have been developed, originated from one of these two classes[8]. There are other classes of BGCs that are known to produce bioactive compounds such as ribosomally synthesised and post translationally modified peptides (RiPP), bacteriocins, lanthipeptides, siderophores and terpenoids. The wonderful thing about BGCs is

that the distinguishing characteristics of each class is conserved very strongly throughout the natural world allowing the advances in computing and sequencing to predict BGCs in genomes in a process known as genome mining[9,85].

Genome mining is one of the main bases for the renewed scientific interest in natural products towards combatting the antimicrobial resistance crisis[84]. The conserved domains present in the different class of BGCs has allowed the development of computer algorithms that have the ability to predict likely BGCs with their boundaries as well as their homology to known BGCs in brand new genomes that haven't been investigated[10, 86]. This prevents unnecessary time wasting on bacteria that may not contain any novel compounds along with the painstaking step by step experiments previously required to determine the length of the BGC[83].

This new genetic approach to natural product detection not only works for new bacterial genomes but ensures the full exploitation of previously thought exhausted sources[81]. The traditional culturing method didn't have the ability to cause bacteria to produce their full range of secondary metabolites. These cryptic BGCs weren't being expressed for many reasons, usually that particular conditions weren't being met to switch on their production. Genome mining is able to identify these previously inaccessible clusters for investigation[87].

Cryptic BGCs are potentially a good source of novel antibiotics. The difficulty in activating the cluster in a laboratory setting, hopefully would mean that it isn't prevalent in the environment, meaning resistance to the compound is unlikely. There is the potential as well that the core structure of the molecule is a completely new class of natural product antibiotic which hasn't happened since daptomycin in 1987[11]. However they are cryptic BGCs for a reason. Most often these clusters are tightly regulated at the DNA level with the traditional culturing methods being ineffective. Therefore to activate these sorts of clusters other methods are required, such as genome editing and heterologous expression which can be done in concert with one another.

Heterologous expression of cryptic BGCs has been used to discover the product of clusters previously. This approach is very popular in natural product experiments. This is because the heterologous hosts are usually much better understood and have been developed to focus on the metabolite production in greater amounts[12]. Problems can arise such as ligand based transcription factors being unable to function due to the ligand missing in the heterologous host. These can be mitigated by using a heterologous host that is found in the same genus as the original bacteria reducing the likelihood of such issues. Many bacteria are genetically intractable as well which necessitates the use of a heterologous host.

However, awakening cryptic BGCs can be done through genome editing, which when applicable can avoid some of the problems of heterologous expression. There are two main approaches in genome based activation of cryptic BGCs (Figure 1.1). Both of these approaches centre around the genes responsible for the control of the cluster's expression[12]. The first is to either delete or inactivate the gene(s) responsible for the repression of the cluster. With this action absent, the gene that activates expression will be free to begin transcription of the cluster machinery. The second approach is slightly simpler and is over expressing the activator gene. These can be achieved by either promoter replacement or by introducing an extra copy of the gene that is constitutively expressed. This overcomes the repressor gene(s) and activates the cluster. The advances in next generation sequencing and development of tools such as CRISPR has made these genome editing approaches easier than ever[14].

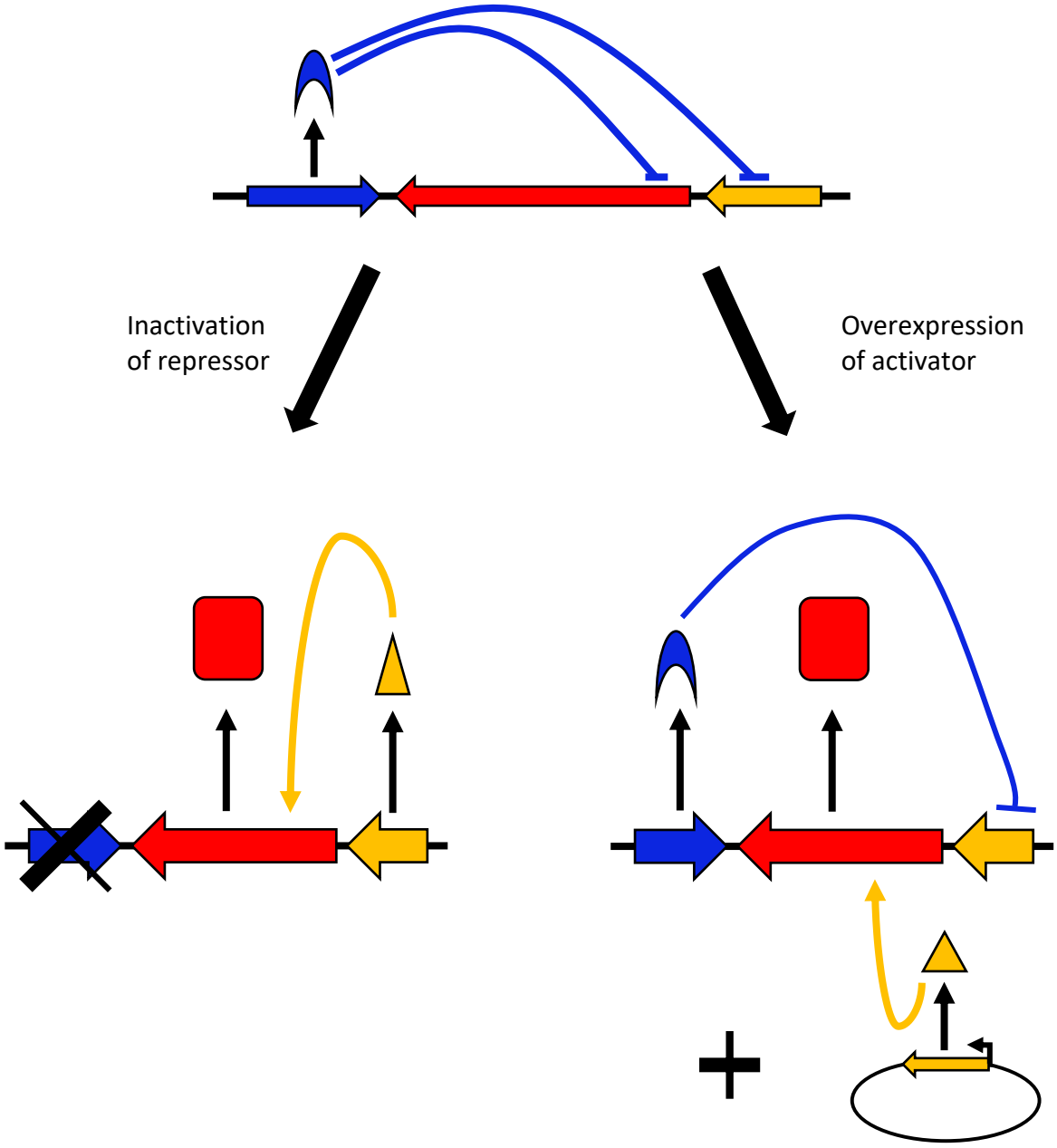


Figure 1.1: Methods of genome editing to cause expression of cryptic biosynthetic gene clusters. The repressor (blue) prevents expression of the cluster (red) or activator (yellow). By inactivating the repressor by deletion or mutation the cluster is free to be expressed. Over expressing the activator with a constitutive promoter overpowers the repression allowing expression of the cluster.

1.2.3 Omics use in Natural Product Discovery

The ability for genome mining to be a successful method for natural product discovery is due to the technology advances in different -omics techniques that allow for in depth detection of the changes that are a result of the genome editing. The use of transcriptomics and metabolomics allow for greater understanding of how these BGCs are working.

Transcriptomics is the study of the transcriptome or the sum of all the RNA. It is a great tool for investigations into the regulation of DNA and how changes introduced by researchers are altering the transcription of different genes. This can be used to identify which genes are being regulated by specific DNA regulators through genome editing (Figure 1.1). Transcriptomics is one of the recent beneficiaries to the better sequencing technologies available and has become a much more prevalent tool in recent years. In the context of natural product discovery, it is very powerful in identifying how regulation of BGCs can be influenced by specific genes[90]. Predictions by genome mining can expect genes to be transcriptional activators or repressors and gene editing experiments can confirm this but transcriptomics can identify which genes these regulators are targeting and allow much more nuanced control of BGC expression[90]. There are limitations to transcriptomics. The process from transcription to translation is extremely complex and not all mRNA is translated to the protein level. Its therefore not a fully indicative view of the outcome as the change in transcription may not result in a BGC producing the compound it encodes for. Therefore, transcriptomics has to be used in concert with other techniques such as metabolomics when used for natural product discovery.

Metabolomics is the study of the metabolome, the entirety of all the small molecule metabolites (<1.5kDa) which have been formed through cellular processes. It is one of the main tools used to identify natural products throughout history[88]. The use of liquid chromatography-mass spectrometry is an easy way to detect alterations in the metabolome and determine if gene editing experiments have had an influence on the secondary metabolism of the cell. The

increased use of metabolomics has led to the creation of platforms that make it easier than ever to identify new natural products[73,74]. Metabolomics does have weaknesses however. While it is a great tool to detect natural products, it does not give particularly in depth understanding of how this compound was produced and ended up secreted from the cell. It is a lot easier to build up from the bottom up than from the top down.

A combination of transcriptomics and metabolomics is a great way to cover the weaknesses of both sets of techniques when used for natural product discovery[91]. The metabolomics indicates if the end goal of detecting a novel natural product has been reached whereas the transcriptomics can provide insight into how the mechanism as to how the novel natural product was created and inform further experimentation. Altogether transcriptomics and metabolomics are a very complementary set of tools for natural product discovery.

1.3 Actinomycetes

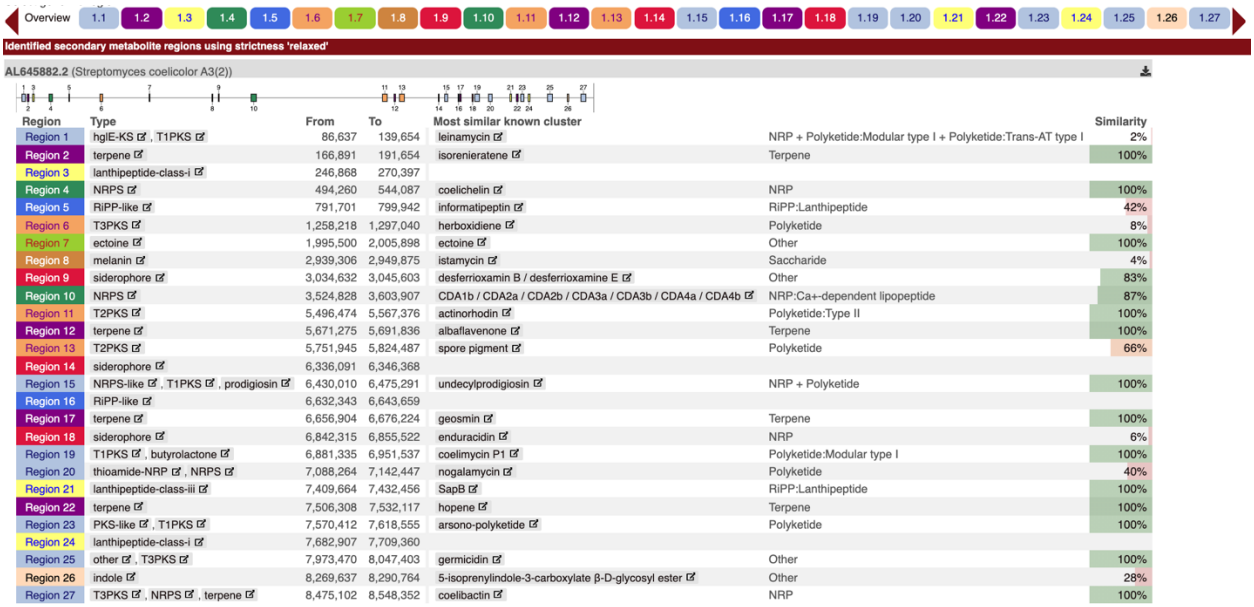
Actinomycetes are a group of filamentous, Gram positive bacteria that grow as mycelium, primarily found in soil although have been discovered across all types of environments[92]. Many of the species are similar to fungi, in that their life cycle involves not only hyphae growth but also sporulation under extreme conditions[16]. They are one of the most well studied and well known group of bacteria due to their history in being such a strong source of antimicrobials[81]. The majority of current commercial drugs were first discovered from an actinomycete and even in the modern day new antibiotic leads are still being found from this group of bacteria[81].

1.3.1 Streptomyces

The main source of natural products that have been viable for drug development has been the actinomycete genus *Streptomyces*[15]. *Streptomyces* bacteria are well known for the high GC content in their genome. They can be found in many environments, although most have been discovered in soil and have quite a complex life cycle compared to many bacteria. They

germinate from a single spore which allow them to withstand many hostile growth conditions and grow out as a base mycelium and aerial hyphae structure[16]. The genus has been discovered to produce a glut of specialised metabolites with a range of biological activities. A large number of these molecules have been antimicrobials that have been developed for clinical use such as streptomycin, tetracycline and chloramphenicol[17].

Due to the variety and number of different antimicrobials that have been found residing in different *Streptomyces* species, the genus is studied extensively with a lot of research focused on a model species, *Streptomyces coelicolor* A3(2) and the different engineered strains that have branched off from it. *S. coelicolor* was chosen due to its high genetic tractability compared to other *Streptomyces* and was one of the first genomes to be fully sequenced[18]. This sequencing showed a chromosomal genome of almost 8.7 Mb and two plasmids; SCP1 a linear plasmid about 365000 base pairs and SCP2 a circular plasmid roughly 31000 base pairs. Now before being sequenced, *S. coelicolor* was thought to close to exhausted as a source of novel compounds.. However initial analysis of the sequence revealed 18 new BGCs, some of which have been since identified. Current prediction software (Figure 1.2), now projects 27 total BGCs in the chromosomal genome with 6 of these showing similarity under 10% to known BGCs[19]. This just goes to show that natural product discovery still has vast untapped potential in extensively studied bacteria, let alone the many genera that have been ignored so far in this area of study.



Region	Type	From	To	Most similar known cluster		Similarity
Region 1	hglE-KS , T1PKS	86,637	139,654	leinamycin	NRP + Polyketide:Modular type I + Polyketide:Trans-AT type I	2%
Region 2	terpene	166,891	191,654	isorenieratene	Terpene	100%
Region 3	lanthipeptide-class-i	246,868	270,397			
Region 4	NRPS	494,260	544,087	coelichelin	NRP	100%
Region 5	RiPP-like	791,701	799,942	informatipeptin	RiPP:Lanthipeptide	42%
Region 6	T3PKS	1,258,218	1,297,040	herboxidiene	Polyketide	8%
Region 7	ectoine	1,995,500	2,005,898	ectoine	Other	100%
Region 8	melanin	2,939,306	2,949,875	istamycin	Saccharide	4%
Region 9	siderophore	3,034,632	3,045,603	desferrioxamin B / desferrioxamine E	Other	83%
Region 10	NRPS	3,524,828	3,603,907	CDA1b / CDA2a / CDA2b / CDA3a / CDA3b / CDA4a / CDA4b	NRP:Ca+-dependent lipopeptide	87%
Region 11	T2PKS	5,496,474	5,567,376	actinorhodin	Polyketide:Type II	100%
Region 12	terpene	5,671,275	5,691,836	albaflavenone	Terpene	100%
Region 13	T2PKS	5,751,945	5,824,487	spore pigment	Polyketide	66%
Region 14	siderophore	6,336,091	6,346,368			
Region 15	NRPS-like , T1PKS , prodigiosin	6,430,010	6,475,291	undecylprodigiosin	NRP + Polyketide	100%
Region 16	RiPP-like	6,632,343	6,643,659			
Region 17	terpene	6,656,904	6,676,224	geosmin	Terpene	100%
Region 18	siderophore	6,842,315	6,855,522	enduracidin	NRP	6%
Region 19	T1PKS , butyrolactone	6,881,335	6,951,537	coelimycin P1	Polyketide:Modular type I	100%
Region 20	thioamide-NRP , NRPS	7,088,264	7,142,447	nogalamycin	Polyketide	40%
Region 21	lanthipeptide-class-iii	7,409,664	7,432,456	SapB	RiPP:Lanthipeptide	100%
Region 22	terpene	7,506,308	7,532,117	hopene	Terpene	100%
Region 23	PKS-like , T1PKS	7,570,412	7,618,555	arsono-polyketide	Polyketide	100%
Region 24	lanthipeptide-class-i	7,682,907	7,709,360			
Region 25	other , T3PKS	7,973,470	8,047,403	germicidin	Other	100%
Region 26	indole	8,269,637	8,290,764	5-isoprenylindole-3-carboxylate β-D-glycosyl ester	Other	28%
Region 27	T3PKS , NRPS , terpene	8,475,102	8,548,352	coelibactin	NRP	100%

Figure 1.2: The output of the *Streptomyces coelicolor* A3(2) chromosomal genome after processing by antiSMASH. The software predicts 27 biosynthetic gene clusters along with the cluster type and similarity to known clusters. Image created using reference 19.

1.3.1.1 Methylenomycin

One of the gene clusters in *S. coelicolor* A3(2) discovered before the genome was fully sequenced was the methylenomycin cluster[21]. This is found on the linear plasmid SCP1 and is responsible for the production of methylenomycin. Methylenomycin is both a Gram positive and negative antibiotic used by *S. coelicolor* and production can be triggered when undergoing acidic shock or limiting alanine conditions[20].

After the full genome sequence of was released this BGC was investigated using a genetic approach which discovered the regulation of this gene cluster is quite interesting. The cluster has autoregulatory molecules encoded within itself and the methylenomycin biosynthetic genes are under the control of a single transcriptional activator, MmyB[21].

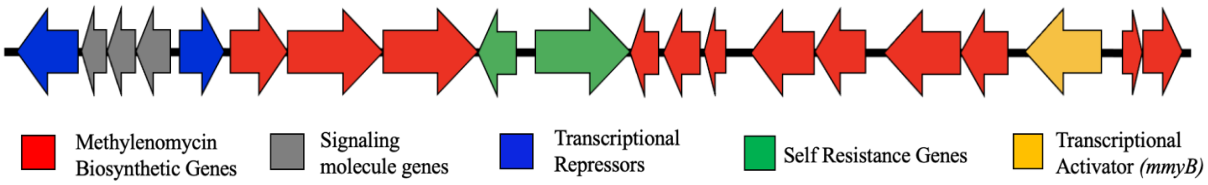


Figure 1.3: The gene cluster responsible for the production of methylenomycin in *Streptomyces coelicolor* A3(2)

The cluster itself, seen in Figure 1.3, is made up of 21 genes with 5 main functions. Three of the genes, *mmfL, mmfH and mmfP,* are involved in the production of the methylenomycin furan molecules, labelled as signalling molecule genes, that are involved in the regulation of the cluster. These furan molecules interact with the transcriptional repressors, *mmyR and mmfR,* that prevent the production of the cluster transcriptional activator, *mmyB,* releasing the promoter site to allow *mmyB* transcription. Two of the other genes, *mmyJ and mmr,* are responsible for the bacteria's resistance to the end product methylenomycin, while the rest of the genes encode the machinery that builds the methylenomycin molecule[21].

It is believed that *mmyB* has a positive feedback loop of expression, being able to activate its own transcription, seen in Figure 1.4. Therefore, the repression of the cluster is much easier to overcome. Previous work in the group has shown the repression can be disabled by introducing an extra copy of *mmyB* under a constitutive promoter into the bacteria[22].

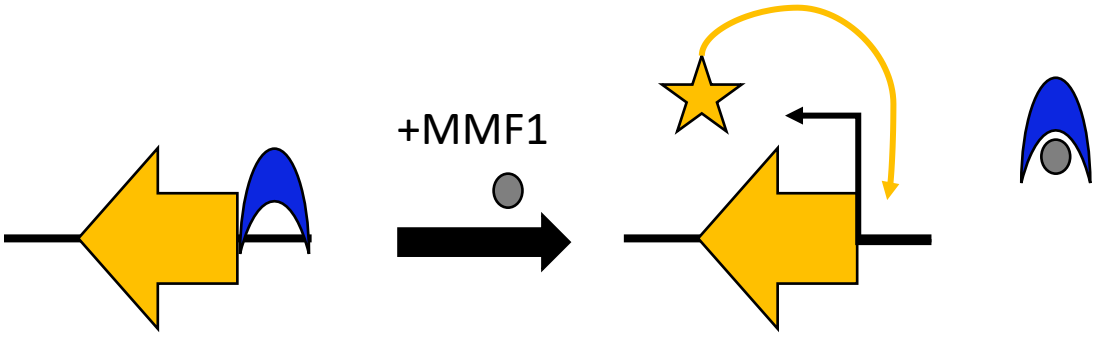


Figure 1.4: Proposed regulation of the *mmyB* gene. The *mmyB* (yellow) gene is repressed by the mmyR/mmfR dimer (blue). The methylenomycin furans (grey) bind to the dimer when expressed releasing the *mmyB* promoter. This can then be expressed and switch on the cluster as well as its own expression

1.3.1.2 MmyB

The MmyB protein that is responsible for the regulation of the methylenomycin BGC is a promising target for natural product discovery. It has been discovered to have numerous homologs that are found predominantly in actinomycetes, the source of many antimicrobials. The crystal structure of MmyB has never been elucidated but predictive models have been used to identify that the transcription factor has two domains. One that is responsible for binding to the DNA while the other has a ligand recognition site that, making the transcription factor ligand dependent. A MmyB homolog, MltR found in *Chloroflexus aurantiacus* has been crystallised and its structure was able to confirm the prediction for MmyB with it having a Xre-type DNA binding N terminal and a ligand binding PAS-like C terminal domain[23](Figure 1.5).

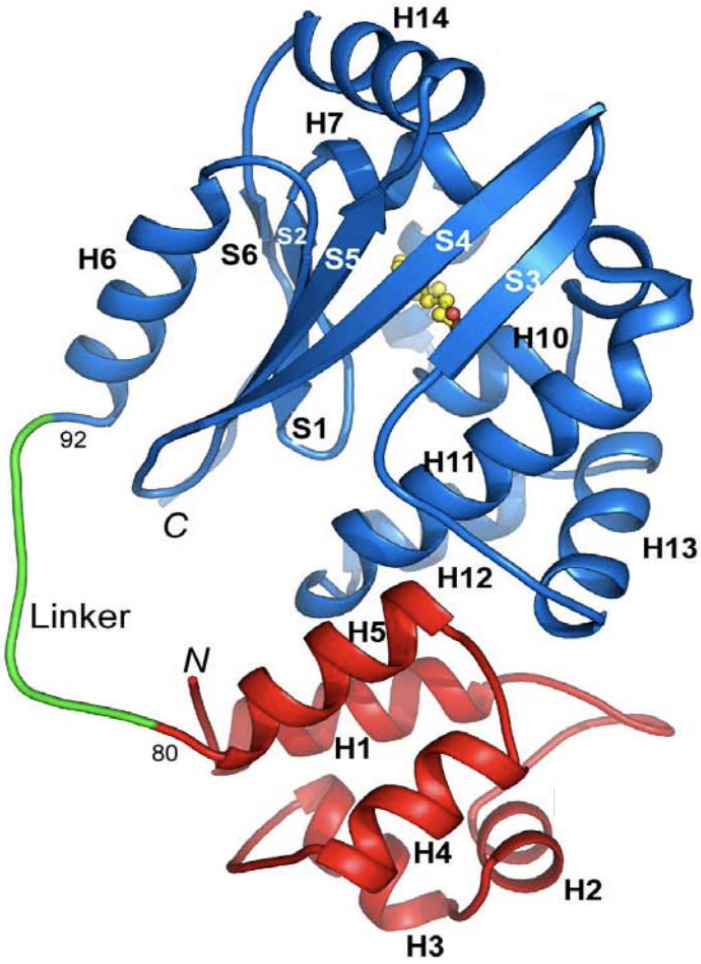


Figure 1.5: Crystal structure of the MmyB-like gene MltR, elucidated in *Chloroflexus aurantiacus.* The protein contains two domains, a DNA binding domain (red) and a ligand binding module (blue) joined by a linker sequences (green). Ball and stick representation of co-crystallised myristic acid within ligand binding module. Adapted from reference 23.

This proposed crystal structure was used to identify sixteen MmyB homologs in the *S. coelicolor* A3(2) genome (Table 1.1) that could be involved in fatty acid or antimicrobial BGC regulation.

MmyB homolog	Distance to nearest cluster prediction (kb)	Nearest cluster prediction
sco0110	Within prediction	Region 1
sco0233	24.3	Region 3
sco0236	21.6	Region 3
sco0307	32.7	Region 3
sco0891	136.5	Region 5
sco2501	239.7	Region 8
sco2537	203.1	Region 8
sco4680	385.7	Region 11
sco4944	117.7	Region 11
sco6539	90.3	Region 20
sco6676	Within prediction	Region 21
sco6926	Within prediction	Region 23
sco7140	38.1	Region 25
sco7706	Within prediction	Region 27
sco7767	48.2	Region 27
sco7817	88.0	Region 27

Table 1.1: MmyB homologs discovered within *Streptomyces coelicolor* A3(2) chromosomal genome alongside their proximity to antiSMASH predicted biosynthetic gene clusters.

The regulatory system of methylenomycin has been exploited previously to discover novel natural products. Homologs of the repressor and furan gene cassette (mmfR-mmfLHP-mmyR) were found in *Streptomyces venezuelae* ATCC10712, with the deletion of the mmyR homolog leading to the discovery of gaburedins[24]. The success of this exploitation of homologs to methylenomycin regulation and the proximity of some MmyB homologs in *S. coelicolor* (Table

1.1) to biosynthetic genes gives the hope that one or more of these could be involved in the regulation of a cryptic gene cluster yet to be characterised.

There are three MmyB homologs of specific interest. The *sco0236* and *sco6926* genes are in such close proximity to the two uncharacterised lanthipeptide clusters, in the case of *sco6926* actually within the cluster prediction, it is reasonable to assume they may be involved in some way with these clusters. Previous experiments have shown that over expression of *sco69626* is able to cause some bioactivity against *Bacillus subtilis*[22]. An interesting property that both *sco0236* and *sco7140* have is that when overexpressed, the extract has shown increased autophagosome accumulation in *Drosphila* flies. This could indicate the triggering of autophagy (cell death) in the flies by these extracts (I. Nezis March 2019 Personal Communication).

1.3.2 *Rhodococcus*

Although the *Streptomyces* genus has been a favourite of the scientific community for bioactivity, closely related genera that potentially could have significant secondary metabolism capabilities have generally fallen by the wayside. One such genus is the *Rhodococcus* bacteria. These are aerobic Gram positive bacteria with mycolic acids in their cell membrane, similar to *Mycobacterium* and have a high GC content genome. *Rhodococcus* can be found in a variety of environments, mainly soil and water and unlike *Streptomyces*, they are non-sporulating, non-motile actinomycetes and grow as rods, cocci and hyphae[25].

The genus is well known to the scientific community in regards to their degradation abilities and that has been the large majority of research focus along with two species, *Rhodococcus equi* and *Rhodococcus fascians* that are a horse and plant pathogen respectively. Across the *Rhodococcus* genus the genomes encode the ability to breakdown a huge variety of compounds many quite complex. These include alkanes, aromatic and cyclic organic molecules, even halogenated metabolites, many with potential microbial toxicity that other bacteria wouldn't be able to tolerate[26]. All these degradation targets are of industrial interest due to how often

these sorts of compounds can be found as waste by-products of industrial chemical reactions and in environment treatment plants[27,28]. This has been where a lot of the research has been centred with the goal of using *Rhodococcus* as bio-degraders for industrial environments as part of the chemical processes[29].

Recently though, the boom in next generation sequencing and the need for new antibiotics has led to a different emphasis towards *Rhodococcus* appearing and more research is starting to centre on whether the secondary metabolism potential of the genus is worth looking in to. There have been previous discoveries of bioactive compounds from *Rhodococcus* bacteria. These have mainly found Gram positive antimicrobials that are particular active against other *Rhodococcus* and *Mycobacterium* species as well as anti-fungal molecules, none of which have been developed for human use [30,31,32,33,34]. The activity of these compounds is unsurprising due to the prevalence of these microorganisms in the environments that *Rhodococcus* are native to. However, the fact that no commercial antimicrobials have been developed from *Rhodococcus* does not mean that it isn't a potential well of unexploited novel antimicrobials. In fact the ability of *Rhodococcus* bacteria to degrade such a wide variety of compounds and survive compounds with higher potential microbial toxicity could indicate that any secondary metabolites are types that wouldn't be viable in other bacteria and are a class yet to be seen.

The secondary metabolism potential of the *Rhodococcus* genus was analysed and discovered a wealth of predicted biosynthetic gene clusters (Figure 1.6)[35].

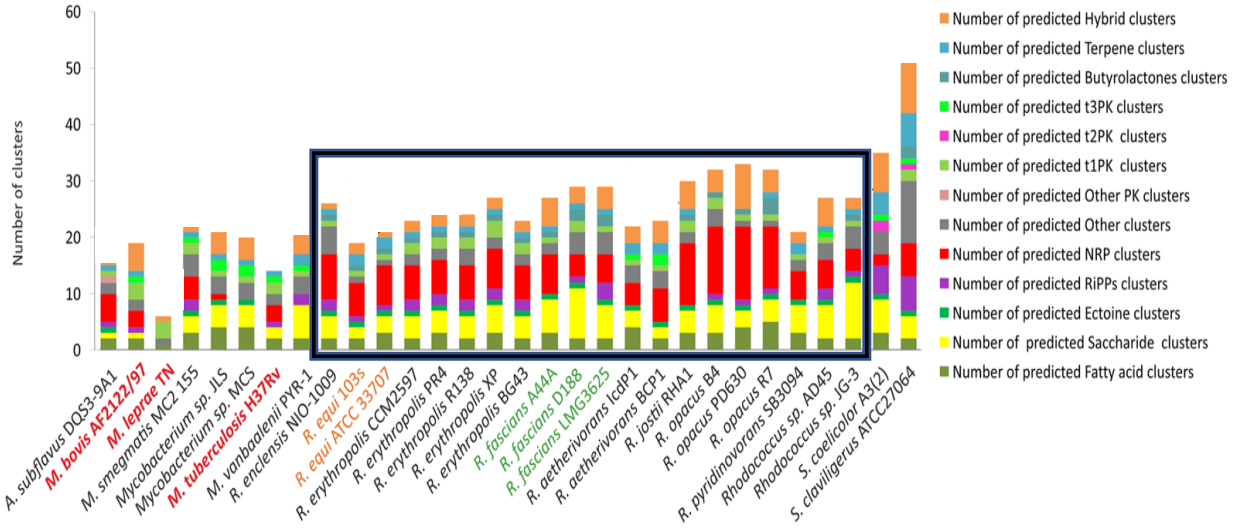
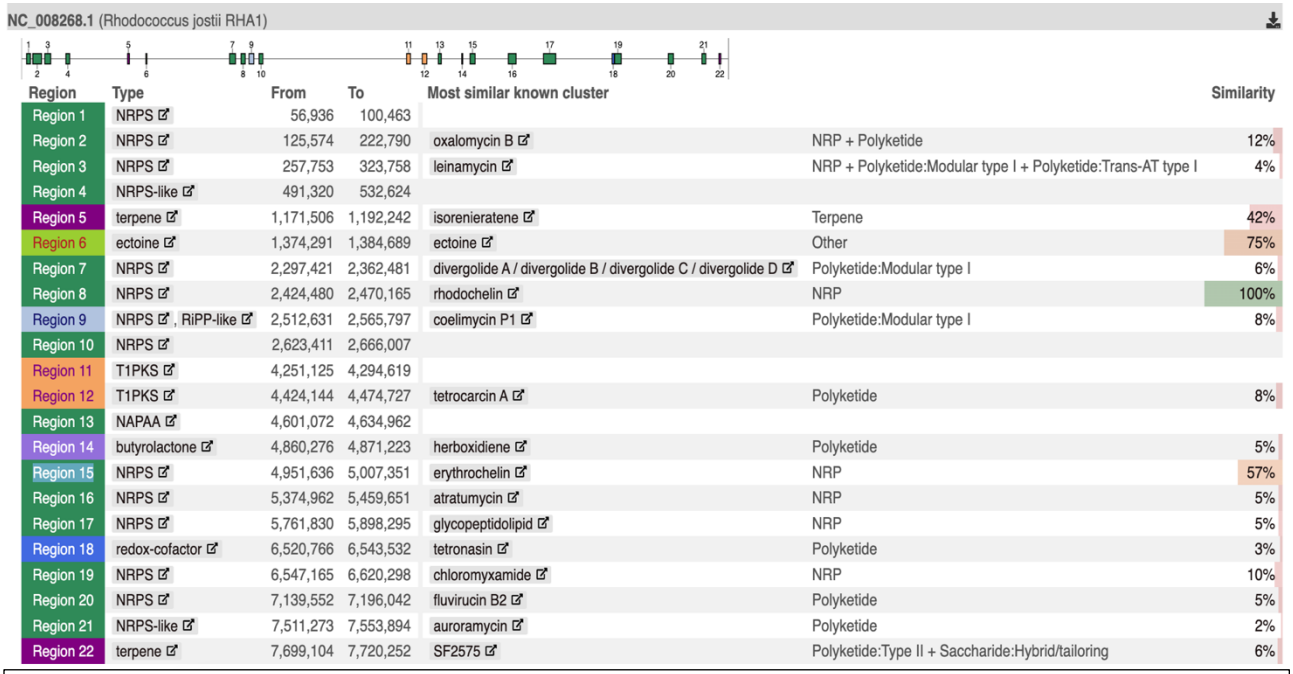


Figure 1.6: The known types of biosynthetic gene clusters predicted by antiSMASH for a selection of different actinomycete bacteria from the genera *Actinobacteria, Mycobacterium, Rhodococcus* and *Streptomyces* with the *Rhodococcus* bacteria reults highlighted by the box. Each colour corresponds to a different type of cluster. Adapted from reference 35.

The high average proportion of NRPS clusters compared to other actinomycetes is of real interest. This type of cluster is the source of many developed antibiotics used in medicine so the large number present in *Rhodococcus* that are uncharacterised are an attractive potential source.

The most well studied *Rhodoccus* bacteria is the *Rhodococcus jostii* RHA1 strain and it is a prime example as to the promise that *Rhodococcus* has in natural product discovery[36]. The core genome on its own contains a prediction of 22 BGCs (Figure 1.7). Fourteen of these are classed as NRPS type clusters, of which only three have known NRPS genes, rhodochelin[33], atratumycin[37] and erythrochelin[38]. The other eleven NRPS clusters are complete mysteries and with how prominent NRPS clusters are in the different natural products that have been developed into commercial drugs, all eleven could in the best case scenario encode brand new classes of antibiotics. The other eight clusters consist of two type I PKS, a butyrolactone, a redox cofactor, a RiPP-like cluster, two terpenes and the ectoine cluster. Of these, only one terpene has any degree of similarity to a known cluster, isorenieratene, and the ectoine cluster has

previously been studied. It still leaves six unknown clusters on top of the eleven NRPS that could all have the potential to be developed into antibiotics.



NC_008268.1 (Rhodococcus jostii RHA1)

Region	Type	From	To	Most similar known cluster		Similarity
Region 1	NRPS	56,936	100,463			
Region 2	NRPS	125,574	222,790	oxalomycin B	NRP + Polyketide	12%
Region 3	NRPS	257,753	323,758	leinamycin	NRP + Polyketide:Modular type I + Polyketide:Trans-AT type I	4%
Region 4	NRPS-like	491,320	532,624			
Region 5	terpene	1,171,506	1,192,242	isorenieratene	Terpene	42%
Region 6	ectoine	1,374,291	1,384,689	ectoine	Other	75%
Region 7	NRPS	2,297,421	2,362,481	divergolide A / divergolide B / divergolide C / divergolide D	Polyketide:Modular type I	6%
Region 8	NRPS	2,424,480	2,470,165	rhodochelin	NRP	100%
Region 9	NRPS , RiPP-like	2,512,631	2,565,797	coelimycin P1	Polyketide:Modular type I	8%
Region 10	NRPS	2,623,411	2,666,007			
Region 11	T1PKS	4,251,125	4,294,619			
Region 12	T1PKS	4,424,144	4,474,727	tetrocarcin A	Polyketide	8%
Region 13	NAPAA	4,601,072	4,634,962			
Region 14	butyrolactone	4,860,276	4,871,223	herboxidiene	Polyketide	5%
Region 15	NRPS	4,951,636	5,007,351	erythrochelin	NRP	57%
Region 16	NRPS	5,374,962	5,459,651	atratumycin	NRP	5%
Region 17	NRPS	5,761,830	5,898,295	glycopeptidolipid	NRP	5%
Region 18	redox-cofactor	6,520,766	6,543,532	tetronasin	Polyketide	3%
Region 19	NRPS	6,547,165	6,620,298	chloromyxamide	NRP	10%
Region 20	NRPS	7,139,552	7,196,042	fluvirucin B2	Polyketide	5%
Region 21	NRPS-like	7,511,273	7,553,894	auroramycin	Polyketide	2%
Region 22	terpene	7,699,104	7,720,252	SF2575	Polyketide:Type II + Saccharide:Hybrid/tailoring	6%

Figure 1.7: The output of *Rhodococcus jostii* RHA1 chromosomal genome after processing by antiSMASH. The software predicts 22 biosynthetic gene clusters and estimates their similarity to known clusters – created using reference 19

With just seeing the untapped potential that *R. jostii* RHA1 contains for novel secondary metabolism, it gives hope that the antibiotic resistance crisis can be combatted by the reignited interest in natural product discovery. There's such a vast well of bacteria criminally under studied in this area of science. It is almost impossible to imagine that not one of the other actinomycetes that are less studied, does not share at least some of the same capabilities for secondary metabolism as *Streptomyces.* Perhaps *Rhodococcus* is that genus that could explode and begin a brand new golden age of antibiotic discovery to rival that of the 20th century.

1.4 Aims and Objectives

The prevalence and spread of MmyB homologs across actinomycetes along with the likelihood that some are involved in natural product biosynthesis makes them an attractive study area.

Being able to further understand the potential binding sites these transcriptional regulators have could lead to exploiting any that are involved in the regulation of secondary metabolism and potentially lead to novel natural product discovery. The investigations into *Streptomyces* using transcriptomics will allow for greater understanding of the way MmyB homologs are able to influence natural product biosynthesis, while metabolomics will indicate if overexpression of the homologs is able to promote activation of BGCs. Experiments focusing on *Rhodococcus* will allow for an indication of the potential this bacteria genus might have to be a new source of natural products for novel antimicrobials and help combat the rising AMR crisis.

The following objectives will be attempted:

- Genome mining *Rhodococcus* bacteria for MmyB homologs
- Bioinformatic analysis of MmyB homologs to predict potential binding sites
- Overexpression of MmyB homologs in both *S. coelicolor* and *Rhodococcus* bacteria to attempt to overcome inherent BGCs repression
- RNA sequencing of *S. coelicolor* mutants to identify areas of action by MmyB homologs
- Metabolomic analysis to identify any new metabolites from *S. coelicolor* and *Rhodococcus* MmyB homolog overexpression

2. Materials & Methods

2.1 Materials

2.1.1 Microbial Strains

Strain	Uses	Source
Escherichia coli		
DH5α	General cloning host	New England Biolabs
Top10	General cloning host	New England Biolabs
ET12567/pUZ8002	Non-methylating host pUZ8002 – non transmissible plasmid for conjugation	John Innes Centre, Norwich, UK
Rhodococcus		
j*ostii* RHA1	Wild type strain	DSMZ, Braunschweig, Germany
erythropolis PR4	Wild type strain	DSMZ, Braunschweig, Germany
CO-213	RHA1 with integrated pOSV556 plasmid	This study
CO-218	RHA1 with integrated pC-0256 plasmid	This study
CO-222	RHA1 with integrated pC-0260 plasmid	This study
CO-231	PR4 with integrated pOSV556 plasmid	This study
CO-236	PR4 with integrated pC-0256 plasmid	This study
CO-240	PR4 with integrated pC-0260 plasmid	This study
Streptomyces coelicolor		

M145	Derivative of *Streptomyces coelicolor* A3(2) (SCP1-, SCP2-)	John Innes Centre, Norwich, UK
CO-121	M145 with integrated pC-0020 plasmid	Nazia Auckloo, Corre group, University of Warwick, UK
CO-124	M145 with integrated pC-0018 plasmid	Nazia Auckloo, Corre group, University of Warwick, UK
CO-154	M145 with integrated pC-0016 plasmid	Nazia Auckloo, Corre group, University of Warwick, UK
CO-250	M145 with integrated pOSV556 plasmid	This study

Table 2.1: List of microbial strains used in this study

2.1.2 Plasmids

Name	**Notes**	**Source**
pOSV556	ampR, hygR, ermE*, oriT, pSAM2 integrase	Pernodet Group, University of Paris-sud, France
pC-0016	pOSV556 derivative with sco6926 cloned in	Nazia Auckloo, Corre group, University of Warwick, UK
pC-0018	pOSV556 derivative with sco0236 cloned in	Nazia Auckloo, Corre group, University of Warwick, UK
pC-0020	pOSV556 derivative with sco7140 cloned in	Nazia Auckloo, Corre group, University of Warwick, UK
pC-0252 (pDR0005)	pOSV556 derivate with O5Y_RS25605 cloned in via XhoI and HindIII	This study
pC-0253 (pDR007)	pOSV556 derivative with	This study

	RER_RS29060 cloned in via XhoI and HindIII	
pC-0254 (pDR008)	pOSV556 derivative with RHA1_RS02730 cloned in via PstI and StuI	This study (Genewiz)
pC-0255 (pDR009)	pOSV556 derivative with RHA1_RS02780 cloned in via PstI and XhoI	This study (Genewiz)
pC-0256 (pDR010)	pOSV556 derivative with RHA1_RS09515 cloned in via PstI and StuI	This study (Genewiz)
pC-0257 (pDR011)	pOSV556 derivative with RHA1_RS40960 cloned in via HindIII and XhoI	This study (Genewiz)
pC-0258 (pDR012)	pOSV556 derivative with PD630_RS08950 cloned in via PstI and XhoI	This study (Genewiz)
pC-0259 (pDR013)	pOSV556 derivative with PD630_RS21525 cloned in via PstI and StuI	This study (Genewiz)
pC-0260 (pDR014)	pOSV556 derivative with PD630_RS21650 cloned in via PstI and XhoI	This study (Genewiz)
pC-0261 (pDR015)	pOSV556 derivative with PD630_RS28605 cloned in via PstI and StuI	This study (Genewiz)
pC-0262 (pDR016)	pOSV556 derivative with PD630_RS34220 cloned in via PstI and XhoI	This study (Genewiz)

pC-0263 (pDR017)	pOSV556 derivative with PD630_RS34435 cloned in via HindIII and PstI	This study (Genewiz)
pC-0264 (pDR018)	pOSV556 derivative with PD630_RS34505 cloned in via PstI and XhoI	This study (Genewiz)
pC-0265 (pDR019)	pOSV556 derivative with PD630_RS34730 cloned in via PstI and XhoI	This study (Genewiz)
pC-0266 (pDR020)	pOSV556 derivative with R1CP_RS19215 cloned in via PstI and XhoI	This study (Genewiz)
pC-0267 (pDR021)	pOSV556 derivative with R1CP_RS28630 cloned in via PstI and StuI	This study (Genewiz)

Table 2.2: List of plasmids used in this study.

2.1.3 Primers

All primers were diluted in sterile dH_2O to a concentration of 100 mM

Primer Name	Primer Sequence	Target	Expected size (bp)
CCM01	CTCGAGTAACGCATCTGAGCTGAT	*O5Y_RS25605*	1002
CCM04	AAGCTTCTGTTGTTCCAGGTAAC		
PR409	CTCGAGGAACCATCGATCGGAT	*RER_RS29060*	1168
PR410	AAGCTTGCTTGGAAGTCGACGT		
pOSV01	AATCCTGTATATCGTGCGAA	*ermE** cloning site	

pOSV02	CTTGTTATCGCAATAGTTGG		Dependent on gene insert
pDR005 01	TGATGGCGTCGAGTAT	*O5Y_RS25605*	722
pDR005 02	CTCATCGGTGAACTTGC		547
pDR007 01	TAGTCGATGCTCACACC	*RER_RS29060*	595
pDR007 02	CATCGGACGGGATTCAA		568
pDR008 01	GTGATCATCGCGTCGA	*RHA1_RS02730*	698
pDR008 02	CCGCCTTCATCTTCCTC		609
pDR009 01	CGTAGTAGTCGACCGAGA	*RHA1_RS02780*	500
pDR009 02	CATGCTCCGACTCGAA		561
pDR010 01	GACAACAGGTAGAGATGG	*RHA1_RS09515*	599
pDR010 02	GAGAGCAGCACCGAAT		446
pDR011 01	GCCAGATTGAATAGGTG	*RHA1_RS40960*	573
pDR011 02	CGTGTTGCGCTGGTTCTT		558
pDR012 01	GCGAATAGAAGGCGTAG	*PD630_RS08950*	784
pDR012 02	TCAACTTCGTCCGGTT		650
pDR013 01	AGAGTGAACAGGTGCTC	*PD630_RS21525*	599
PDR013 02	ACATCGTCGCGGTCAAT		688
pDR014 01	AAGAGGTGGTCGCGTT	*PD630_RS21650*	605
pDR014 02	CGACTACGACACCATC		594
pDR015 01	TGGTCCAGAACCGCAT	*PD630_RS28605*	704
pDR015 02	GCGTGTAGTAGTCCGT		429
pDR016 01	TGCGAGGTCGAACAGAT	*PD630_RS34220*	606
pDR016 02	TACGTCAACTGGGAACG		620
pDR017 01	CGAGGACCAGTTGCACA	*PD630_RS34435*	680
pDR017 02	ACAACTCATCGAGCTGA		569
pDR018 01	CGATGCTGTCGAGTATT	*PD630_RS34505*	550
pDR018 02	AGAATCCGCTCGACAAG		532

pDR019 01	CTTCGTCGAGTTGTAGA	*PD630_RS34730*	577
pDR019 02	TTCGTCTTCCTCAGTC		614
pDR020 01	TTGGCGGCGTCATAGA	*R1CP_RS19215*	563
pDR020 02	GTCAACTTCGTCCGGTT		647
pDR021 01	CAACAGGTAGAGATGGTC	*R1CP_RS28630*	597
pDR021 02	CTGAAGACGATCACGA		556

Table 2.3: List of primers used in this study. All pDR- 01 primers are paired with pOSV 02 and pDR- 02 primers are paired with pOSV 01

2.1.4 Antibiotic Solutions

Antibiotic	Stock Solution (mg/mL)	E. coli (µg/mL)	Rhodococcus (µg/mL)	Streptomyces (µg/mL)
Ampicillin	100	100	-	-
Hygromycin	50	-	150	150
Nalidixic Acid	25	25	-	25

Table 2.4: List of antibiotics used in this work.

2.1.5 Solid Media

<u>Lysogeny Broth (LB) Agar</u>

Tryptone	10g
Yeast Extract	5g
NaCl	10g
Agar	15g
dH_2O	make up to 1L

Yeast Extract Malt Extract (YEME) Agar

Yeast Extract	3g
Malt Extract	3g
Bactopeptone	5g
Glucose	10g
Agar	15g
dH_2O	make up to 1L

Soya Flour Mannitol (SFM) Agar

Soya Flour	20g
Mannitol	20g
Agar	20g
Tap water	make up to 1L

2.1.6 Liquid Media

LB Medium

Tryptone	10g
Yeast Extract	5g
NaCl	10g
dH_2O	make up to 1L

YEME Medium

Yeast Extract	3g
Malt Extract	3g
Bactopeptone	5g
Glucose	10g
dH_2O	make up to 1L

2xYT Medium

Tryptone	NaCl
Yeast Extract	dH_2O
16g	5g
10g	make up to 1L

Supplemented Minimal Medium (SMM)

Casamino acids	2g
TES buffer	8.68g
dH_2O	make up to 1L

After autoclave add the following sterilised solutions:

$NaH_2PO_4 + K_2HPO_4$	
(50mM each)	10mL
$MgSO_4$ (1M)	5mL
Glucose (50% w/v)	18mL
Trace element solution	1mL

Trace Element Solution

$ZnSO_4.7H_2O$	0.1g
$FeSO_4.7H_2O$	0.1g
$MnCl_2.4H_2O$	0.1g
$CaCl_2.6H_2O$	0.1g
NaCl	0.1g
dH_2O	make up to 1L

2.2 Methods

2.2.1 Bioinformatics analysis

The *Rhodococcus* bacteria MmyB homologs were identified using the Basic Local Alignment Search Tool (BLAST). The MmyB gene (NCBI accession: NC_003903.1) was compared to all NCBI uploaded *Rhodococcus* genomes

The genomes of *Rhodococcus* bacteria and *Streptomyces coelicolor* M145 (NCBI accession: AL645882.2) were analysed by antiSMASH 6.0.1 (https://antismash.secondarymetabolites.org/) to predict the biosynthetic gene clusters responsible for secondary metabolite production. The identified clusters were then further annotated in *S. coelicolor* M145 using BLAST.

Phylogeny trees of *Streptomyces coelicolor* and *Rhodococcus* MmyB homologs were created using the Clustal Omega multiple sequence alignment tool provided by EMBL European Bioinformatics Institute (https://ebi.ac.uk/Tools/msa/clustalo/) with default settings[70].

The potential binding sites of MmyB homologs were discovered using the Motif Em for Motif Elicitation (MEME) tool from MEME Suite Version 5.4.1 (https://meme-suite.org/)[63]. Promoters of MmyB homologs in *S. coelicolor* M145 (500bp upstream of ORF start codon) were submitted to MEME with the programme enriching for up to 20 motifs using all the default parameters provided. The top motifs identified by MEME were further investigated with submission to Find Individual Motif Occurrences (FIMO) using default parameters. The FIMO output then underwent comparison the RNA sequencing data of CO-121, CO-124 and CO-154.

2.2.2 *Escherichia coli* bacteria

2.2.2.1 *E. coli* growth

E. coli was grown overnight in liquid LB media or on solid LB agar at 37°C. For liquid cultures, flasks/tubes were shaken at 200rpm. If required 100µg/mL ampicillin was added to the media before inoculation for *E. coli* mutants.

2.2.2.2 *E. coli* storage

E. coli bacteria was stored as glycerol stocks. These were prepared from overnight *E. coli* LB cultures, which were diluted with sterile 50% w/v glycerol to a final concentration of 25%. These glycerol stocks were then stored at -80°C.

2.2.2.3 *E.coli* electrocompetency

Electrocompetent *E. coli* cells were prepared by washing mid exponential phase bacteria (OD_{600} ~ 0.5 - 0.7) with ice cold 10 w/v glycerol twice. This wash was performed by pelleting the bacteria, discarding the supernatant and resuspending in glycerol. After the second wash, the pellet is resuspended in the residual glycerol and then aliquoted (100µL) into cryogenic tubes for storage at -80°C.

2.2.2.4 *E. coli* transformation

Electroporation was executed to introduce plasmids into *E. coli* bacteria. An aliquot of electrocompetent *E. coli* was defrosted and added to an ice cold electroporation cuvette (0.2 cm) along with 100-200 ng of plasmid. Electroporation was performed at 200 Ω, 25µF and 2.5 kV using BioRad Gene Pulser II. Immediately after electroporation 1mL of LB medium was added to the cuvette for cell recovery. This mixture was transferred to a new tube and

incubated at 37°C for 1 hour at 200rpm. The mixture was then plated onto LB agar with 100µg/mL ampicillin and incubated overnight at 37°C. Positive transformants were then used to make glycerol stocks.

2.2.2.5 Plasmid extraction

The Thermo Scientific GeneJet Plasmid Miniprep kit was used for the plasmid purification from *E. coli* bacteria following the manufacturer's instructions.

2.2.3 *Rhodococcus* bacteria

2.2.3.1 *Rhodococcus* growth

Rhdococcus bacteria was grown overnight in liquid LB media or on solid LB agar at 30°C. For liquid cultures, flasks/tubes were shaken at 200rpm. If required 150µg/mL hygromycin was added to the media before inoculation of *Rhodococcus* mutants.

2.2.3.2 *Rhodococcus* storage

Rhodococcus bacteria was stored as glycerol stocks. These were prepared from overnight LB cultures, which were diluted with sterile 50% w/v glycerol to a final concentration of 25%. These glycerol stocks were then stored at -80°C.

2.2.3.3 *Rhodococcus* electrocompetency

Electrocompetent *Rhodococcus* cells were prepared by washing mid exponential phase bacteria (OD_{600} ~ 0.5 - 0.7) with ice cold 10% w/v glycerol four times reducing the amount of glycerol each wash. The wash was performed by pelleting the bacteria, discarding the supernatant and

resuspending in glycerol. After the final wash, the pellet is resuspended in the residual glycerol and then aliquoted (100µL) into cryogenic tubes for storage at -80°C.

2.2.3.4 *Rhodococcus* transformation

Electroporation was executed to introduce plasmids into *Rhodococcus* bacteria. An aliquot of electrocompetent *Rhodococcus* was defrosted and added to an ice cold electroporation cuvette (0.2 cm) along with 100-200 ng of plasmid. Electroporation was performed at 200 Ω, 25µF and 2.5 kV using BioRad Gene Pulser II. Immediately after electroporation 1mL of LB medium was added to the cuvette for cell recovery. This mixture was transferred to a new tube and incubated at 30°C for 4 hours at 200rpm. The mixture was plated on LB agar with 150µg/mL hygromycin and incubated at 30°C overnight. Positive transformants were then used to make glycerol stocks.

2.2.3.5 *Rhodococcus* mutant construction

Mutant *Rhodococcus* bacteria were constructed using the RHA1 and PR4 parent strains. CO-213 was constructed by electroporation of RHA1 with pOSV556, while CO-231 was constructed by electroporation of PR4 with pOSV556 as described in 2.2.3.4. CO-218 was constructed by electroporation of pC0256 into RHA1, while CO-236 was constructed by electroporation of pC0256 into PR4. The pC0256 plasmid contains the RHA1_RS09515 gene which was synthetically built according to the NCBI Genbank entry by GeneWiz before being ligated into pOSV556 via PstI and StuI restriction enzymes. CO-222 was constructed by electroporation of pC0260 into RHA1, while CO-236 was constructed by electroporation of pC0260 into PR4. The pC0260 plasmid contains the PD630_RS21650 gene which was synthetically built according to the NCBI Genbank entry by GeneWiz before being ligated into pOSV556 via PstI and XhoI restriction enzymes.

2.2.3.6 DNA extraction

Genomic DNA extraction was completed using the MP Biomedicals FastDNA Spin Kit for Soil DNA Extraction. The protocol was carried out using the manufacturer's instructions with the change that prewarmed dH_2O was used instead of DES solution.

2.2.4 *Streptomyces coelicolor* bacteria

2.2.4.1 *S. coelicolor* growth

S. coelicolor bacteria was grown for 4 days in liquid or on solid agar, incubated at 30°C. For liquid cultures, baffled flasks were used and shaken at 200rpm. If required 150µg/mL hygromycin was added to the media before inoculation for *S. coelicolor* mutants.

2.2.4.2 *S. coelicolor* storage

S. coelicolor hyphae glycerol stocks were prepared from a 4 day liquid culture growth in YEME at 30°C, 200 rpm. This was diluted with sterile 50% w/v glycerol to a final concentration of 25%. These stocks were then stored at -80°C.

S. coelicolor spore glycerol stocks were prepared from a lawn of spores grown on SFM plates. 3 mL of sterile dH_2O was added to the plate and a sterile loop was used to suspend the spores. This was transferred to a universal and vortexed for mycelia separation. This was then filtered through a syringe filled with non-absorbent wool and centrifuged to pellet the spores. The supernatant was discarded and the pellet was resuspended in the residual liquid. This was diluted with 50% w/v glycerol to a final concentration of 25%. These stocks were then stored at -80°C.

2.2.4.3 Conjugation

Bacterial conjugation was completed to transfer pOSV556 from *E. coli* ET12567/pUZ8002 into *Streptomyces coelicolor* M145. The *E. coli* cells were grown to a OD_{600} ~ 0.4 – 0.6 and centrifuged (4000 rpm, 10 minutes) to pellet the cells. This pellet was washed twice with clean LB and then resuspended in 500 µL of LB. The *Streptomyces* spores were germinated by thawing stocks on ice and adding 5 µL of spores to 500 µL 2xYT for incubation at 50°C for 10 minutes.

After spore germination, the *E. coli* cells and *Streptomyces* spores were combined, mixed and centrifuged for 2 minutes. The supernatant was discarded and the pellet was resuspended in 100 µL sterile dH_2O. This was serially diluted four times by a factor of 10 each dilution. 100 µL of each serial dilution was plated onto SFM + 10 mM $MgCl_2$ and incubated overnight at 30°C. After the overnight incubation, each plate was overlayed with 1 mL of 25 µg/ml nalidixic acid, 150 µg/ml hygromycin dissolved in dH_2O. The plates were incubated again at 30°C until growth was observed. Positive colonies were then sub-cultured on fresh SFM and glycerol stocks were made.

2.2.4.4 *Streptomyces* mutant construction

The mutant *Streptomyces coelicolor* CO-250 was constructed from the parent M145 strain. The empty pOSV556 vector was electroporated into *E. coli* ET12567/pUZ8002 and then transferred into *S. coelicolor* M145 by conjugation (2.2.4.3)

2.2.4.5 DNA extraction

Genomic DNA extraction was completed using the MP Biomedicals FastDNA Spin Kit for Soil DNA Extraction. The protocol was carried out using the manufacturer's instructions with the change that prewarmed dH_2O was used instead of DES solution.

2.2.5 Molecular Biology

2.2.5.1 Polymerase chain reaction

In all polymerase chain reactions (PCR), New England Biolabs (NEB) Phusion High-Fidelity DNA polymerase was used in the following reaction mixture:

5x Phusion GC buffer	10 µL
dNTP (10 mM)	1 µL
Primers (100 mM)	1 µL each
DNA template (100 – 200 ng)	1 µL
Phusion Taq polymerase	1 unit
Sterile dH_2O	make up to 50 mL

All PCR reactions were completed using the following cycle:

Step 1: 98°C for 2 minutes
Step 2: 98°C for 30 seconds
Step 3: 55°C for 30 seconds
Step 4: 72°C for 1 minutes/1kb
Step 5: 72°C for 10 minutes
Step 6: 4°C hold

The steps 2 to 4 were repeated 30 times, with step 4 timing variable depending on the length of the required product.

2.2.5.2 Agarose gel electrophoresis

All agarose gel electrophoresis was completed using a 1% agarose dissolved in 1x TAE buffer. Each gel was stained using ThermoFisher Scientific Sybr Safe DNA Gel Stain according the

manufacturer's protocol. Each sample was combined with NEB 6X Gel Loading Dye, Purple to a final concentration of 1X Gel Loading Dye before loading into the gel. Electrophoresis was carried out at 120V for 1 hour.

2.2.5.3 Restriction enzyme digest

Restriction digests of plasmids and PCR products were performed using NEB restriction enzymes. The following mixture was prepared for each reaction:

CutSmart Buffer	5 µL
DNA	20 µL
Enzyme.	1 µL of each enzyme
Sterile dH_2O	make up to 20 µL

In the case of a restriction digest of pOSV556 to later ligate a gene product into the plasmid, 3 µL Shrimp Alkaline Phosphatase (rSAP) was added to the reaction mixture.

Once the mixture was prepared, it was incubated at 37°C for 1 hour before enzyme inactivation per manufacturer's instructions. After inactivation the digest products were analysed by agarose gel electrophoresis and purified using the GeneJet Gel Extraction kit following the manufacturer's protocol.

2.2.5.4 DNA ligation

Plasmids and PCR products were ligated together using Invitrogen DNA ligase. A ratio of 3:1 PCR digest:plasmid digest was used for all reactions. The following mixture was prepared for each reaction:

5X T4 DNA ligase buffer	4 µL
DNA (3:1 ratio)	8 µL
T4 DNA ligase	1 unit
Sterile dH_2O	make up to 20 µL

The mixture once prepared was incubated at room temperature for 15 minutes before enzyme inactivation per manufacturer's instructions. This was then used for transformation of *E. coli* bacteria.

2.2.5.5 Plasmid construction

All plasmids constructed used the parent plasmid pOSV556 seen in Figure 2.1. Plasmids pC-0252 and pC-0253 were constructed using PCR replication of the gene to be used in a restriction enzyme digest and subsequent ligation reaction (using the method described in 2.2.5.4). The restriction enzymes used are specified in Table 2.2.

Plasmids pC-0254 – pC-0267 were constructed by Genewiz (https://genewiz.com). Each gene was synthetically constructed using the available nucleotide sequence on NCBI Genbank. These were then cloned into the pOSV556 vector (using the method described in 2.2.5.4) using the restriction enzymes specified in Table 2.2.

Figure 2.1: Plasmid map of pOSV556 used as the parent plasmid for all subsequent plasmids used for mutant construction. The plasmid contains both hygromycin (dark blue) and ampicillin (yellow) resistance genes for use in actinomycetes and *E. coli* respectively. The ermE* is the constitutive promoter and is the location where all genes have been cloned in. The plasmid contains a pSAM2 integrase (light blue) and attP site (pink) for integration into the bacterial genome to prevent loss of the plasmid. The oriT feature (grey) is required for conjugation, while ColE1 origin (green) is needed for plasmid replication in *E. coli*.

2.2.6 Sequencing

2.2.6.1 RNA Sequencing

The first RNA sequencing was undertaken by Genewiz. Cell cultures of *S. coelicolor* M145, CO-121, CO-124 and CO-154 were grown in triplicate for 4 days at 30°C in 50 mL liquid SMM, shaking at 200 rpm. These cells were harvested by centrifugation, flash frozen and shipped to

Genewiz for RNA extraction, depletion, library construction, sequencing and bioinformatic analysis.

The second RNA sequencing was undertaken by Eurofins Genomics (https://eurofinsgenomics.eu). Cell cultures of *S. coelicolor* M145, CO-121, CO-124, CO-154 and CO-250 were grown in triplicate for 4 days at 30°C in 50 mL liquid YEME, shaking at 200 rpm. These cells were harvested by centrifugation, flash frozen and shipped to Eurofins for RNA extraction, depletion, library construction, sequencing and bioinformatic analysis.

Bioinformatic analysis for both sets of RNA sequencing was completed using the Bowtie2[71] pipeline for read alignment to the reference genome of *S. coelicolor* A3(2) (NCBI Genbank: AL645882.2). Differential expression was identified using the DESeq2[72] software. All raw data, workflow and results of the bioinformatic analysis were provided via file transfer and stored on the Warwick sharepoint.

2.2.7 Analytical Chemistry

2.2.7.1 Organic Extraction

Liquid-liquid ethyl acetate organic extraction was completed on the supernatant of the cultures sent for RNA sequencing. The first set of RNA sequencing cell cultures of *S. coelicolor* M145, CO-121, CO-124 and CO-154 were grown in triplicate for 4 days at 30°C, 200 rpm in 50 mL liquid SMM. The cells were harvested by centrifugation and separated from the supernatant. 25 mL of the supernatant was combined with an equal volume of ethyl acetate. This was allowed to settle and the upper ethyl acetate layer was separated and evaporated. The residual of the evaporation was dissolved in a 1:1 mixture of HPLC grade methanol:water and stored at 4°C.

The second set of RNA sequencing cell cultures of *S. coelicolor* M145, CO-121, CO-124, CO-154 and CO-250 were grown in triplicate for 4 days at 30°C, 200 rpm in 50 mL liquid YEME. The cells

were harvested by centrifugation and separated from the supernatant. 25 mL of the supernatant was combined with an equal volume of ethyl acetate. This was allowed to settle and the upper ethyl acetate layer was separated and evaporated. The residual of the evaporation was dissolved in a 1:1 mixture of HPLC grade methanol:water and stored at 4°C.

2.2.7.2 Liquid chromatography – mass spectrometry

The liquid chromatography – mass spectrometry (LCMS) analysis was performed on the Bruker Compact operated in positive electrospray ionisation. Organic extracts were filtered on a spin column (4500 rpm for 5 minutes) and 20 µL of the filtered extract was injected. The following solvent and linear gradient profile with a flow rate of 1mL/minute was used for all LCMS (Table 2.5).

Time (minutes)	Water + 0.1% formic acid (%)	Methanol + 0.1% formic acid (%)
0	95	5
5	95	5
30	0	100
35	0	100

Table 2.5: Gradient profile for LCMS analysis

3. Results

3.1 Investigation of *Streptomyces coelicolor* MmyB homologs

The MmyB homologs that reside in *S. coelicolor* M145 seem to have the potential to potentially be involved in at least one of the cryptic biosynthetic clusters that have yet to be characterised. The success of previous studies that centre around homologs of the repressors of the methylenomycin cluster causing the expression of novel natural products[24] coupled with the evidence that overexpression of the *mmyB* activator overcomes repression[22], indicate there is potential for the combination of these approaches to have a successful outcome.

3.1.1 Minimal Media RNA sequencing

The three mutants CO-121 (overexpressed *sco7140*), CO-124 (overexpressed *sco0236*) and CO-154 (overexpressed *sco7140*) along with the control parent strain (M145) were grown in supplemental minimal media (SMM) in triplicate. Cultures were grown for four days at 30°C, 200 rpm before being pelleted and frozen. These pellets were then sent to Genewiz for RNA extraction, depletion, sequencing and bioinformatic analysis.

Analysis discovered multiple genes being over and under expressed in each of the three mutants compared to the M145 parent strain (Table 3.1). The CO-124 had almost three times as many unique genes overexpressed when compared to the other two mutants, while also having the most under expressed.

Strain	Number of genes overexpressed	Number of genes under expressed	Number of unique genes overexpressed	Number of unique genes under expressed
CO-121	243	258	61	143
CO-124	198	316	158	203
CO-154	250	205	56	62

Table 3.1: The total number of genes flagged as differentially expressed in CO-121, CO-124 and CO-154 compared to *S. coelicolor* M145 along with how many genes were unique to each mutant.

3.1.1.1 MmyB homologs

Streptomyces coelicolor contains sixteen homologs of the methylenomycin biosynthetic gene cluster transcriptional activator MmyB. The three mutants CO-121, CO-124 and CO-154 each had an additional copy of a different MmyB homolog under the constitutive promoter ErmE* (CO-121 = *sco7140*, CO-124 = *sco0236*, CO-154 = *sco6926*). As can be seen in Figure 3.1, each of the mutants has overexpressed the MmyB homolog that was added.

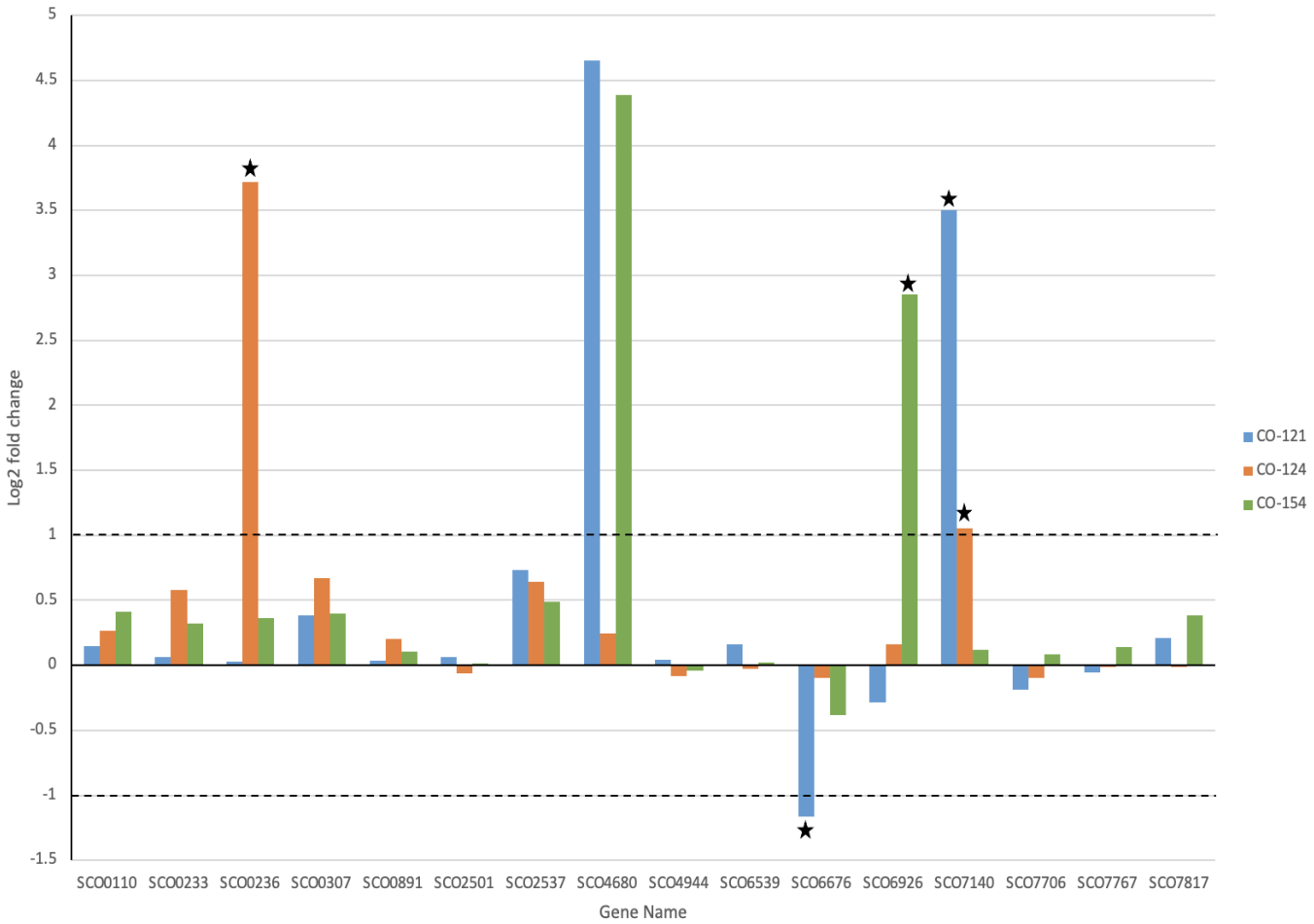


Figure 3.1: The log2 fold change of the sixteen MmyB homologs in the chromosome genome of *Streptomyces coelicolor* A3(2) in CO-121, CO-124 and CO-154 compared to *S. coelicolor* M145 grown in SMM. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05.

Other MmyB homologs were detected with differential expression in both CO-121 and CO-124, with *sco6676* being under expressed in CO-121 while *sco7140* was over expressed in CO-124. The *sco4680* was detected to have overexpression but the statistical analysis found them to have a p-value greater than 0.05 which means it cannot be said for certain that this is correct.

MmyB homolog	**Directly Downstream Gene**	**Function**	**Log2 fold change**		
			CO-121	**CO-124**	**CO-154**
sco0110	*sco0111*	SDR oxidoreductase	0.23124464	0.85206519	0.24198328
sco0233	*sco0234*	SDR oxidoreductase	-0.7813885	0.95051056	-0.2063009
sco0236	*sco0237*	SDR oxidoreductase	-0.1282383	0.65977025	0.0840808
sco0307	*sco0306*	Pseudogene	0.38714511	N/A	N/A
sco0891	*sco0892*	SDR oxidoreductase	-0.1362946	0.53747665	-0.0997532
sco2501	*sco2502*	Transport protein	-1.1857145	0.41952831	-0.4752849
sco2537	*sco2538*	Unknown function	-0.1151079	0.04103599	-0.2077434
sco4680	*sco4681*	SDR	4.43969311	-0.24353432	4.34830378
sco4944	*sco4945*	SDR alcohol dehydrogenase	0.07641419	0.06158177	-0.0416736
sco6539	*sco6540*	Unknown function	N/A	N/A	N/A
sco6676	*sco6677*	ABC transporter	-1.2783448	-0.0117764	-0.4541524
sco6926	*sco6925*	Membrane protein	-0.6665062	0.76590891	-0.32222953
sco7140	*sco7141*	Nitroreductase	-0.1644282	0.80264951	-0.4248155
sco7706	*sco7707*	Unknown function	-0.3002616	0.18933137	0.30371693
sco7767	*sco7768*	Unknown function	-0.4930199	1.0584636	0.48588659
sco7817	*sco7818*	SDR oxidoreductase	-0.2944854	0.26230985	-0.2347388

Table 3.2: The log2 fold change of the directly downstream genes to MmyB homologs in the three mutants compared to the wildtype, grown in SM media. Values in red indicate a p-value < 0.05 for log2 fold changes greater than -1, while values in green indicate a p-value <0.05 for log2 fold changes greater than 1.

The most likely genes to be regulated by the various MmyB homologs are those that are directly downstream to them in the genome seen in Table 3.2. Almost half of the homologs are found nearby to a reductase enzyme, the majority of which are part of the short chain dehydrogenase/reductase family (SDR).

Only four homologs have significant altered expression to the downstream gene; *sco2501, sco4680, sco6676* and *sco7767.* Further investigation into these four finds that for all but *sco4680,* in all three mutants there are no subsequent genes with altered expression. However, in *sco4680* there is a cluster of increased expression in both CO-121 and CO-154 from *sco4671* - *sco4698* visualised in Figure 3.2.

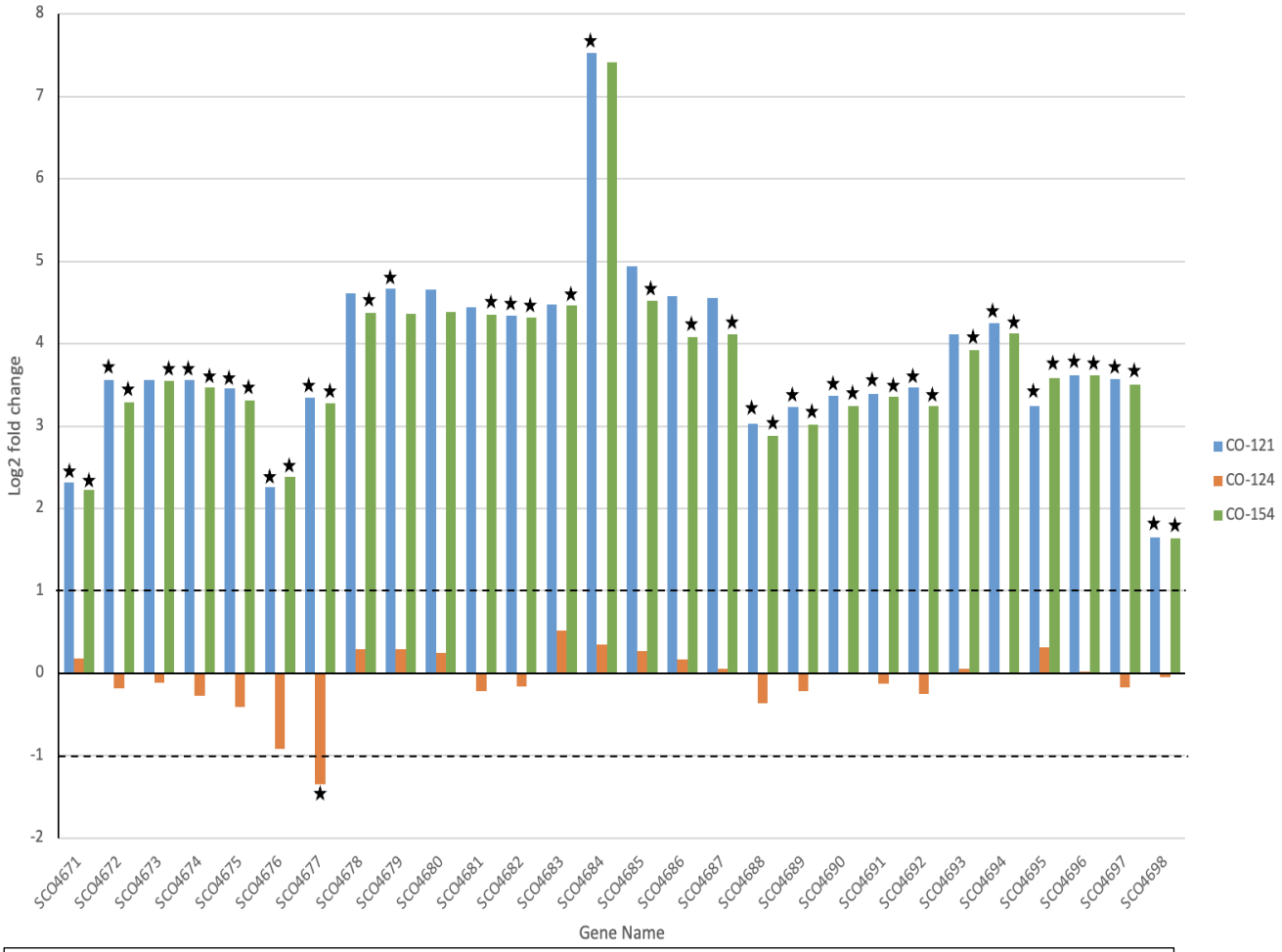


Figure 3.2: The log2 fold change of RNA transcription in CO-121, CO-124 and CO-154 of a potential cluster centred around the MmyB homolog *sco4680,* compared to *S. coelicolor* M145 grown in SM media. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05.

Gene Name	Function
sco4671	lysR family regulatory protein
sco4672	Secreted protein
sco4673	deoR family transcriptional regulator
sco4674	Unknown function
sco4675	Unknown function
sco4676	Unknown function
sco4677	Regulatory protein
sco4678	Unknown function
sco4679	Unknown function
sco4680	MmyB homolog
sco4681	Short chain dehydrogenase
sco4682	Tautomerase
sco4683	NADP specific glutamate dehydrogenase
sco4684	Cold shock protein
sco4685	DEAD-box helicase
sco4686	Unknown function
sco4687	Unknown function
sco4688	Unknown function
sco4689	Unknown function
sco4690	Membrane protein
sco4691	Membrane protein
sco4692	Unknown function
sco4693	Membrane protein
sco4694	Unknown function
sco4695	Unknown function
sco4696	Unknown function
sco4697	Integral membrane protein
sco4698	Insertion element IS1652 transposase

Table 3.3: List of genes and their functions in the cluster centred around the *sco4680* MmyB homolog.

This cluster of genes with drastic overexpression has been termed the supercoiling hypersensitive cluster (SHC)[39]. It has a number of uncharacterised genes as seen in Table 3.3 but many involved in DNA regulation as well as chromosomal manipulation such as the cold shock protein *sco4684* and the DEAD-box helicase *sco4685*.

The core genome of *S. coelicolor* A3(2) has twenty-seven biosynthetic gene clusters predicted by antiSMASH (Figure 1.2). Many of these are well studied and characterised but not all of them have had their products discovered.

Region	First Gene	Last Gene	Total cluster size (kb)
1	*sco0104*	*sco0147*	53.0
2	*sco0178*	*sco0201*	24.8
3	*sco0257*	*sco0278*	23.5
4	*sco0473*	*sco0508*	49.8
5	*sco0750*	*sco0755*	8.2
6	*sco1186*	*sco1225*	38.8
7	*sco1862*	*sco1870*	10.4
8	*sco2694*	*sco2706*	10.6
9	*sco2781*	*sco2789*	11.0
10	*sco3215*	*sco3250*	79.1
11	*sco5057*	*sco5119*	70.9
12	*sco5212*	*sco5231*	20.6
13	*sco5282*	*sco5354*	72.5
14	*sco5797*	*sco5803*	10.3
15	*sco5874*	*sco5910*	45.3
16	*sco6041*	*sco6050*	11.3
17	*sco6065*	*sco6080*	19.3
18	*sco6221*	*sco6231*	13.2
19	*sco6259*	*sco6291*	70.2

20	*sco6419*	*sco6450*	54.2
21	*sco6669*	*sco6686*	22.8
22	*sco6750*	*sco6773*	25.8
23	*sco6808*	*sco6844*	48.1
24	*sco6919*	*sco6943*	26.5
25	*sco7176*	*sco7239*	73.9
26	*sco7455*	*sco7476*	21.1
27	*sco7648*	*sco7711*	73.3

Table 3.4: List of the 27 antiSMASH predicted biosynthetic gene clusters in *Streptomyces coelicolor* A3(2) with the boundaries of the predictions and how large the cluster is.

3.1.1.2 Region 2

The second cluster in the antiSMASH prediction is one that the product is known, isorenieratene[40]. However this cluster is of potential interest due to its proximity to *sco0236*, the gene overexpressed in CO-124. However there is not any altered expression in CO121, CO-124 and only three in CO-154, none of which are predicted biosynthetic genes (Figure 3.3).

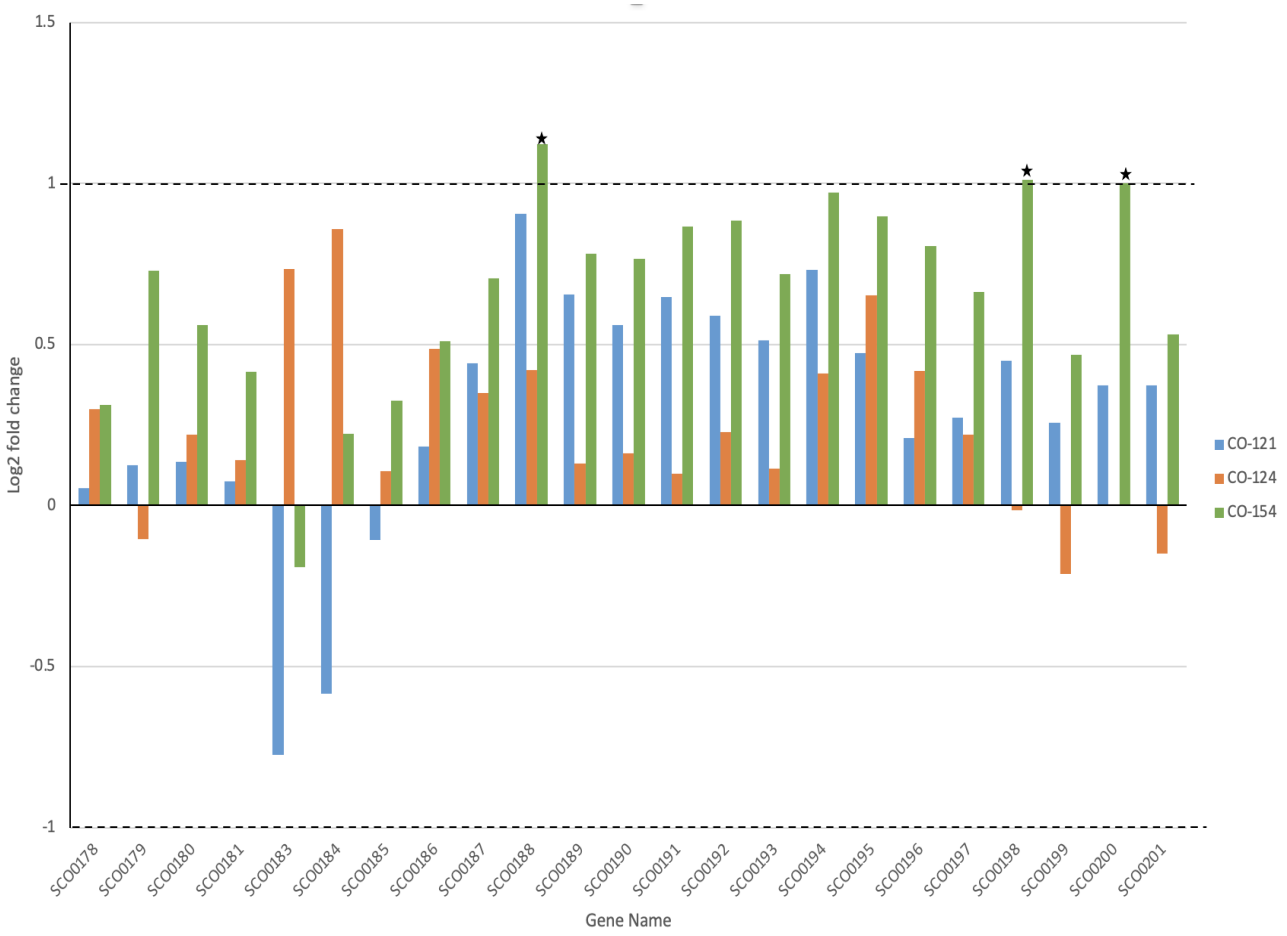


Figure 3.3: The log2 fold change of RNA transcription for the second antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to *S. coelicolor* M145 grown in SMM. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05.

3.1.1.3 Region 3

The third cluster is of interest again similar to cluster 2, due to the proximity of the sco0236 MmyB homolog and the fact that there is no similarity found with by the antiSMASH algorithm to any known BGCs.

All three of the mutants have some changes in expression within the cluster (Figure 3.4). Both CO-121 and CO-154 have overexpression of the same gene, *sco0269*, which is one of the two core biosynthesis genes of the cluster, encoding a post-translational lanthipeptide modifier.

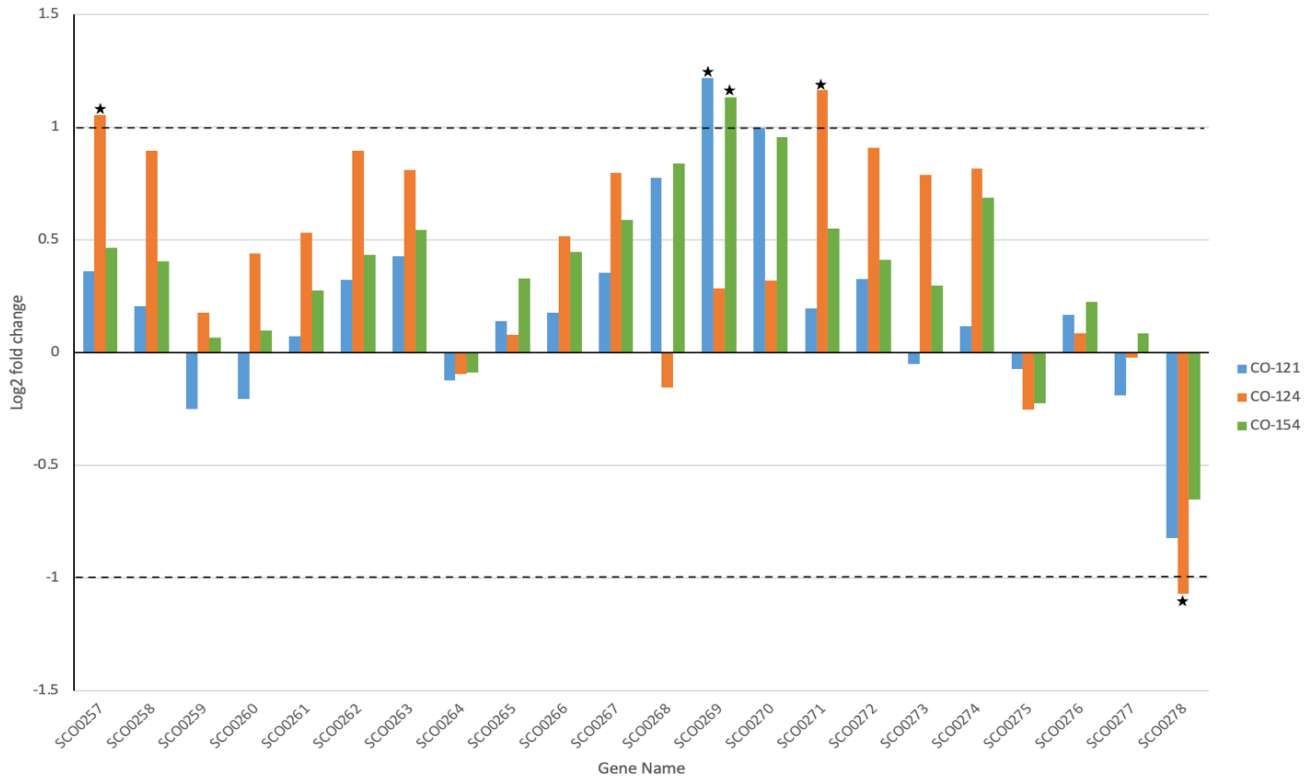


Figure 3.4: The log2 fold change of RNA transcription for the third antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to *S. coelicolor* M145 grown in SMM. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05.

3.1.1.4 Region 11

This cluster is one of the best studied in *S. coelicolor*, the actinorhodin BGC[46]. In all three of the mutants there is under expression extending from *sco5069 – sco5095,* which encompasses most of the biosynthetic genes for actinorhodin (Figure 3.5). The CO-124 mutant is the only one with any genes overexpressed which is *sco5109*.

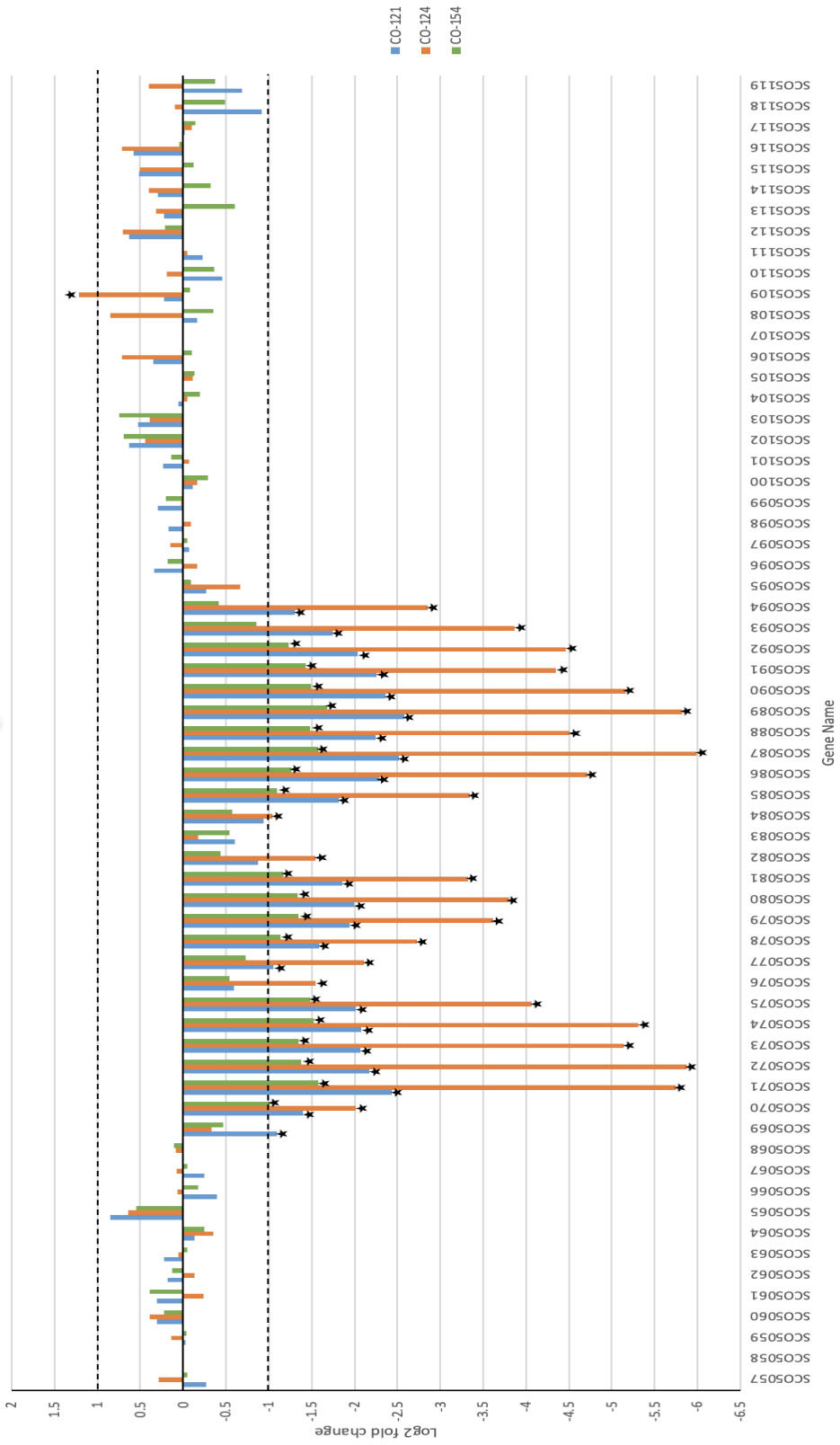


Figure 3.5: The log2 fold change of RNA transcription for the eleventh antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to *S. coelicolor* M145 grown in SMM. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

3.1.1.5 Region 13

The thirteenth cluster in the antiSMASH prediction has a two thirds similarity to a known spore pigment[43,49] from *Streptomyces avermitilis* but also contains all the genes that produce curamycin[50], an antimicrobial discovered in *Streptomyces curacoi*. In this cluster, CO-154 has no differential expressed genes but CO-121 has four genes under expressed, one being a core biosynthetic gene, *sco5318*, while CO-124 has two genes over expressed (Figure 3.6).

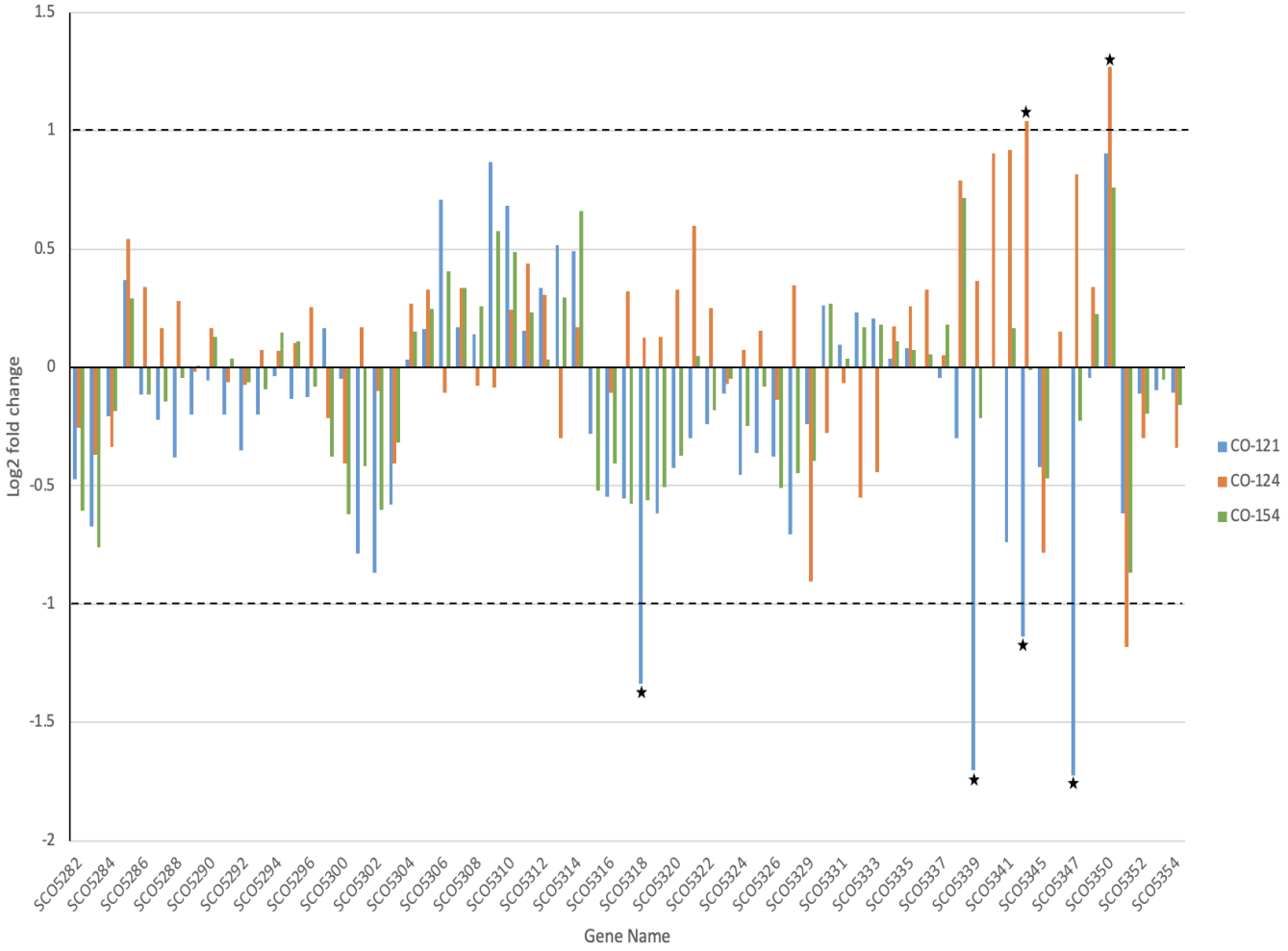


Figure 3.6: The log2 fold change of RNA transcription for the thirteenth antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to *S. coelicolor* M145 grown in SMM. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05.

3.1.1.6 Region 15

The fifteenth cluster is one that has been previously studied and is known to produce undecylprodigiosin, a compound that has been investigated for various bioactivities[51]. The only mutant that has an effect on the transcription of the genes in this cluster is CO-121, which has four genes that are under expressed, with all four involved in the biosynthesis (Figure 3.7).

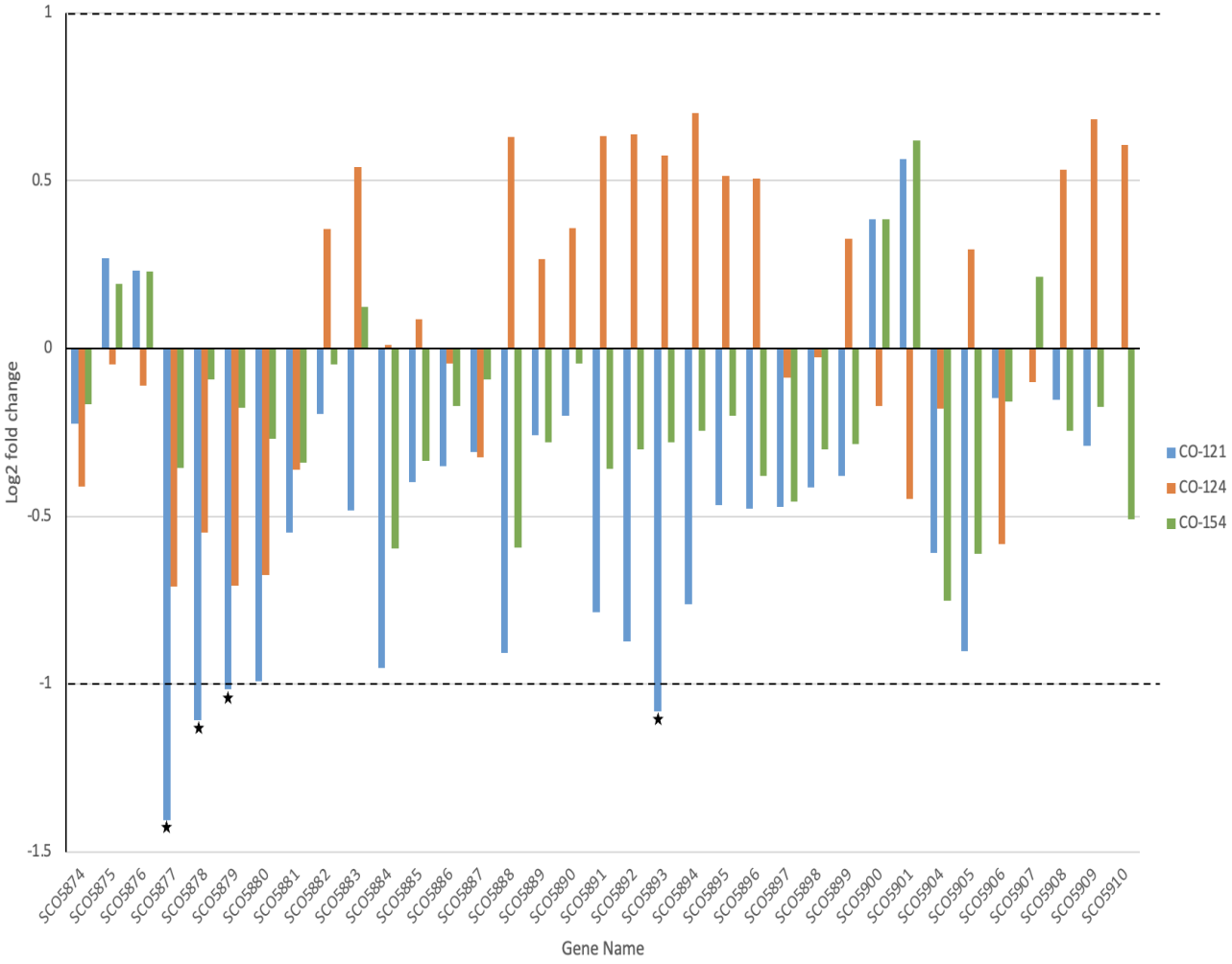


Figure 3.7: The log2 fold change of RNA transcription for the fifteenth antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to *S. coelicolor* M145 grown in SMM. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

3.1.1.7 Region 21

This cluster is another one of interest due to the presence of the MmyB homology, *sco6676* within the prediction. The lanthipeptide product of the cluster has been discovered as SapB with the genes responsible identified[55]. All three mutants have differential expression found within the cluster (Figure 3.8). CO-121 has *sco6676* under expressed along with the adjacent gene, while CO-124 has *sco6682* under expressed. CO-154 has the opposite effect to CO-124 with *sco6682* being over expressed as well as the adjacent gene.

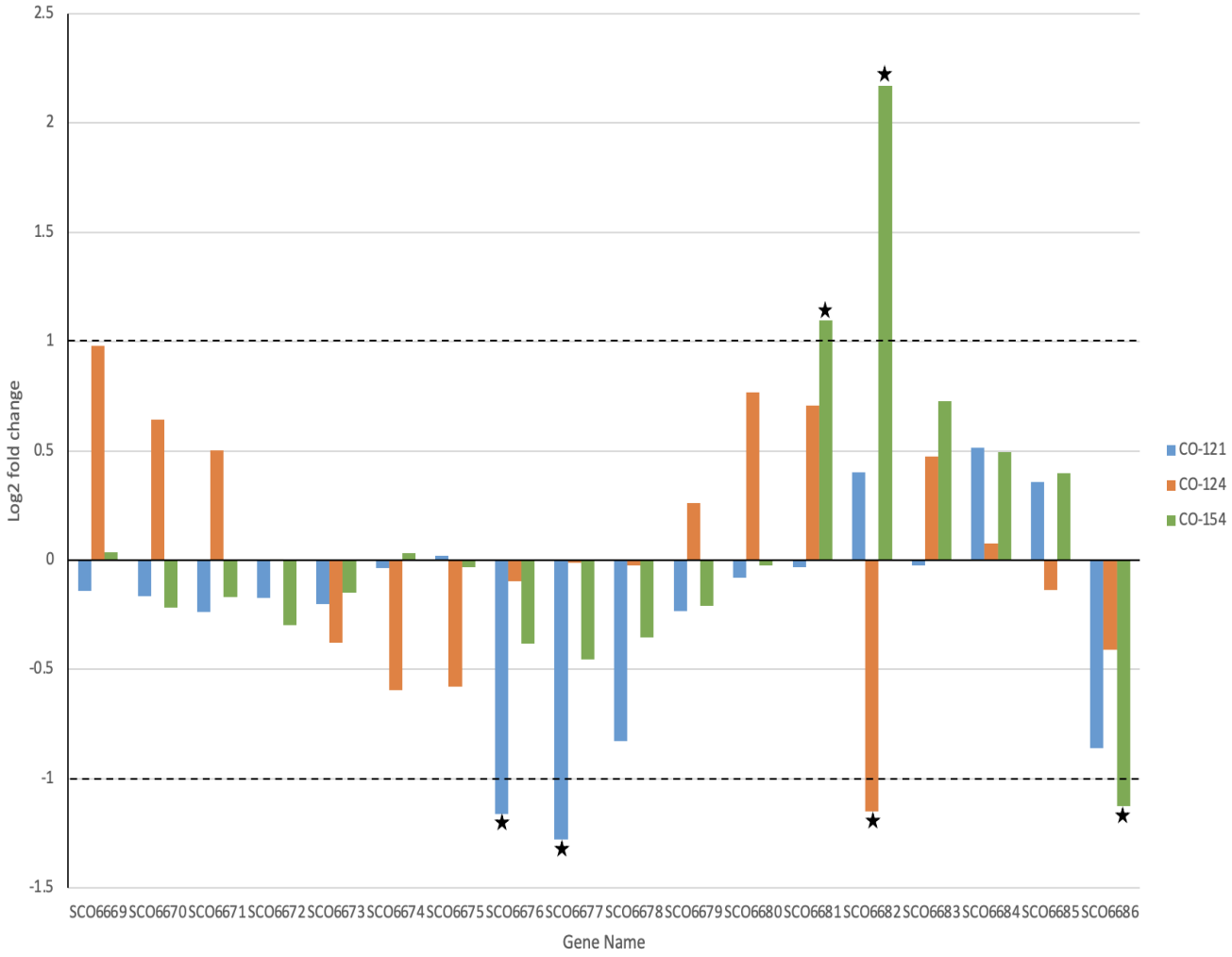


Figure 3.8: The log2 fold change of RNA transcription for the twenty-first antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to *S. coelicolor* M145 grown in SMM. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05.

3.1.1.8 Region 24

This lanthipeptide cluster is another of interest due to the MmyB homolog *sco6926* being found within its limits, especially as CO-154 has the extra copy of *sco6926* under a constitutive promoter. In CO-121 there are fifteen genes all under expressed while CO-124 has one gene over expressed. The CO-154 has the *sco6926* overexpressed and another four that are under expressed.

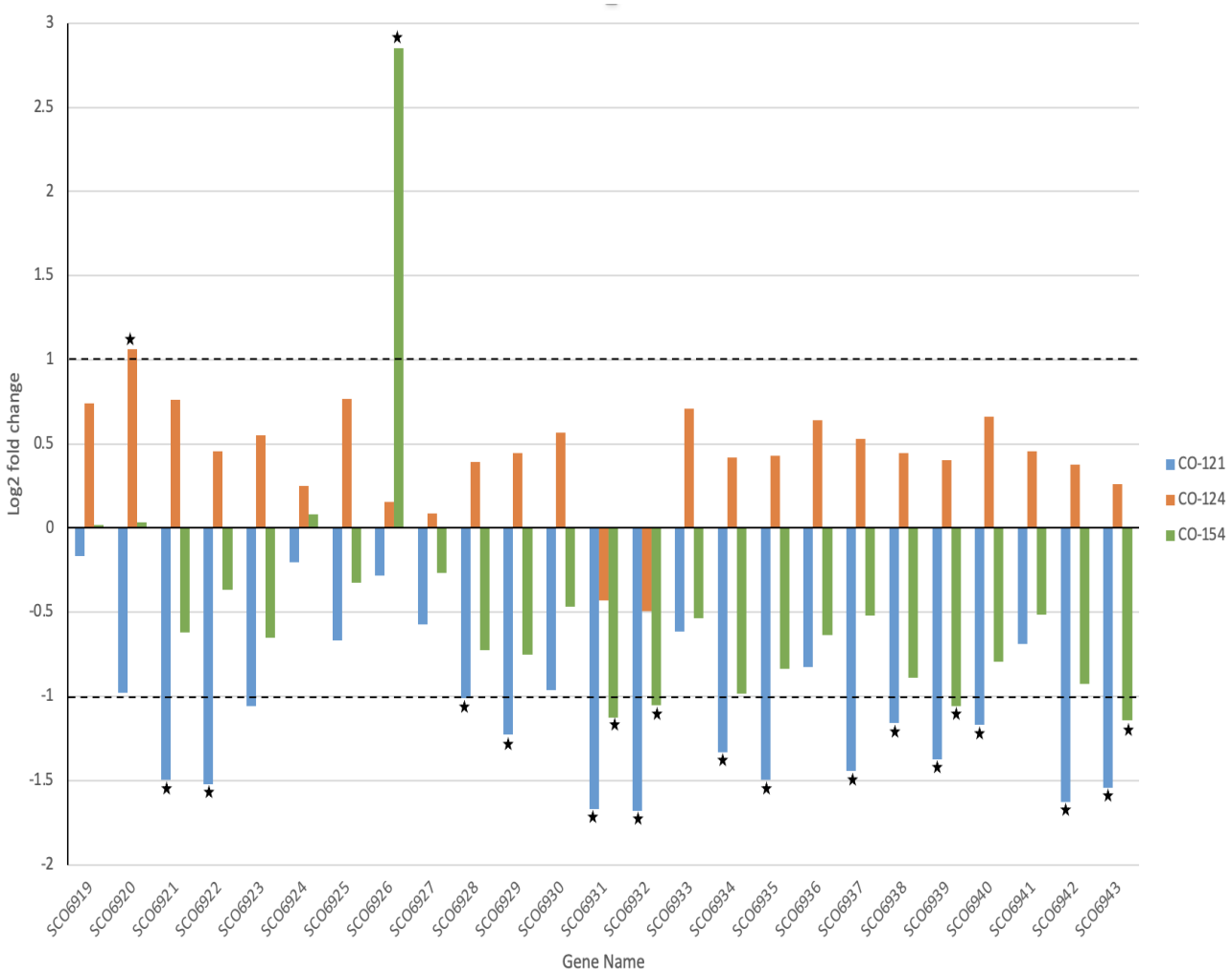


Figure 3.9: The log2 fold change of RNA transcription for the twenty-fourth antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to *S. coelicolor* M145 grown in SMM. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

3.1.1.9 Region 26

The twenty-sixth cluster is the only indole class predicted. All three mutants have only one gene with differential expression (Figure 3.10), with CO-124 under expressing *sco7460*. Both CO-121 and CO-154 have the *sco7467* gene overexpressed which is the predicted core biosynthetic gene.

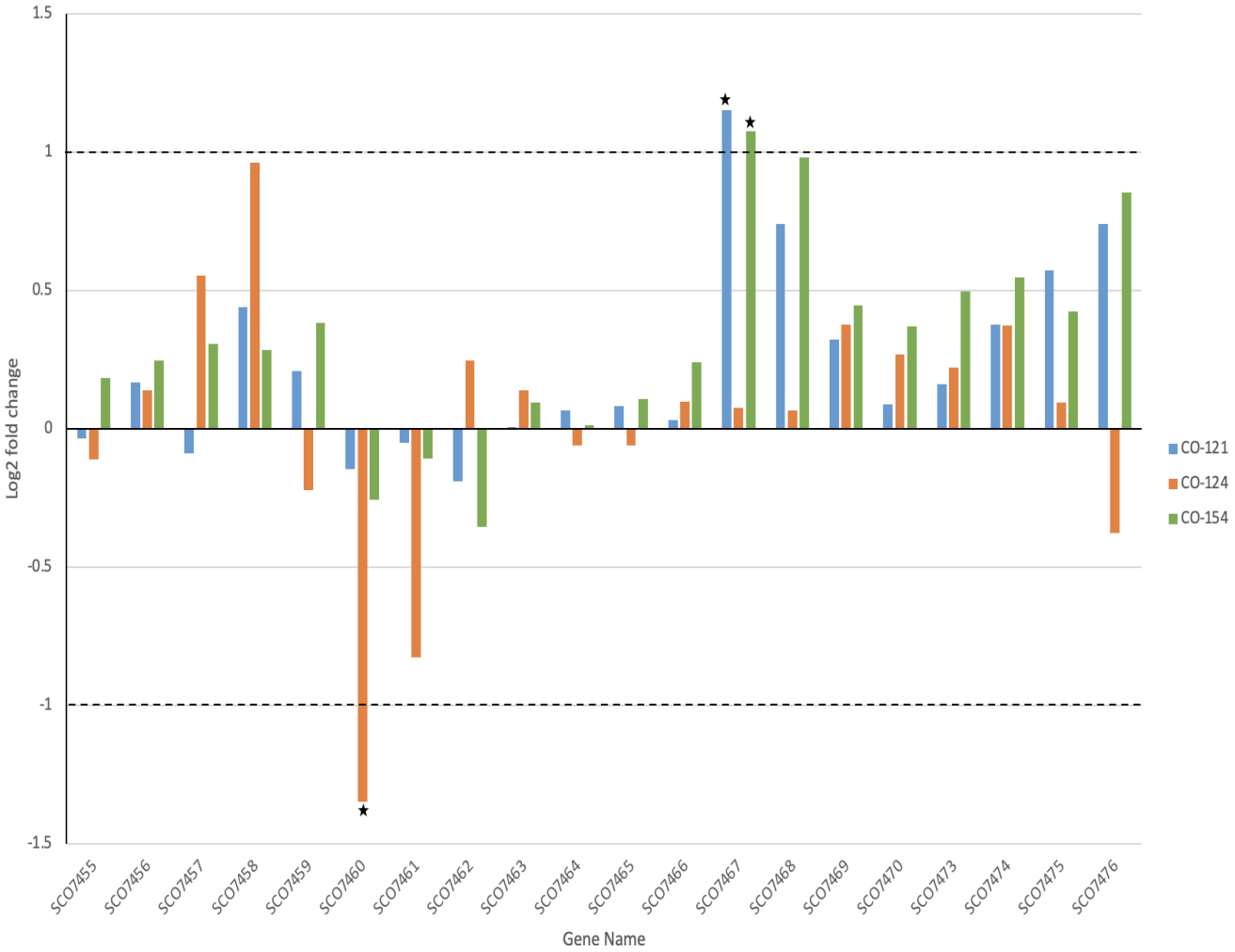


Figure 3.10: The log2 fold change of RNA transcription for the twenty-sixth antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to *S. coelicolor* M145 grown in SMM. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

3.1.2 YEME Media RNA sequencing

The three mutants (CO-121, CO-124 and CO-154) along with the parent strain (M145) were grown in yeast extract malt extract (YEME) media. The empty vector control (CO-250) was grown alongside the mutants and wildtype to look at the impact the plasmid integration had on transcription. Cultures were grown for four days at 30°C, 200 rpm before being pelleted and frozen. These pellets were then sent to Eurofins Genomics for RNA extraction, depletion, sequencing and bioinformatic analysis

The CO-250 empty vector mutant was compared to the parent strain M145 to investigate how the vector integration can influence transcription. In total 202 genes were identified to have differential expression in CO-250. All of these were detected as overexpressed. When focusing on predicted BGCs and MmyB homologs, only sixteen of these genes were over expressed in CO-250 (Table 3.5).

Gene Name	Gene Function	Predicted region
sco0121	ABC transport system ATP binding protein	1
sco0179	Zinc containing dehydrogenase	2
sco1190	Export protein	6
sco1199	Oxidoreductase	6
sco3239	Unknown function	10
sco4198	MmyB homolog	-
sco5073	Oxidoreductase	11
sco5074	Dehydratase	11
sco5101	Unknown function	11
sco5103	Regulatory protein	11
sco5896	RedH (phosphoenolpyruvate utilising enzyme)	15
sco6068	Unknown function	17
sco6073	Cyclase	17

sco6231	Sugar transport system sugar binding lipoprotein	18
sco6752	Integral membrane transferase	22
sco7656	Unknown function	27

Table 3.5: The genes overexpressed in CO-250 compared to *S. coelicolor* M145 grown in YEME media, that are either MmyB homologs or predicted by antiSMASH within a cluster

When the three mutants CO-121, CO-124 and CO-154 expression was compared against the CO-250 control roughly three thousand genes were identified in each mutant to have over expression while CO-124 and CO-154 had almost triple the number of genes under expressed that CO-121 had (Table 3.6).

Strain	**Number of genes overexpressed**	**Number of genes under expressed**
CO-121	3419	531
CO-124	2982	1501
CO-154	3084	1419

Table 3.6: The total number of genes flagged as differentially expressed in mutants compared to the empty vector control CO-250

3.1.2.1 MmyB homologs

All three of the mutants CO-121, CO-124 and CO-154 have overexpressed the MmyB homolog that they have an additional copy of (CO-121 = *sco7140*, CO-124 = *sco0236*, CO-154 = *sco6926*). In both the cases of CO-121 and CO-124, the MmyB homolog that they are specific to does not have altered transcription in either of the other two mutants (Figure 3.11). The CO-154 mutant on the other hand, the homolog *sco6926* has increased transcription across all three mutants.

In all three mutants there are at least eight of the homologs that have differential expression. The CO-121 mutant is the only one of the three that doesn't have any of the homologs under expressed and has ten homologs overexpressed. The CO-124 mutant has nine homologs over expressed and two under expressed, while CO-154 has seven of the homologs overexpressed and just one under expressed. Many of the homologs are shared between all three mutants. All of CO-121, CO-124 and CO-154 have *sco0110, sco0233, sco0307, sco6926* and *sco7767* overexpressed compared to CO-250. The CO-121 mutant has two homologs that have differential expression unique to the mutant, *sco6676* and the extra copy *sco7140*. The CO-121 mutant also shares overexpression of the *sco6539* gene with CO-124. The CO-124 mutant has one unique bit of differential expression which is the under expressed *sco4680* gene. It shares under expression of the *sco7706* homolog with the CO-154 mutant.

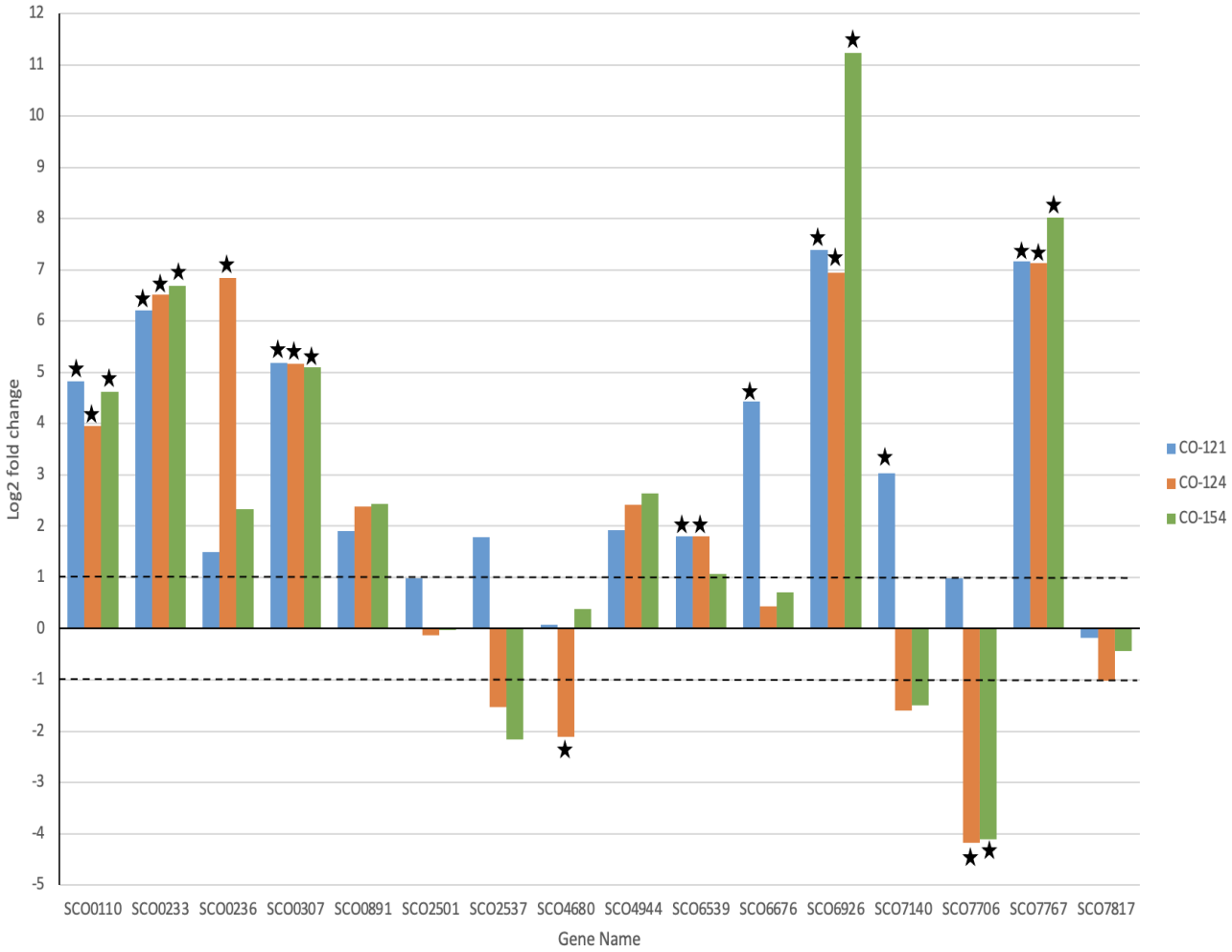


Figure 3.11: The log2 fold change of RNA transcription for the sixteen MmyB homologs in the chromosomal genome of *Streptomyces coelicolor* A3(2) in CO-121, CO-124 and CO-154 compared to CO-250 in YEME media. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

The directly downstream genes to the MmyB homologs have some differentially expressed genes in the YEME media with two genes conserved across the mutants (Table 3.7). The *sco6540* gene is overexpressed in all three of the mutants while the *sco7140* gene is under expressed in all three. CO-121 has the least differential expression of the directly downstream genes with three while CO-124 has the most with seven in total and CO-154 has five differentially expressed genes.

MmyB homolog	Directly Downstream Gene	Function	Log2 fold change		
			CO-121	CO-124	CO-154
sco0110	*sco0111*	SDR oxidoreductase	5.11542458	3.64465143	3.91534713
sco0233	*sco0234*	SDR oxidoreductase	N/A	N/A	N/A
sco0236	*sco0237*	SDR oxidoreductase	6.45809692	N/A	5.08237742
sco0307	*sco0306*	Pseudogene	N/A	N/A	N/A
sco0891	*sco0892*	SDR oxidoreductase	-2.4821151	-3.6309406	-3.8552053
sco2501	*sco2502*	Transport protein	0.96004547	-3.3924834	-2.3266206
sco2537	*sco2538*	Unknown function	1.24467008	0.64236906	0.98917866
sco4680	*sco4681*	SDR	1.26612918	-1.9595171	2.08398003
sco4944	*sco4945*	SDR alcohol dehydrogenase	-1.6243729	-3.5214959	-1.5359671
sco6539	*sco6540*	Unknown function	8.23592217	8.7632603	8.7968528
sco6676	*sco6677*	ABC transporter	-0.6178594	-5.3790118	-7.4717781
sco6926	*sco6925*	Membrane protein	N/A	N/A	N/A
sco7140	*sco7141*	Nitroreductase	-1.9898281	-4.944647	-5.7921707
sco7706	*sco7707*	Unknown function	4.47175964	4.33236206	4.91065544
sco7767	*sco7768*	Unknown function	N/A	N/A	N/A
sco7817	*sco7818*	SDR oxidoreductase	-0.4756438	0.06232804	-1.1337448

Table 3.7: The log2 fold change of the directly downstream genes to MmyB homologs in the CO-121, CO-124 and CO-154 compared to the CO-250 control, grown in YEME media. Values in red indicate a p-value < 0.05 for log2 fold changes greater than -1, while values in green indicate a p-value <0.05 for log2 fold changes greater than 1.

The supercoiling hypersensitive cluster has differential expression in the YEME media as well (Figure 3.12). There are seventeen genes overexpressed in CO-121 and one under expressed while CO-124 only has one along with four genes under expressed. CO-154 has the most differential expression with twenty-three genes over expressed within the supercoiling cluster.

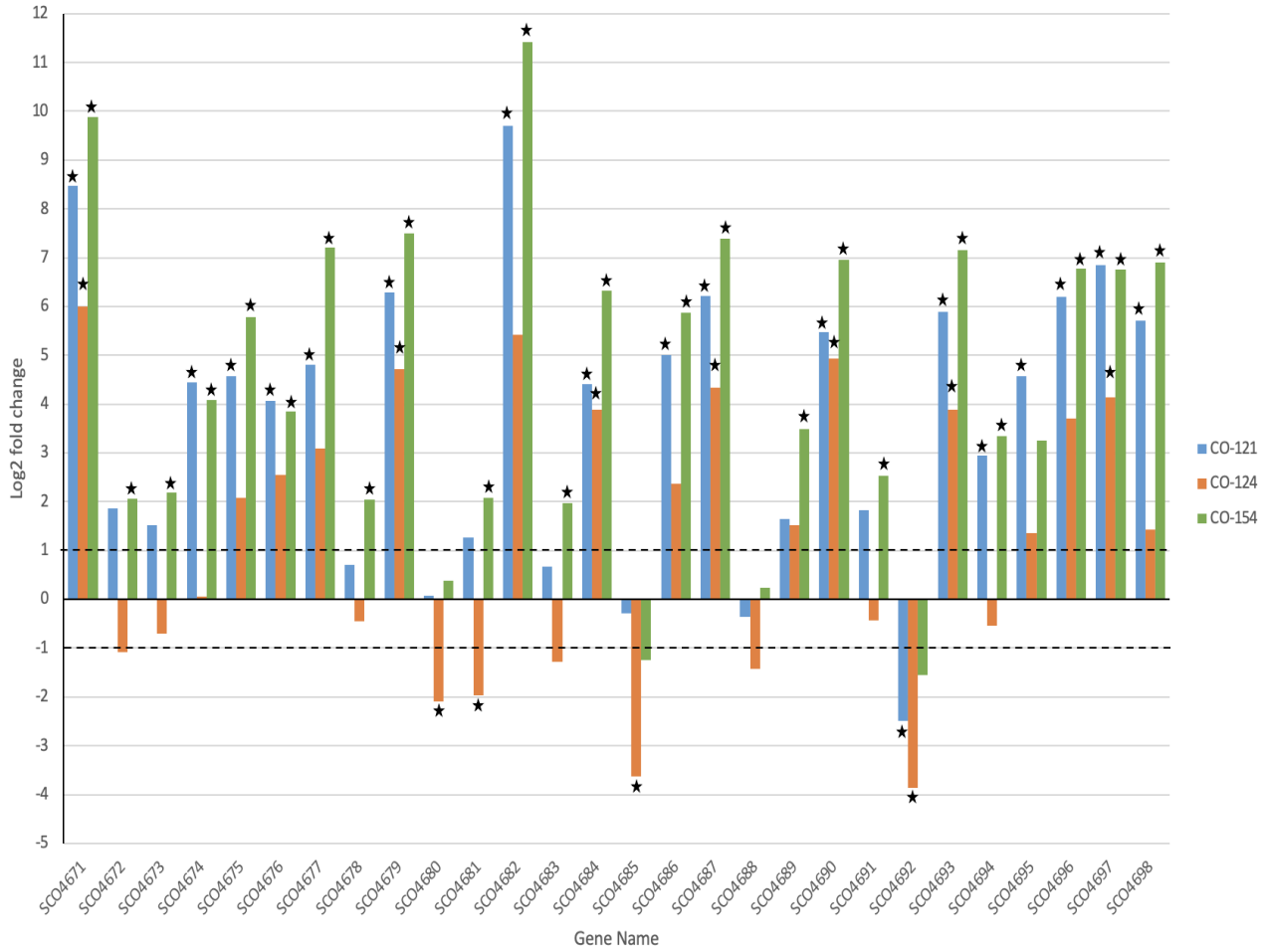


Figure 3.12: The log2 fold change of RNA transcription for the supercoiling hypersensitive cluster in CO-121, CO-124 and CO-154 compared to the empty vector control CO-250 grown in YEME media. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

3.1.2.2 Region 2

The second antiSMASH predicted region, close to the *sco0236* MmyB homolog, has differentially expressed genes in all three mutants (Figure 3.13). CO-121 has seventeen genes overexpressed, while CO-124 has fourteen genes overexpressed and one under expressed. CO-154 has the same gene under expressed as CO-124 along with another fifteen genes overexpressed.

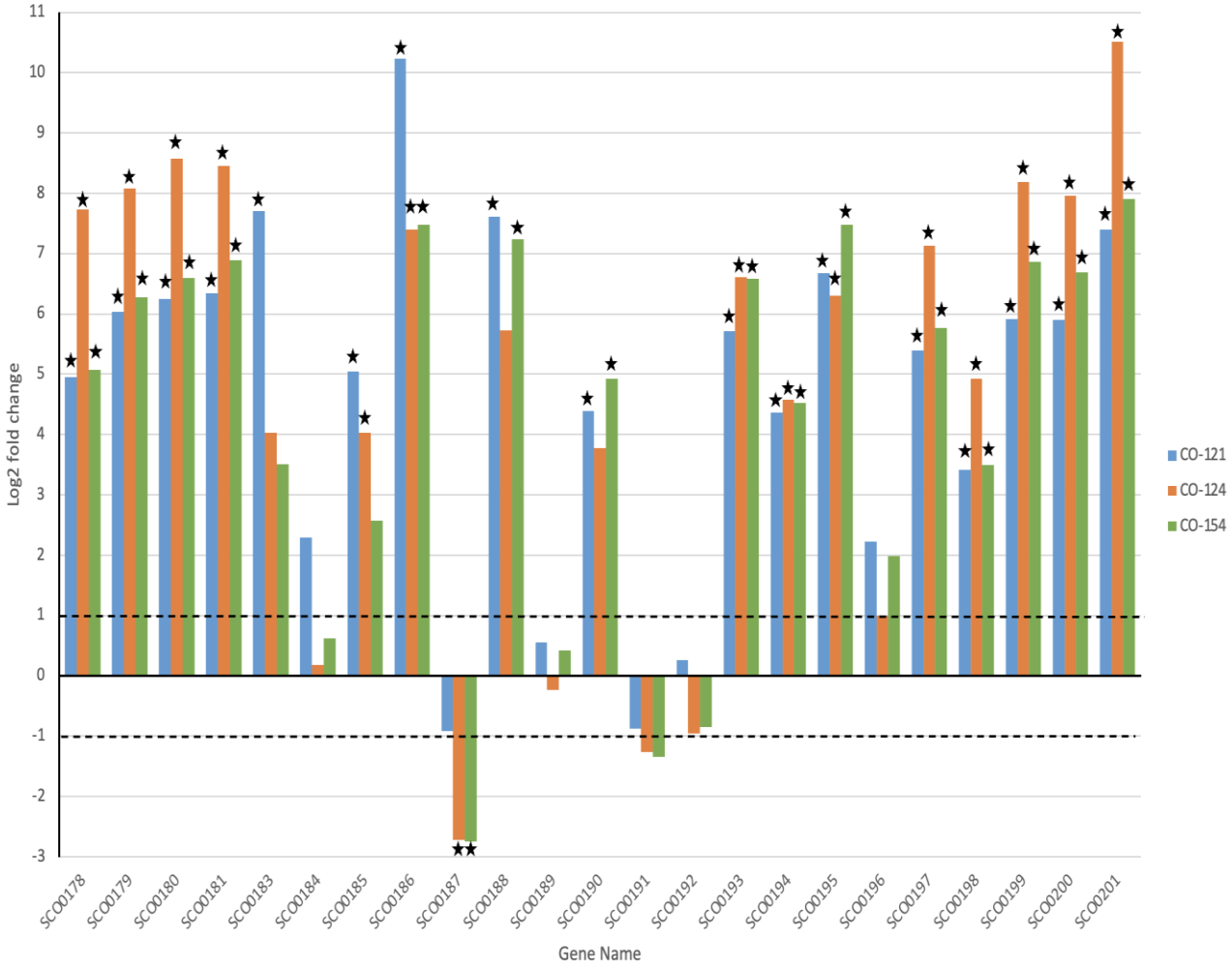


Figure 3.13: The log2 fold change of RNA transcription for the second antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to the empty vector control CO-250 grown in YEME media. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

3.1.2.3 Region 3

All three mutants have altered expression within the third predicted cluster, close to the *sco0236* homolog (Figure 3.14). CO-121 has the most change in expression with ten genes overexpressed and one under expressed while CO-124 has seven overexpressed and three under expressed. CO-154 has the least number of genes overexpressed with six but the most under expressed with four.

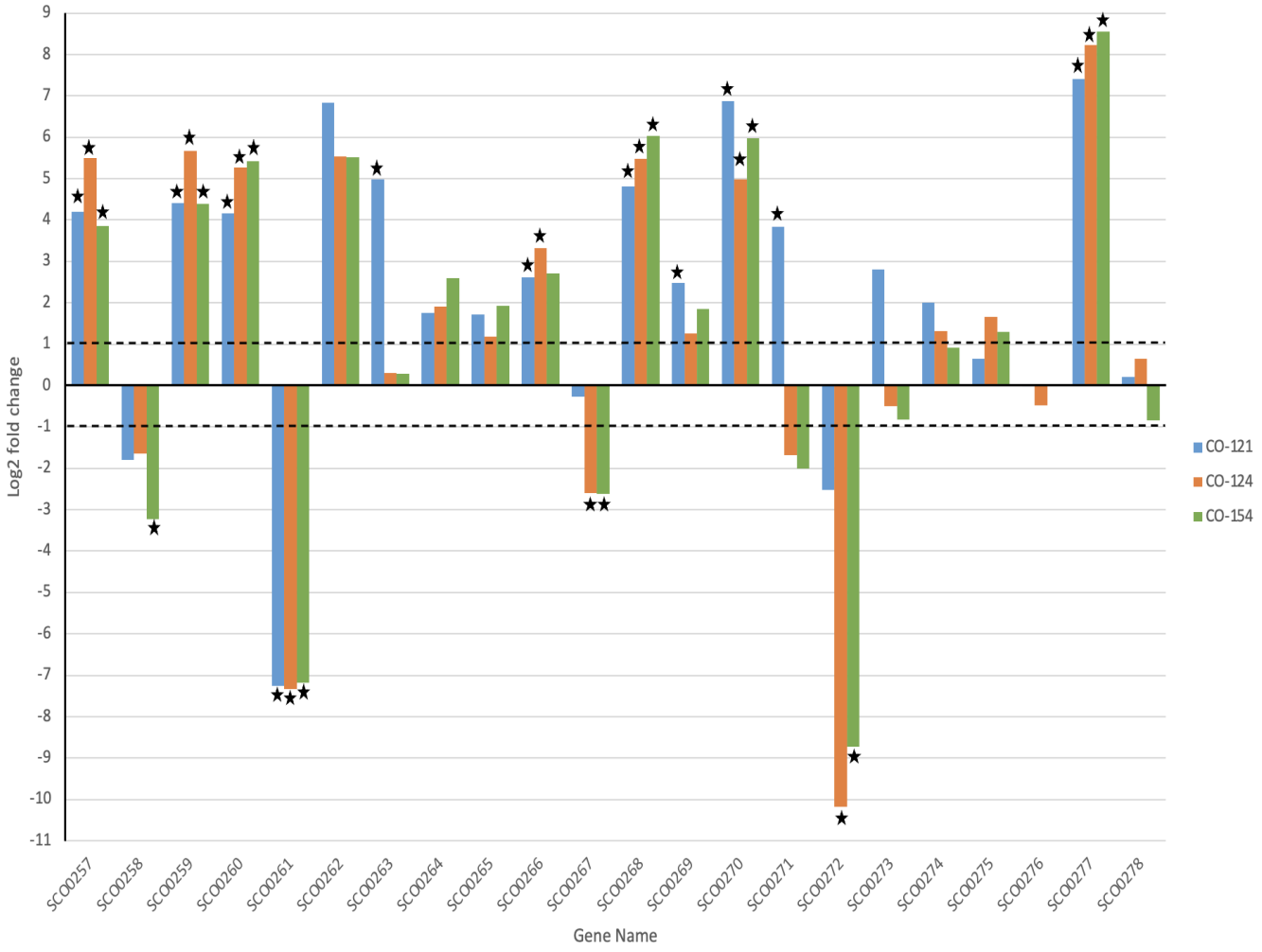


Figure 3.14: The log2 fold change of RNA transcription for the third antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to the empty vector control CO-250 grown in YEME media. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

3.1.2.4 Region 11

The actinorhodin cluster had large amounts of differential expression across all three mutants with the majority being overexpressed, including nearly all of the biosynthetic genes (Figure 3.15). CO-121 has thirty one genes overexpressed but only one under expressed. This is less than both CO-124 and CO-154 that have eleven and eight genes under expressed respectively.

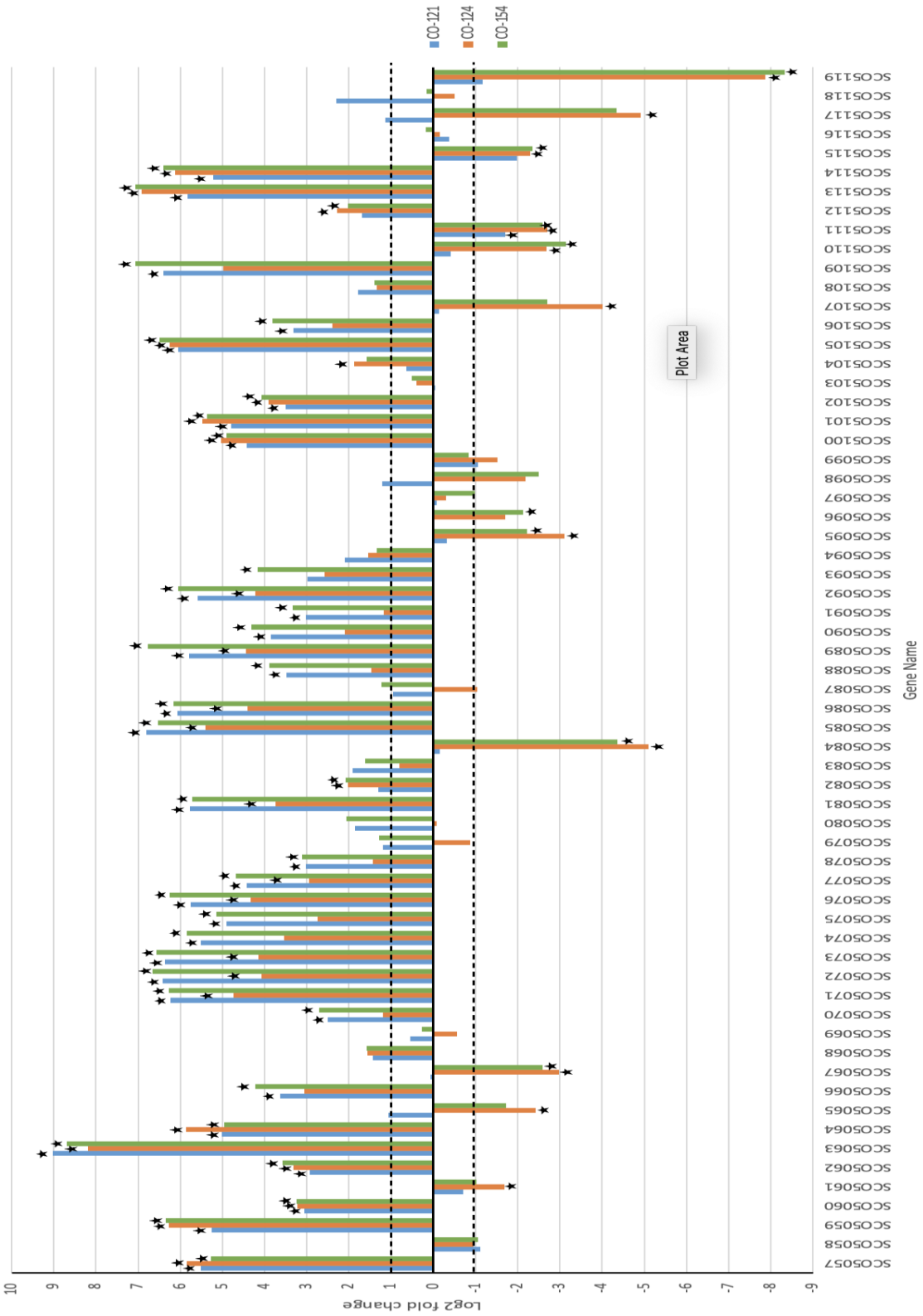


Figure 3.15: The log2 fold change of RNA transcription for the eleventh antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to the empty vector control CO-250 grown in YEME media. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

3.1.2.5 Region 21

The SapB cluster which contains the MmyB homolog *sco6676,* has a large amount of differential expression across the cluster in each of the mutants (Figure 3.16). While each mutant has differentially expressed genes only CO-121 has *sco6676* overexpressed along with four other genes and another gene under expressed. CO-124 and CO-154 have six and five genes under expressed respectively with many of them the same between the two mutants.

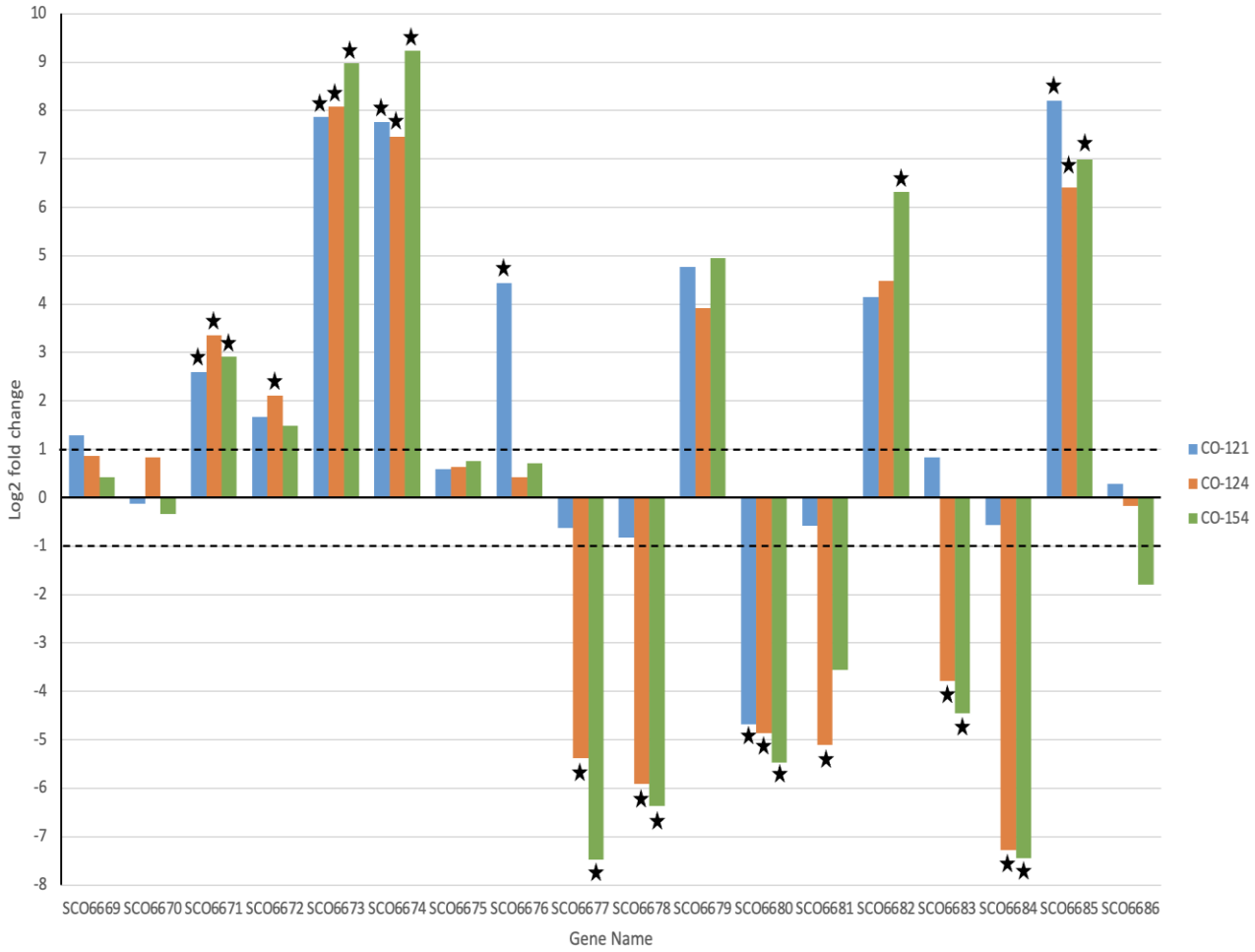


Figure 3.16: The log2 fold change of RNA transcription for the twenty-first antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to the empty vector control CO-250 grown in YEME media. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

3.1.2.6 Region 24

The twenty fourth cluster is of interest due to the presence of the *sco6926* MmyB homolog within the BGC prediction and CO-154 having an extra copy of the gene overexpressed. The *sco6926* is overexpressed across all three of the mutants with CO-121 having five more genes overexpressed while CO-124 and CO-154 both only have one other gene overexpressed (Figure 3.17). CO-124 and CO-154 also have one gene, *sco6935* under expressed.

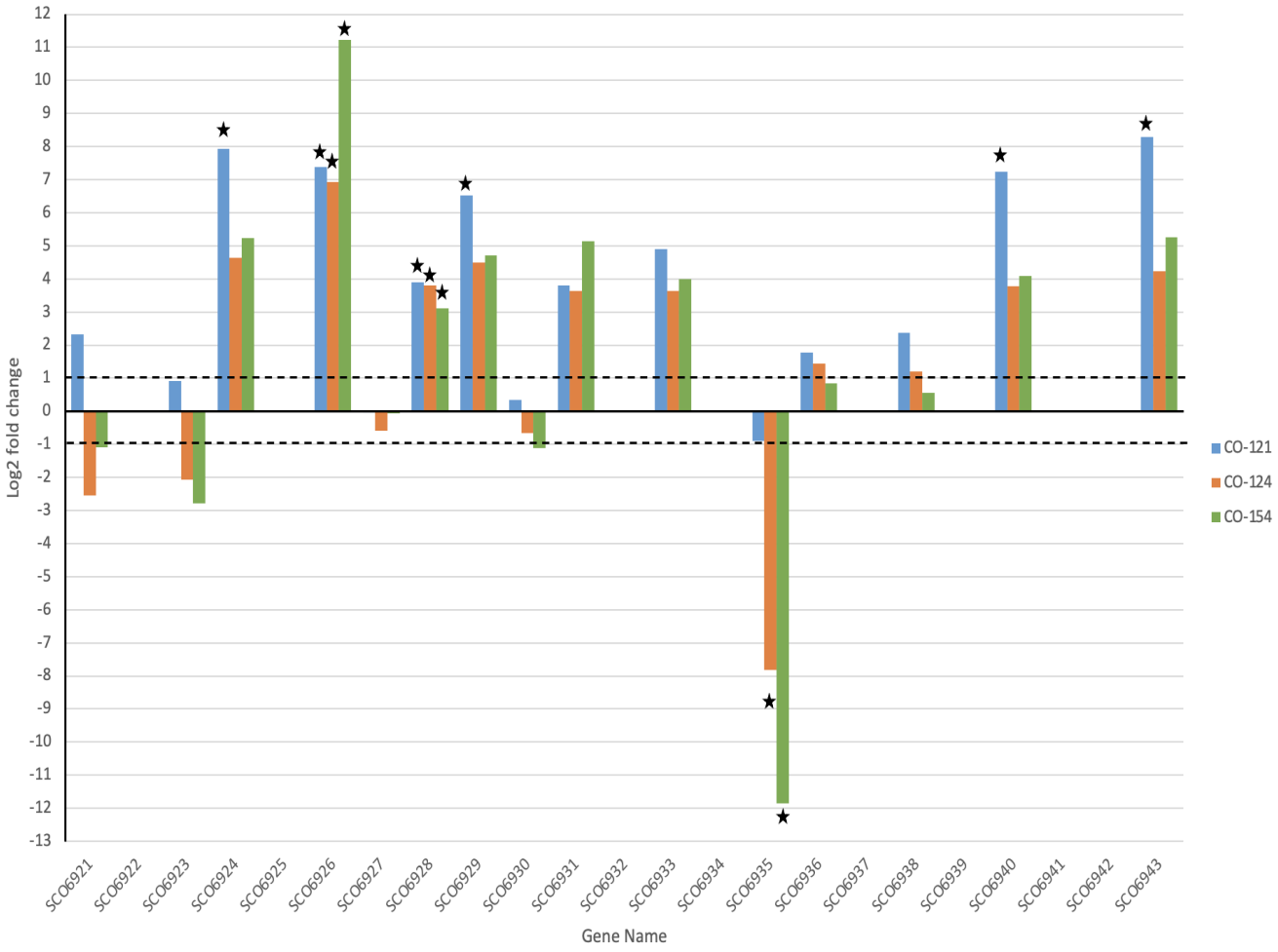


Figure 3.17: The log2 fold change of RNA transcription for the twenty-fourth antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to the empty vector control CO-250 grown in YEME media. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

3.1.3 Metabolomics of RNA sequenced cultures

3.1.3.1 SMM cultures

The metabolome of the three mutants CO-121, CO-124 and CO-154 were compared against the parent strain M145 by LCMS. The three mutants and M145 parent strain were cultured in triplicate for 4 days at 30°C, 200 rpm in SMM before the cells were harvested by centrifugation. The supernatant was separated and ethyl acetate liquid-liquid extraction was completed before running high resolution LCMS/MS.

In CO-121, two peaks were identified in the mutant that were not in M145, with retention times of 17 and 23.5 minutes (Figure 3.18).

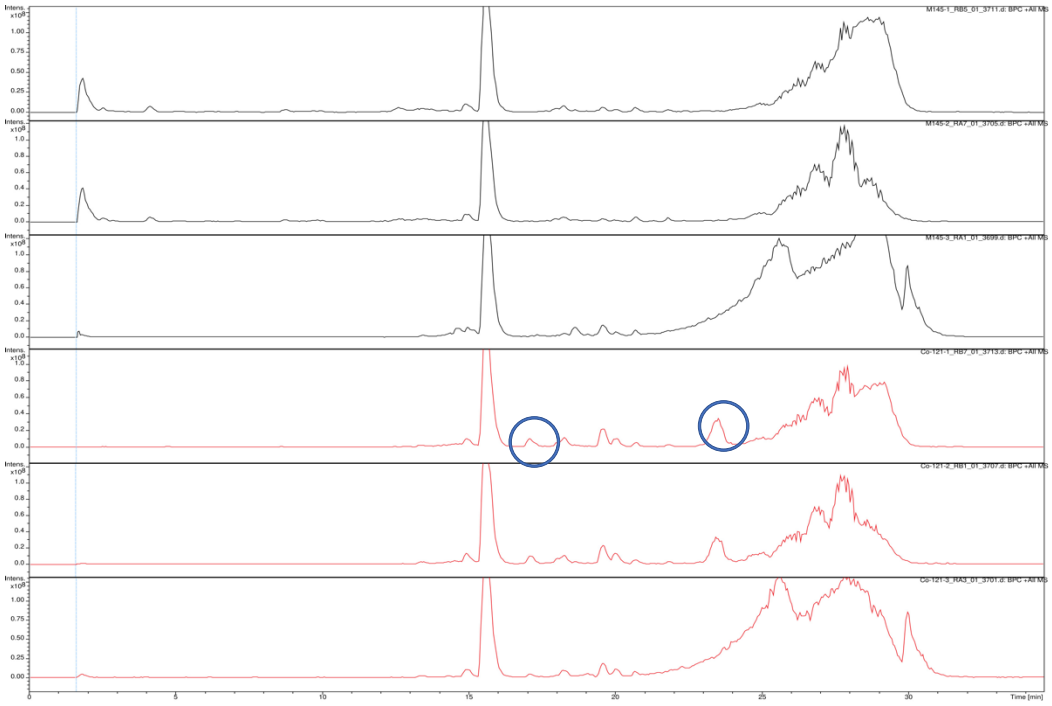
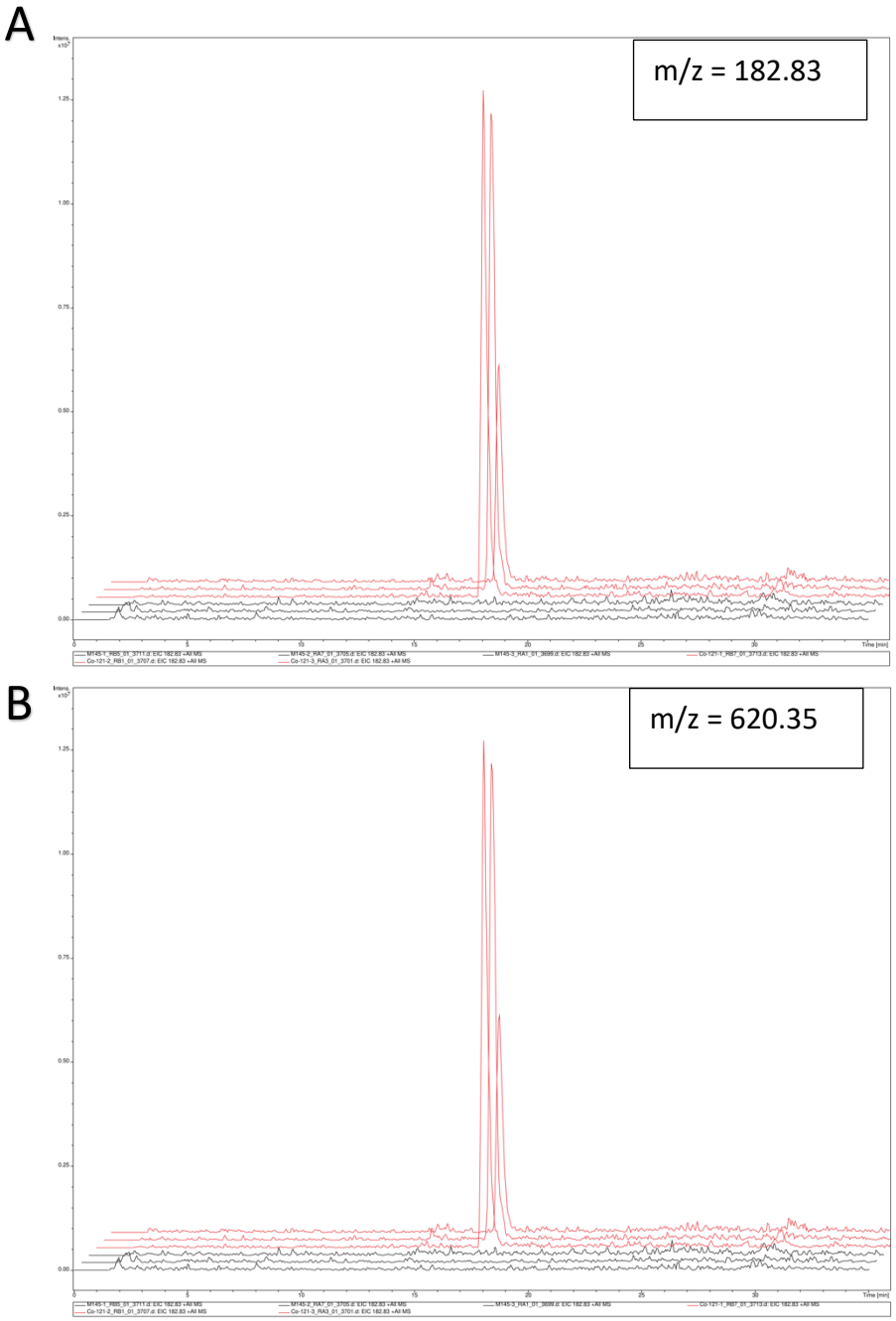


Figure 3.18: Base peak chromatogram of *S. coelicolor* M145 (black) and CO-121 (red) after growth in SMM, in triplicate. The traces are similar across all replicates with new peaks identified within the circle for CO-121.

Extracted ion chromatograms were isolated for each of the peaks (Figure 3.19) with Bruker Data Analysis software used to predict the molecular formula of the metabolites. The chemical formula was only able to be predicted for one of the peaks with a m/z 620.35 with a predicted formula of $C_{28}H_{46}N_9O_7$.



Molecular formula prediction: $C_{28}H_{46}N_9O_7$

Figure 3.19: Extracted ion chromatograms (A: m/z = 182.83 B: m/z = 620.35) for *S. coelicolor* M145 (black) and CO-121 (red).

In the CO-124 (Figure 3.20) and CO-154 mutants (Figure 3.21) in depth analysis of the LCMS peaks identified a peak in both of the mutants at a retention time of 11 minutes that was not in M145.

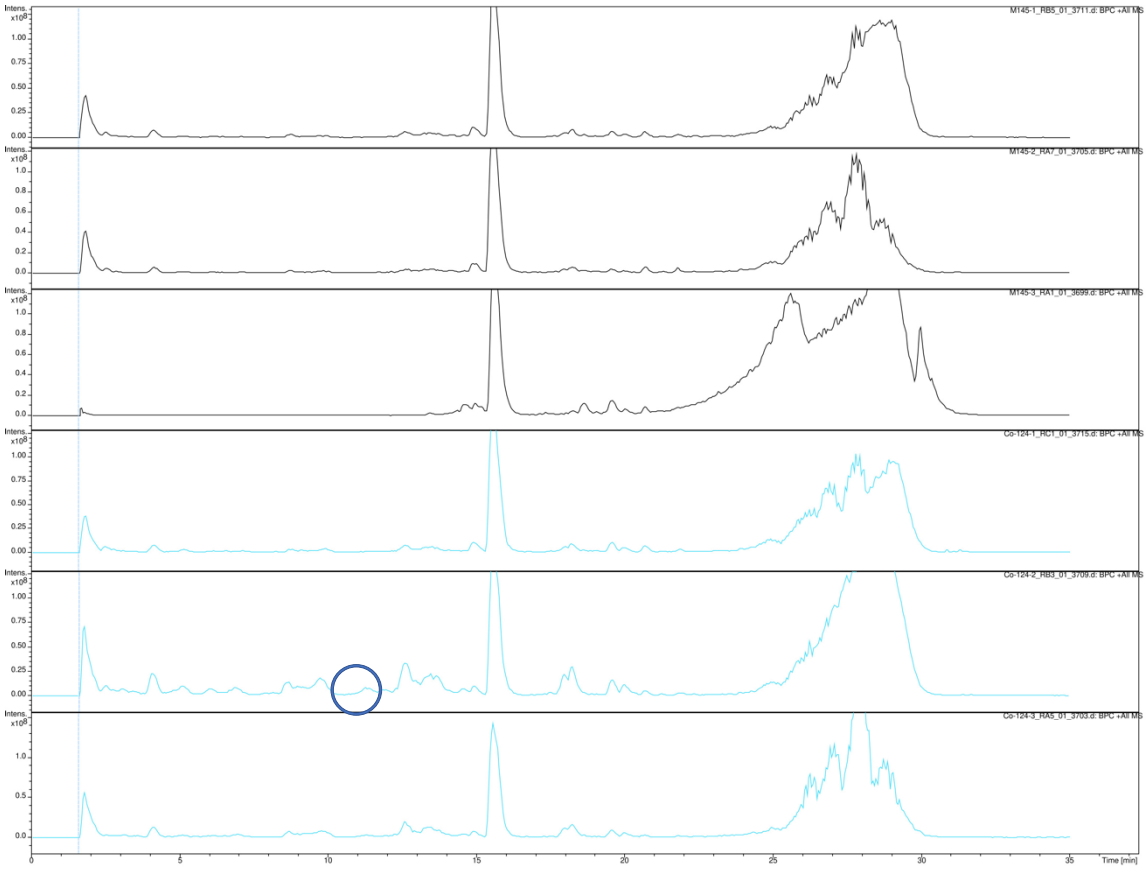


Figure 3.20: Base peak chromatogram of *S. coelicolor* M145 (black) and CO-124 (blue) after growth in SMM, in triplicate. The traces are similar across all replicates with new peaks identified within the circle for CO-124.

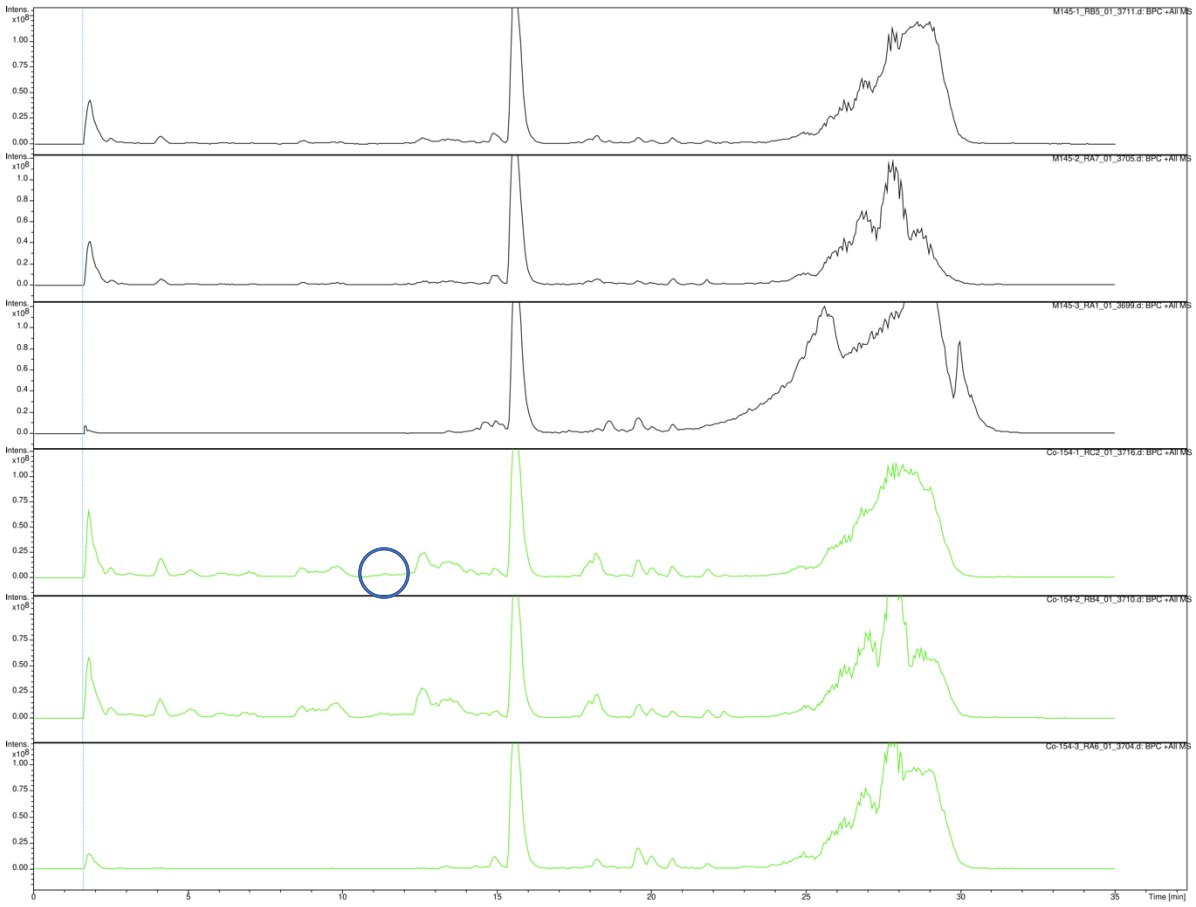
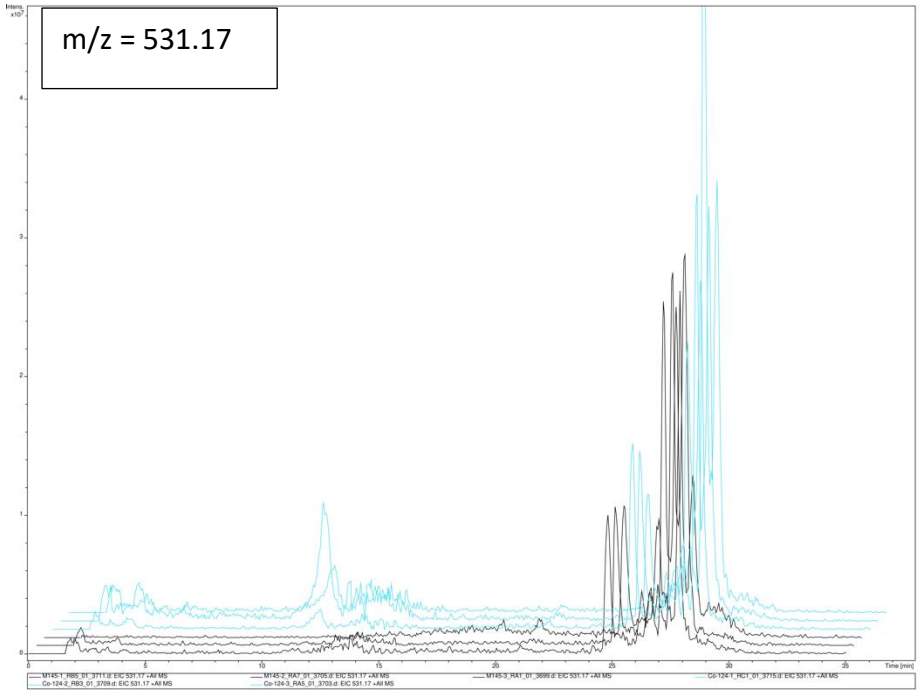


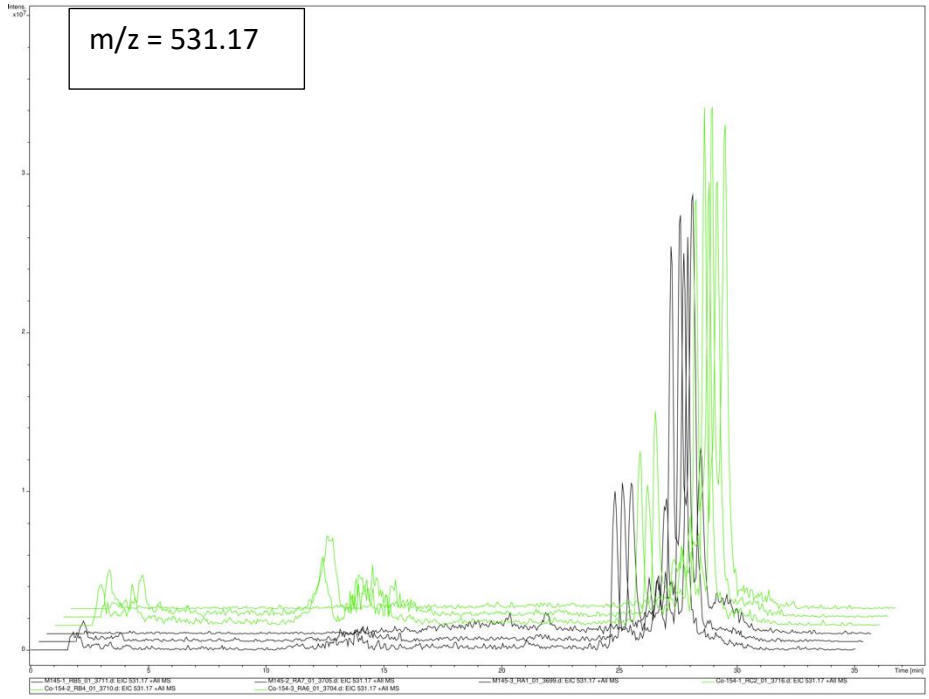
Figure 3.21: Base peak chromatogram of *S. coelicolor* M145 (black) and CO-154 (green) after growth in SMM, in triplicate. The traces are similar across all replicates with new peaks identified within the circle for CO-154.

Extracted ion chromatograms were isolated for each of the peaks in CO-124 and CO-154 (Figure 3.22) with Bruker Data Analysis software used to predict the molecular formula of the metabolite. The chemical formula was predicted as $C_{36}H_{23}N_2O_3$.

A



B



Molecular formula prediction: $C_{36}H_{23}N_2O_3$

Figure 3.22: Extracted ion chromatograms for a m/z of 531.17 comparing *S. coelicolor* M145 (black) against CO-124 (A: blue) and CO-154 (B: green)

3.1.3.2 YEME cultures

The metabolome of the four mutants CO-121, CO-124, CO-154 and CO-250 along with the parent strain M145 were investigated for the YEME RNA sequencing cultures. The four mutants and M145 were cultured in triplicate for 4 days at 30°C, 200 rpm in liquid YEME before the cells were harvested by centrifugation. The supernatant was separated and ethyl acetate liquid-liquid extraction was completed before running high resolution LCMS/MS.

The M145 and CO-250 samples were compared to each other to determine if plasmid integration altered the metabolome. The BPC of M145 and CO-250 are very similar with M145 having peaks of much greater intensity (Figure 3.23).

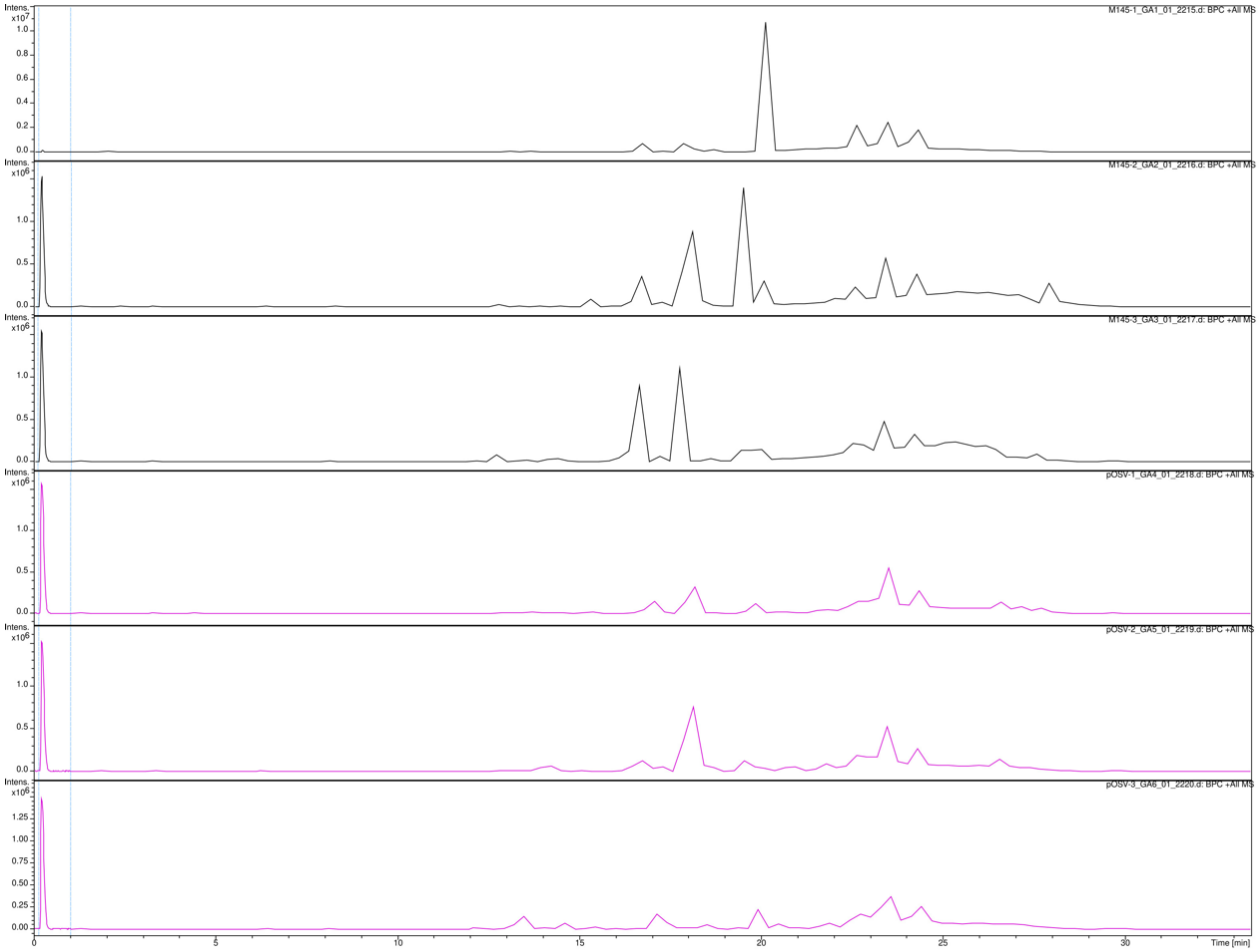


Figure 3.23: Base peak chromatograms of *S. coelicolor* M145 (black) and CO-250 (purple) after growth in YEME, in triplicate. The traces are similar across all replicates with no new peaks unique to CO-250.

The similarity between M145 and CO-250 led to CO-250 being used as the control for comparison against the other three mutants. When comparing CO-121 and CO-250 (Figure 3.24), there is a drastic change in the BPC at a retention time of 15 minutes in one of the CO-121 replicates.

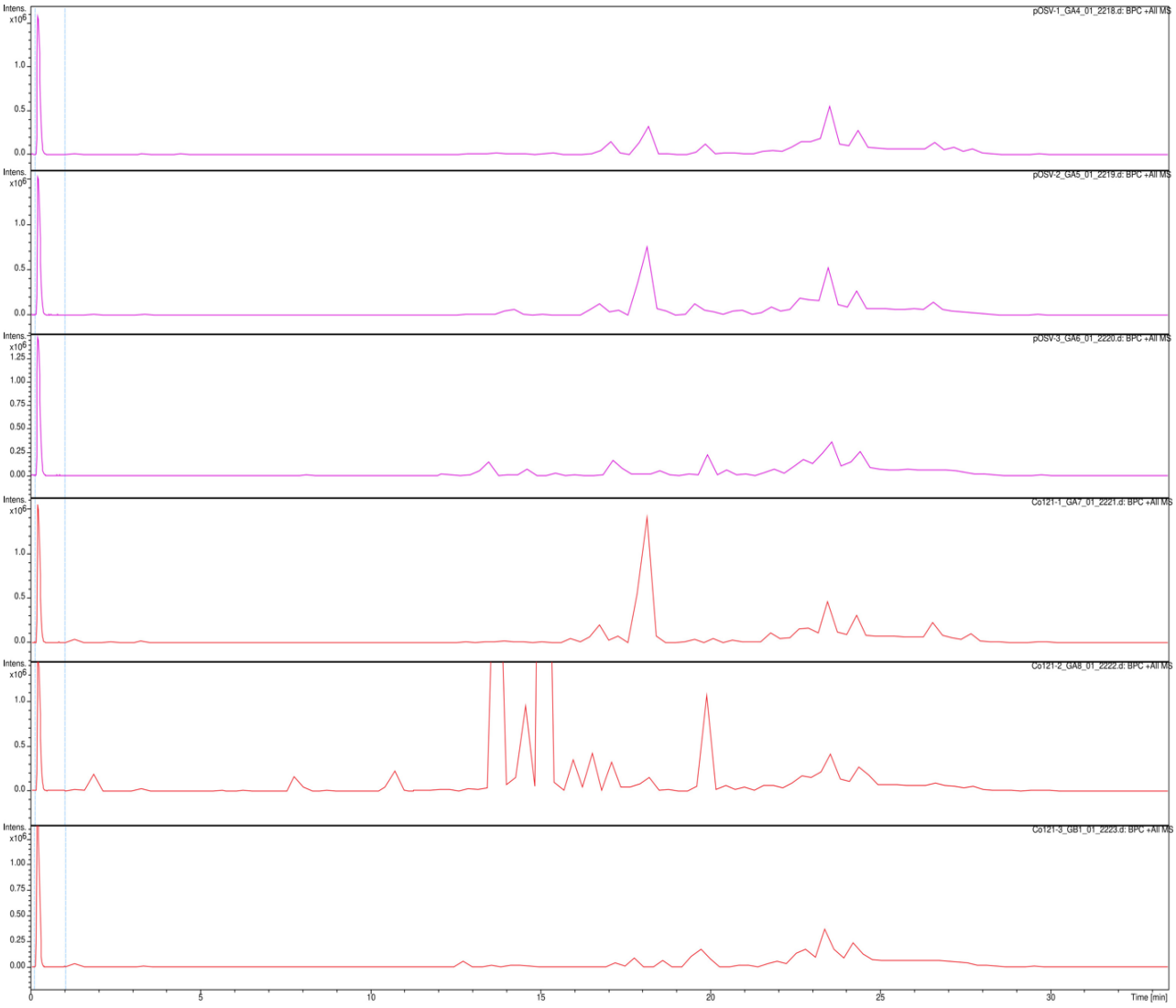


Figure 3.24: Base peak chromatograms of CO-250 (purple) and CO-121 (red) after growth in YEME, in triplicate. The traces are similar across all replicates with an anomalous peak in the second CO-121 replicate

The LCMS traces of CO-124 and CO-154 are very similar to CO-250 with no novel peaks presenting themselves in either mutant (Figure 3.25) with only the intensities differing between the BPC.

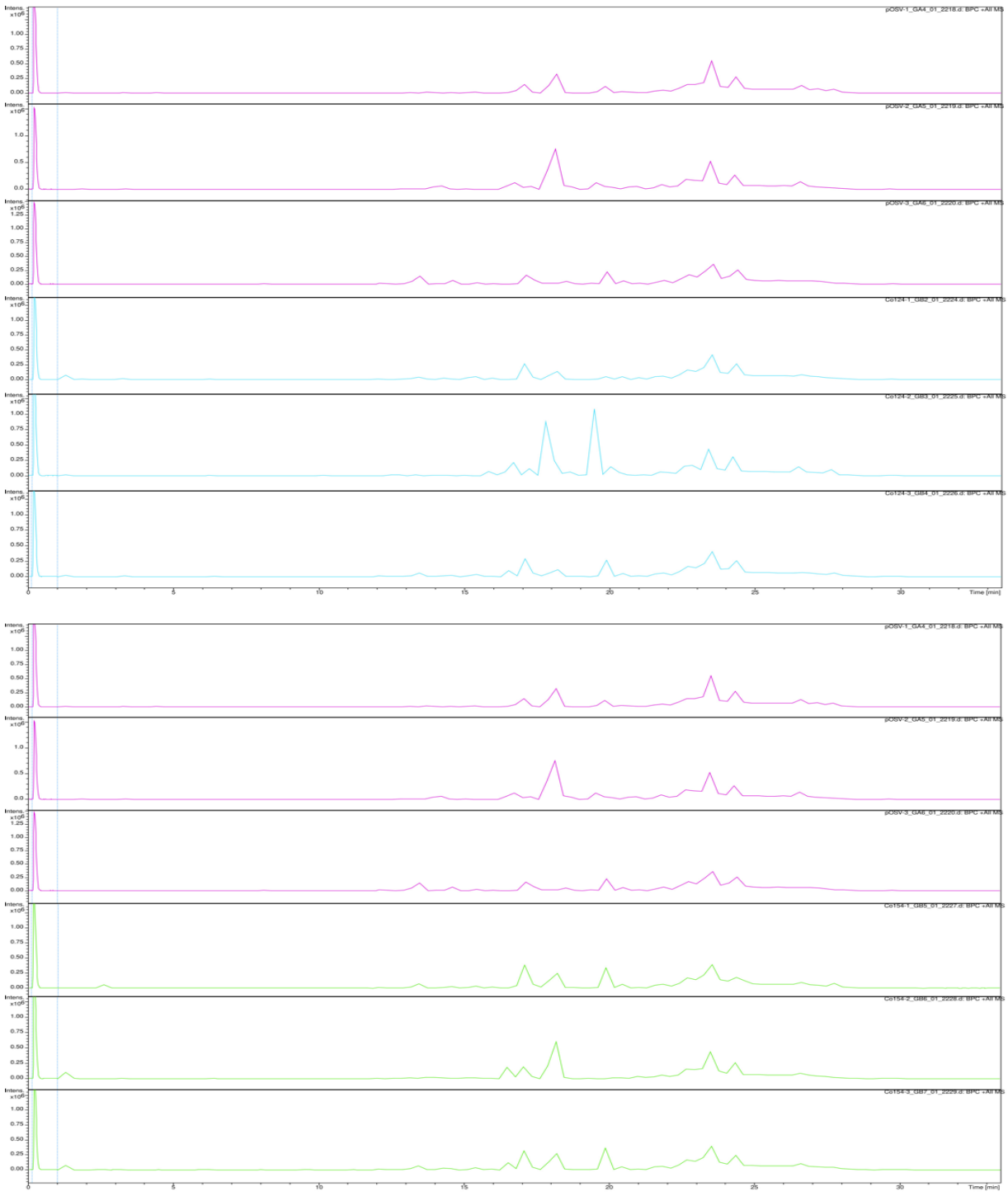


Figure 3.25: Base peak chromatograms of CO-250 (purple), CO-124 (blue) and CO-154 (green) after growth in YEME, in triplicate. The traces are similar across all replicates with no unique peaks in either CO-124 and CO-154.

3.2 *Rhodococcus* investigations

3.2.1 Genome mining

An antiSMASH analysis was performed against twenty-eight publicly available *Rhodococcus* genomes sourced from the NCBI Genbank. The predictive modelling estimated at least thirteen clusters for each genome with up to twenty eight at the most for *R. opacus* PD630 (Table 3.8). In total across the twenty eight different strains of bacteria 519 clusters were predicted.

Strain	Predicted cluster number
R. aetherivorans IcdP1	18
R. erythropolis BG43	18
R. erythropolis CCM2595	17
R. erythropolis PR4	18
R. erythropolis R138	17
R. equi 103S	15
R. equi DSSKP-R-001	15
R. fascians D188	18
R. jostii RHA1	22
R. opacus 1CP	25
R. opacus B4	26
R. opacus PD630	28
R. opacus R7	24
R. qingshengii djl-6-2	18
R. rhodochrous NCTC10210	16
R. ruber P14	17
R. ruber YYL	16
R. sp. 008	17
R. sp. 2G	14

R. sp. B7740	17
R. sp. BH4	17
R. sp. H-CA8f	18
R. sp. MTM3W5.2	16
*R. sp.*p52	13
R. sp. PBTS1	13
R. sp. PBTS2	19
R. sp. WB1	17
R. sp. WMMA185	14

Table 3.8: Publicly available *Rhodococcus* genomes from NCBI Genbank and the antiSMASH cluster prediction number

3.2.2 MmyB homolog prevalence

BLAST[69] analysis was used to compare the twenty eight *Rhodococcus* genomes used for genome mining against the MmyB protein. BLAST was able to identify seventy-six homologs across the twenty-eight different strains. The prevalence of MmyB to an individual species can vary from only one homolog all the way to eight.

As seen in Figure 3.26, the MmyB homologs in *Rhodococcus* group into three distinct clades. There are some genes found across multiple species. There is a group of three genes; *RER_RS29060, XU06_RS25955* and *C1M55_RS27190* found in *R. erythropolis* PR4, *R. erythropolis* BG43 and *R. qingshengii* djl-6-2 respectively. There are two other pairs that are found in multiple strains of the same species (REQ_RS22620 and C7H75_RS23790 are both in *R. fascians* strains while XU06_RS28080 and O5Y_RS27670 are both in *R. erythropolis* strains). There's one final pair found in two species, the *A3L23_RS07750* gene from *R. fascians* D188 and the *A3Q41_RS08810* found in *R. sp.* PBTS2.

Figure 3.26: Phylogeny tree of *Rhodococcus* MmyB homologs alongside *mmyB* and *Chloroflexus aurantiacus* MltR. The branch length of each homolog is represented by the boxed number by the gene name which indicates the average number of amino acid substitutions per site.

Strain	Gene Name	Distance to nearest BGC (kb)
R. aetherivorans IcdP1	*AAT18_RS01415*	1.7
	AAT18_RS21070	27.8
R. erythropolis BG43	*XU06_RS25955*	2.7
	XU06_RS28080	185.9
R. erythropolis CCM2595	*O5Y_RS25605*	4.9
	O5Y_RS27670	154.3
R. erythropolis PR4	*RER_RS05870*	194.8
	RER_RS29060	2.7
	RER_RS30975	148.2
pREL1	*RER_RS00200*	No predicted clusters
R. erythropolis R138	*H351_RS04340*	155.1
	H351_RS06590	2.7
R. equi 103S	*REQ_RS22620*	147.7
R. equi DSSKP-R-001	*C7H75_RS23790*	162.4
R. fascians D188	*A3L23_RS00815*	82.9
	A3L23_RS02225	Within predicted cluster
	A3L23_RS06235	22.4
	A3L23_RS07750	Within predicted cluster
	A3L23_RS14960	94.0
R. jostii RHA1	*RHA1_RS02730*	101.6
	RHA1_RS02780	111.0
	RHA1_RS09515	224.2
pRHL2	*RHA1_RS40960*	46.4
R. opacus 1CP	*R1CP_RS19215*	Within predicted cluster
	R1CP_RS28630	244.2
R. opacus B4	*ROP_RS08095*	275.2

pROB01	*ROP_RS36215*	73.4
R. opacus PD630	*PD630_RS08950*	109.1
	PD630_RS21525	36.9
	PD630_RS21650	64.3
	PD630_RS28605	234.5
	PD630_RS34220	356.9
	PD630_RS34435	408.4
	PD630_RS34505	421.1
	PD630_RS34730	469.3
R. opacus R7	*EP51_01710*	91.8
	EP51_05660	240.4
	EP51_26910	590.2
	EP51_26970	605.0
	EP51_26995	609.4
	EP51_27025	615.0
	EP51_33145	229.8
pPDG3	*EP51_44330*	No predicted clusters
R. qingshengii djl-6-2	*C1M55_RS27190*	2.7
	C1M55_RS29565	172.8
R. rhodochrous NCTC10210	*CKW34_RS06975*	226.5
R. ruber YYL	*CSW53_RS17445*	30.6
R. ruber P14	*CS378_RS10510*	22.5
R. sp. 008	*AOT96_RS07160*	4.9
	AOT96_RS09230	171.1
R. sp. 2G	*BO226_RS18795*	22.6
R. sp. B7740	*NY08_RS02975*	505.5
	NY08_RS05860	125.9
	NY08_RS08690	82.4

	NY08_RS10435	57.7
	NY08_RS11105	202.7
	NY08_RS17795	137.9
	NY08_RS23685	231.6
R. sp. BH4	*A0W34_RS25975*	2.7
	A0W34_RS28075	172.2
R. sp. H-CA8f	*CPI83_RS01265*	159.1
	CPI83_RS03325	4.9
R. sp. MTM3W5.2	*BTZ20_RS17675*	16.2
R. sp. p52	*IM25_RS05365*	350.3
R. sp. PBTS1	*A3Q40_RS14345*	70.7
R. sp. PBTS2	*A3Q41_RS01460*	270.6
	A3Q41_RS07355	118.4
	A3Q41_RS08810	Within predicted cluster
	A3Q41_RS10390	22.3
	A3Q41_RS14475	24.4
	A3Q41_RS16035	80.5
	A3Q41_RS22790	121.8
R. sp. WB1	*A4U64_RS03520*	35.0
	A4U64_RS23355	1.7
R. sp. WMMA185	*BFN03_RS05680*	408.3

Table 3.9: List of MmyB homologs detected using BLAST and their proximity to a cluster predicted by antiSMASH. Genes found within a cluster prediction are highlighted in green, while those within 5 kilobases are highlighted in orange.

The antiSMASH and BLAST analysis was combined to link the MmyB homologs to the closest BGC to predict if it may be involved in cluster regulation (Table 3.9). There are ten homologs within five kilobases of a cluster and there is four that are found within the cluster prediction.

3.2.3 Mutant construction

The plasmids for creation of the *Rhodococcus* mutants were created and insertion of the targeted gene into the pOSV556 backbone was confirmed using PCR amplification with a primer pair binding to the flanking region of the *ermE** cloning site (Figure 2.1). All plasmids pC0254 – pC0267 bar pC0258 were confirmed to have the gene insert by PCR and electrophoresis (Figure 3.27). The pC0258 plasmid was confirmed using a restriction digest and electrophoresis along with the other plasmids again (Appendix 1).

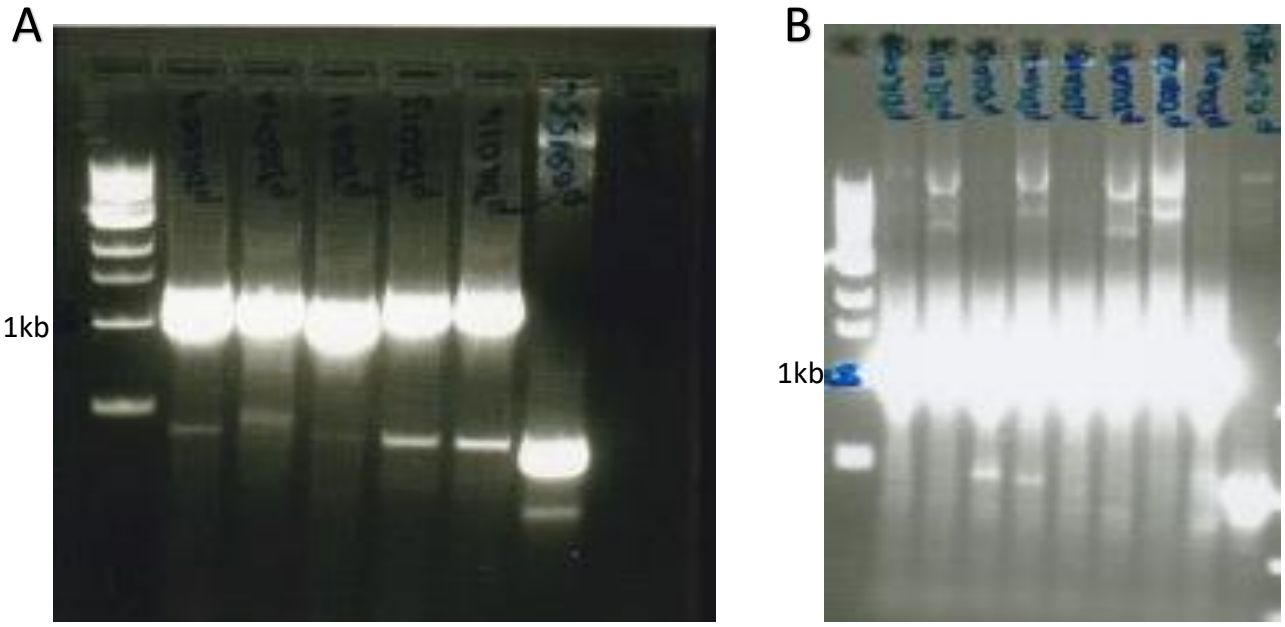


Figure 3.27: PCR testing of plasmids to check ligation of gene insert. The primer pair (pOSV02: pOSV03) was used for all reactions with a band expected corresponding to approximately 1.2kb was obtained for all plasmids minus pC-0258. The band expected for pOSV556 is just under 500bp

A) Lanes correspond to: **1.** NEB 1kb ladder **2.** pC-0255 **3.** pC-0256 **4.** pC-0257 **5.** pC-0259 **6.** pC-0262
7. pOSV556 positive control

B) Lanes correspond to: **1.** NEB 1kb ladder **2.** pC-0254 **3.** pC-02560 **4.** pC-02561 **5.** pC-02563 **6.** pC-02564
7. pC-02565 **8.** pC-02566 **9.** pC-02567 **10.** pOSV556 positive control

The mutant *Rhodococcus* were created using electroporation of *R. jostii* RHA1 and *R. erythropolis* PR4 for all sixteen plasmids. After antibiotic culturing, successful integration of the plasmid was confirmed using a primer pair targeted between the cloning site and the genome

integration site (Figure A1 and A2). Two plasmids (pC0256 and pC0260) were confirmed to have successfully integrated creating *R. jostii* RHA1 mutants (CO-218 and CO-222) alongside *R. erythropolis* PR4 mutants (CO-236 and CO-240). Control mutants containing the empty pOSV556 vector were also created for *R. jostii* RHA1 (CO-213) and *R. erythropolis* PR4 (CO-231).

3.2.4 Metabolomics of mutants

The metabolic profile of CO-213 and CO-231 were compared to *R. jostii* RHA1 and *R. erythropolis* PR4 respectively. Cultures were grown in triplicate for four days at 30°C, 200 rpm in liquid LB media. The cells were harvested by centrifugation and the supernatant underwent ethyl acetate liquid-liquid organic extraction before analysis by LCMS.

Comparing CO-213, CO-218 and CO-222 metabolome against RHA1 showed no new and unique peaks for the mutants. The same result was seen when comparing CO-231, CO-236 and CO-240 to *R. erythropolis* PR4, with no new unique peaks occurring in the mutants (Figure A43 – A45).

3.3 MmyB homolog bioinformatic investigation

3.3.1 Phylogenetic analysis

The protein sequences of the sixteen MmyB homologs in *S. coelicolor* plus MmyB itself was aligned using Clustal Omega (Figure 3.28). The phylogeny grouped the homologs within three distinct clades.

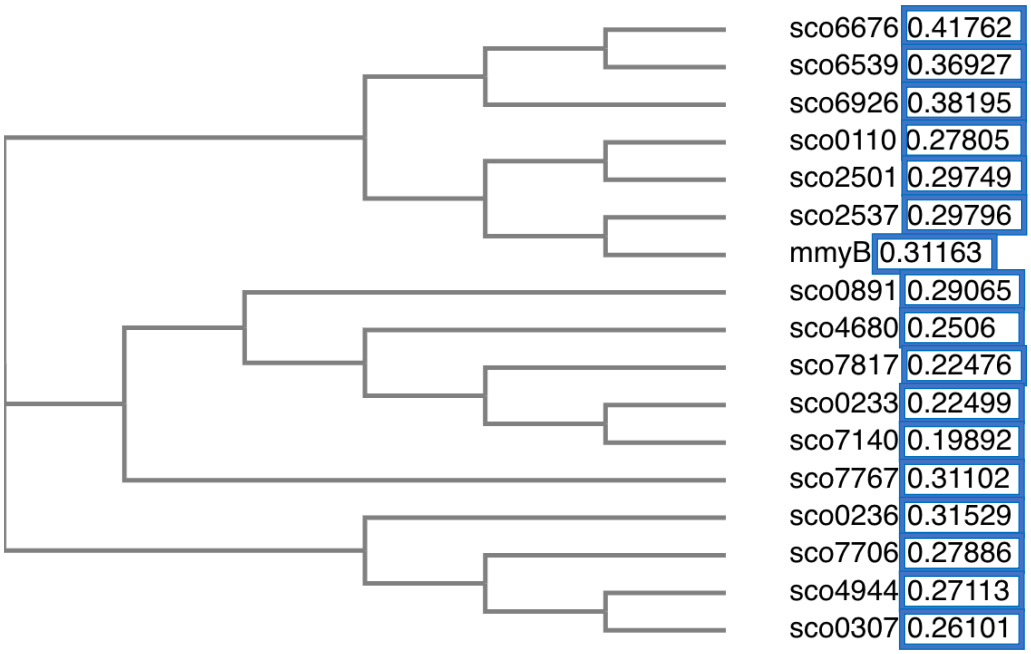


Figure 3.28: Phylogeny tree of the sixteen homologs of MmyB discovered in *S. coelicolor* alongside MmyB. The branch length of each homolog is represented by the boxed number by the gene name which indicates the average number of amino acid substitutions per site.

The only known crystal structure of a MmyB homolog, MltR from *Chloroflexus aurantiacus*, identified two domains within the protein[23], a N-terminal binding doman and a C-terminal ligand binding module. Using the MltR crystal structure as a guide, the first 90 amino acid residues were identified for each homolog and used for multiple sequence alignment by ClusterOmega (Figure 3.29A), while the rest of the residues for each protein were used for the ligand binding module (Figure 3.29B). The phylogenetic distribution of the homologs does vary between the DNA binding domain and the ligand binding domain. However, each tree consistently groups the homologs into three distinct clades.

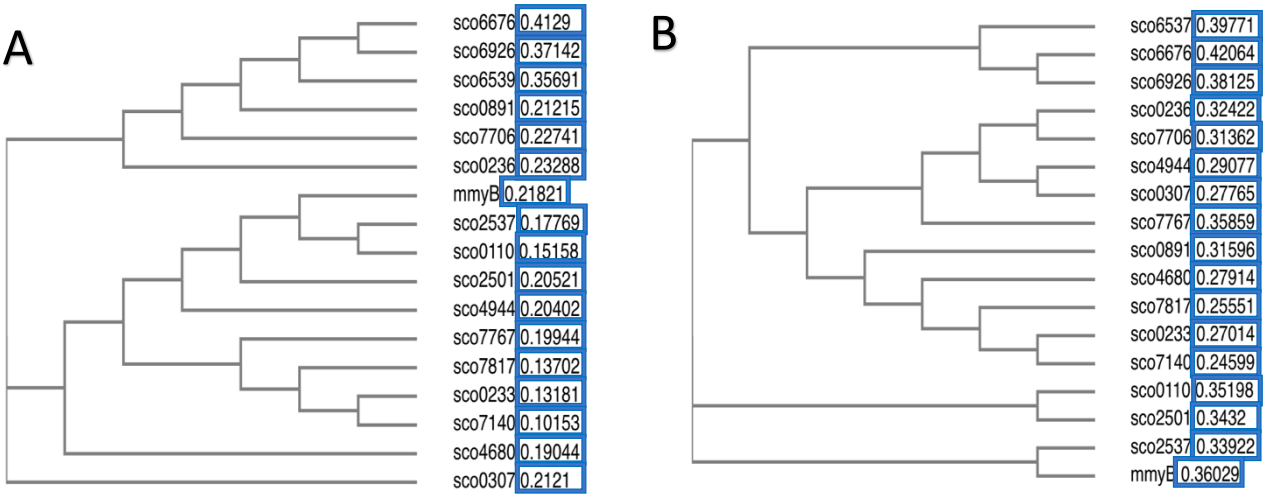


Figure 3.29: Phylogeny tree of the sixteen MmyB homologs two distinct domains plus MmyB. The branch length of each homolog is represented by the boxed number by the gene name which indicates the average number of amino acid substitutions per site.

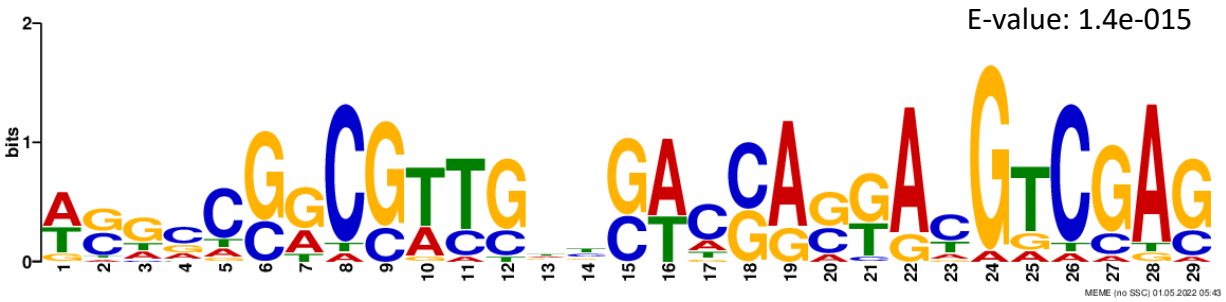
A) N terminal DNA binding domain **B)** C terminal ligand binding module

3.3.2 MEME analysis

In the methylenomycin cluster, *mmyB* is the activator for the cluster and is hypothesised to be able to activate its own expression. The binding sequence for MmyB is still unknown but it was suggested that the *mmyB* promoter contained three B-boxes which were potentially the binding sites for MmyB although this hasn't been confirmed[21]. These B-boxes were checked against the *S. coelicolor* MmyB homologs, looking at if they were present within 500 base pairs upstream of the start codon of each homolog. However, none of the B-boxes were present for any of the homologs.

In order to attempt to identify a shared targeted binding site in DNA for the MmyB-like transcription family, the sixteen *S. coelicolor* MmyB homologs promoter sequences (500 base pairs upstream of the start codon) were submitted for MEME enrichment[63]. One motif of significance was found to be conserved through all sixteen promoter sequences (Figure 3.30).

This motif was submitted to FIMO[64] to identify individual occurrences of the motif in the genome of *S. coelicolor* M145. FIMO identified 68403 hits of the motif within the genome.



S. coelicolor M145 FIMO hits: 68403

Figure 3.30: The motif identified by MEME in the promoter sequences of the sixteen MmyB homologs and the number of hits detected in *S. coelicolor* M145 by FIMO

3.3.3 Motif presence in RNA sequencing

The motif identified by MEME enrichment was compared against the top ten genes with the greatest positive log2 fold change in the RNA sequencing data for both SMM and LB. The promoter sequences (500 base pairs upstream of start codon) of these genes were tested using FIMO against the motif to detect any matches.

CO-121 SMM RNA Sequencing		**CO-121 YEME RNA Sequencing**	
Top 10 positive log2 fold change genes	**FIMO hit number (number discounting overlapping hits)**	**Top 10 positive log2 fold change genes**	**FIMO hit number (number discounting overlapping hits)**
sco4672	11 (5)	*sco0995*	8 (3)
sco4674	3 (2)	*sco1341*	2 (2)
sco4679	0 (0)	*sco1988*	9 (5)
sco4682	10 (6)	*sco2423*	1 (1)
sco4684	5 (2)	*sco2693*	4 (3)

sco4694	6 (3)	*sco3504*	4 (3)
sco4696	8 (5)	*sco3833*	6 (4)
sco4697	3 (1)	*sco4793*	2 (2)
sco5921	2 (1)	*sco6973*	3 (3)
sco7140	15 (6)	*sco7732*	1 (1)

Table 3.10: The number of FIMO hits for the conserved motif against the promoter sequences of the top ten genes in CO-121 with the greatest positive log2 fold change against *S. coelicolor* M145 in SMM and CO-250 in YEME.

In CO-121, of the twenty genes only one does not contain the motif, *sco4679*. The rest contain a minimum of one motif hit and a maximum of fifteen although this does drop to six when overlapping motif hits are discounted (Table 3.10). The maximum number of hits occurs in the *sco7140* promoter which is the gene that an extra copy of was introduced for in CO-121.

CO-124 SMM RNA Sequencing		**CO-124 YEME RNA Sequencing**	
Top 10 positive log2 fold change genes	**FIMO hit number (number discounting overlapping hits)**	**Top 10 positive log2 fold change genes**	**FIMO hit number (number discounting overlapping hits)**
sco0236	4 (3)	*sco0201*	6 (3)
sco1359	4 (3)	*sco2920*	0 (0)
sco1773	4 (3)	*sco3558*	1 (1)
sco2027	2 (1)	*sco3866*	0 (0)
sco4568	7 (4)	*sco4042*	5 (4)
sco4569	6 (3)	*sco4174*	6 (3)
sco4983	1 (1)	*sco4175*	3 (2)
sco5525	5 (2)	*sco4990*	23 (6)
sco7078	13 (3)	*sco5868*	13 (4)
sco7565	25 (6)	*sco5915*	1 (1)

Table 3.11: The number of FIMO hits for the conserved motif against the promoter sequences of the top ten genes in CO-124 with the greatest positive log2 fold change against *S. coelicolor* M145 in SMM and CO-250 in YEME.

In CO-124, there are two genes in the YEME RNA sequencing that do not contain a hit for the motif. FIMO detected a minimum of one for the rest of the genes up to a maximum of twenty-five although this again decreases to six once overlapping hits are discounted (Table 3.11).

CO-154 has the most number of genes that do not contain a motif hit in the top ten most overexpressed in each RNA sequencing, with three. The maximum number of motif hits is lower than in CO-121 and CO-124 with only a maximum of thirteen although it does reach the same level once overlapping hits are discounted (Table 3.12).

CO-154 SMM RNA Sequencing		**CO-154 YEME RNA Sequencing**	
Top 10 positive log2 fold change genes	**FIMO hit number (number discounting overlapping hits)**	**Top 10 positive log2 fold change genes**	**FIMO hit number (number discounting overlapping hits)**
sco4678	9 (4)	*sco1029*	5 (3)
sco4681	0 0)	*sco1571*	5 (3)
sco4682	10 (6)	*sco1744*	5 (4)
sco4683	10 (3)	*sco2920*	0 (0)
sco4685	2 (2)	*sco3082*	10 (2)
sco4686	7 (3)	*sco4042*	5 (3)
sco4687	0 (0)	*sco4682*	10 (5)
sco4693	10 (4)	*sco5868*	13 (4)
sco4694	6 (3)	*sco6093*	11 (4)
sco5291	2 (1)	*sco6926*	2 (2)

Table 3.12: The number of FIMO hits for the conserved motif against the promoter sequences of the top ten genes in CO-154 with the greatest positive log2 fold change against *S. coelicolor* M145 in SMM and CO-250 in YEME.

3.3.4 Motif presence in *Rhodococcus*

In order to investigate if the motif identified from the promoter sequences of the *S. coelicolor* MmyB homologs, the promoters (500 base pairs upstream of start codon) of each of the seventy-six MmyB homologs found within the twenty eight *Rhodococcus* genomes that were previously analysed were run against the motif using FIMO. Out of the seventy-six homologs, only four did not return a hit for the motif in their promoter sequence, with a maximum of fifteen hits being returned for one of the promoter sequences.

Strain	Gene Name	Motif hits in FIMO
R. aetherivorans IcdP1	*AAT18_RS01415*	4
	AAT18_RS21070	8
R. erythropolis BG43	*XU06_RS25955*	1
	XU06_RS28080	9
R. erythropolis CCM2595	*O5Y_RS25605*	3
	O5Y_RS27670	8
R. erythropolis PR4	*RER_RS05870*	3
	RER_RS29060	1
	RER_RS30975	9
pREL1	*RER_RS00200*	1
R. erythropolis R138	*H351_RS04340*	9
	H351_RS06590	1
R. equi 103S	*REQ_RS22620*	6
R. equi DSSKP-R-001	*C7H75_RS23790*	6
R. fascians D188	*A3L23_RS00815*	4
	A3L23_RS02225	2
	A3L23_RS06235	5
	A3L23_RS07750	8
	A3L23_RS14960	8

R. jostii RHA1	*RHA1_RS02730*	3
	RHA1_RS02780	2
	RHA1_RS09515	3
pRHL2	*RHA1_RS40960*	1
R. opacus 1CP	*R1CP_RS19215*	11
	R1CP_RS28630	4
R. opacus B4	*ROP_RS08095*	4
pROB01	*ROP_RS36215*	1
R. opacus PD630	*PD630_RS08950*	9
	PD630_RS21525	9
	PD630_RS21650	2
	PD630_RS28605	10
	PD630_RS34220	8
	PD630_RS34435	6
	PD630_RS34505	13
	PD630_RS34730	2
R. opacus R7	*EP51_01710*	3
	EP51_05660	0
	EP51_26910	2
	EP51_26970	8
	EP51_26995	3
	EP51_27025	3
	EP51_33145	6
pPDG3	*EP51_44330*	1
R. qingshengii djl-6-2	*C1M55_RS27190*	1
	C1M55_RS29565	13
R. rhodochrous NCTC10210	*CKW34_RS06975*	7
R. ruber YYL	*CSW53_RS17445*	9

R. ruber P14	*CS378_RS10510*	10
R. sp. 008	*AOT96_RS07160*	1
	AOT96_RS09230	13
R. sp. 2G	*BO226_RS18795*	8
R. sp. B7740	*NY08_RS02975*	9
	NY08_RS05860	5
	NY08_RS08690	3
	NY08_RS10435	0
	NY08_RS11105	5
	NY08_RS17795	2
	NY08_RS23685	9
R. sp. BH4	*A0W34_RS25975*	0
	A0W34_RS28075	13
R. sp. H-CA8f	*CPI83_RS01265*	9
	CPI83_RS03325	1
R. sp. MTM3W5.2	*BTZ20_RS17675*	9
R. sp. p52	*IM25_RS05365*	15
R. sp. PBTS1	*A3Q40_RS14345*	3
R. sp. PBTS2	*A3Q41_RS01460*	8
	A3Q41_RS07355	1
	A3Q41_RS08810	8
	A3Q41_RS10390	1
	A3Q41_RS14475	2
	A3Q41_RS16035	4
	A3Q41_RS22790	2
R. sp. WB1	*A4U64_RS03520*	7
	A4U64_RS23355	4
R. sp. WMMA185	*BFN03_RS05680*	0

Table 3.13: Number of hits for the enriched motif against the promoter sequences of *Rhodococcus* MmyB homologs.

4. Discussion

4.1 *Streptomyces* investigations

The introduction of extra copies of MmyB homologs was not particularly successful in causing the deregulation of any BGC. However, there are multiple reasons as to why this may have occurred. The overexpression of the homologs was successful as seen in Figure 3.1 and Figure 3.11 with the introduced gene being overexpressed in each mutant. This was as expected with the constitutive promoter allowing the positive feedback loop that is currently hypothesised as to how MmyB control works to begin. This could seem to indicate that the homologs could directly influence the expression of other homologs so the over expression of other homologs in the mutants was not a surprise. The most likely genes to be regulated by MmyB homologs are those that are directly downstream (Tables 3.2 and 3.7). Almost half of these are reductase enzymes that are members of the short chain dehydrogenase/reductase family (SDR). These SDR enzymes are quite commonly found in primary metabolism so this could be an indication that the MmyB homologs may be involved in primary metabolism rather than secondary.

It was interesting to see the overexpression in both sets of media in the supercoiling hypersensitive cluster (Figures 3.2 and 3.12). The fact that it was conserved in both medias would seem to indicate that the overexpression of these homologs is leading to some sort of extra stress being put on the cells causing supercoiling of the DNA that is switching on the expression of this cluster. It's possible that if these homologs are involved in primary metabolism that all the extra transcription being forced by the abundance of the MmyB homolog is creating issues for the DNA.

The aim going into the experiments was to study if the three MmyB homologs selected were involved in any secondary metabolism. Even though many of the twenty-seven BGCs in *S. coelicolor* A3(2) have been characterised (Figure 1.2), there are still some cryptic BGCs. Many genes in the predicted gene clusters showed differential expression (Figure A3-A42), however the majority of these were not predicted to be involved in the biosynthesis. This is likely the result of transcriptional regulators that are within the clusters being affected in some way by

the MmyB homolog overexpression and having a downstream effect on the genes within the cluster. Even the BGCs that contained MmyB homologs such as the SapB BGC (Figure 3.8 and 3.16) did not have consistent differential expression that could be used to make predictions. An interesting observation was how the antiSMASH prediction for certain characterised clusters overestimated the limits of the cluster compared to what had been experimentally validated and recorded in databases such as MIBIG[73] which antiSMASH uses. This is an indication that although prediction tools used in natural product discovery have come a long way and antiSMASH is a great tool, especially for novel genomes, there are still certain limitations to the technology.

One of the most interesting BGCs in these experiments was the actinorhodin BGC due to the fact that all three mutants in SMM have a conserved area of genes that have all been under expressed (Figure 3.5). This would seem to indicate that either the integration of the pOSV556 vector has disrupted the regulation of the cluster or that the core homology of MmyB is able to recognise and down regulate the core biosynthetic genes within the cluster. Although there's twenty-five genes under expressed between the three mutants it is unlikely that any are under the direct influence of one of the MmyB homologs. It is much more probable that it is the *sco5085* gene that is having the influence on the cluster[47]. The *sco5085* gene, also known as actII-4, has been characterised as the actinorhodin cluster activator protein, similar to how *mmyB* activates the methylenomycin cluster by itself. Therefore it isn't a stretch to hypothesise that the under expression of the genes involved in actinorhodin production is due to the under expression of its activator gene by the mutant bacteria whether due to direct control by the homologs but much more likely to be a downstream effect of the overexpression. However, the overexpression in the actinorhodin BGC seen in the YEME media sequencing (Figure 3.15) indicates that the MmyB homologs are unlikely to be the cause of under expression of actII-4 and may have the opposite effect on the cluster. It could be that the stress from growing in minimal media led to the cells prioritising transcription away from actinorhodin and that the levels seen in the minimal media may have been even lower if not for the MmyB homolog overexpression.

Unfortunately not too many inferences can be made on the SMM RNA sequencing overall. The data does show that they are some similarities with many of the genes affected, occurring in multiple of the mutants. This however could be due to the integration of the pOSV556 plasmid into the genome affecting certain transcriptional regulators around the integration site leading to any number of untold downstream effects. It is impossible to discern if the integration of the pOSV556 plasmid has disrupted some regulatory genes and how much of an effect that may have elsewhere in the genome. Therefore trying to identify the differentially expressed genes specific to the MmyB homologs is difficult. There is also the potential that the minimal media that the bacteria was grown in is minimising the transcription changes in biosynthetic gene clusters from the homolog overexpression. The stress that this media exerts onto the bacteria encourages the activation of secondary metabolism due to competition faced for the lack of nutrients. This priming of the bacteria means that any clusters that may be influenced by any of the homologs could already have an increased expression in the wildtype, so the overexpression doesn't have the same impact. On top of this because the bacteria is already beginning the process of switching on secondary metabolism the building blocks for these clusters are in high demand. One of the reasons cryptic clusters haven't been characterised yet is that in natural environments the bacteria are prioritising the already discovered compounds as these are the most reliable. These cryptic clusters could be very niche molecules in their targeting and the bacteria wants to reduce the biosynthetic stress while ensuring maximum efficiency. Therefore the broad targets compounds are preferable and what the bacteria will funnel resources towards. Nevertheless, the fact that there is a number of differences between the three mutants indicates that the overexpression of the MmyB homologs is having an effect on gene expression. Unfortunately it doesn't seem that the alteration in expression is centred on the biosynthetic gene clusters expected. Now this could be for a variety of factors. The biggest is that the MmyB homologs that have been overexpressed just may not be involved in the regulation of any of the clusters predicted by antiSMASH. Another scenario could be that certain triggers are not being met for precursors. So although the MmyB homolog could be involved in a cluster's regulation, it may not be the only regulator needed for expression.

The CO-250 mutant with the empty pOSV556 vector should have been the control from the beginning and it is unfortunate that it was not available for the minimal media sequencing. The identification of 202 genes with differential expression in CO-250 compared to M145 clearly shows that the integration of the plasmid does have an effect on the cell's transcription. The switch from minimal media to the nutrient rich YEME had a drastic effect. There is a huge increase in genes with differential expression with the total number in YEME (Table 3.6) being roughly eight times greater than the SMM (Table 3.1). The large amount of genes makes it extremely difficult to determine the genes directly influenced by the homologs as the DNA binding target is unknown. The far reaching effects are so large that they must be interacting with either a wide variety of different types of genes, potentially many regulatory genes or are somehow encouraging a global regulator to increase expression. Using the Abasy Atlas tool it predicts all three of the homologs are associated in some way with a global regulator[61]. The CO-121 homolog *sco7140* is linked with the RNA polymerase sigma factor *sigR*, *sco0236* introduced in CO-124 is connected to the RNA polymerase sigma factor *hrdB* and the CO-154 homolog *sco6926* interacts with the global regulator *argR*. The likely answer to these interactions is that these sigma factors and regulator are part of the regulation of the homologs but it could be that somehow the homologs are actually encouraging these protein's activation and they are the cause of the widespread differential expression. The reduced stress from growing in nutrient rich YEME may also be a factor in allowing a much stronger downstream effect from the MmyB homolog overexpression.

All the changes in gene expression have ultimately been shown to have little effect on the metabolome of the cell. The SMM cultures, which due to the stress on the bacteria are more primed for secondary metabolism, only had three new peaks occur across the three mutants. CO-121 had two peaks detected that were unique when compared to M145 (Figure 3.18). One of these peaks, the formula was unable to be predicted but the second peak was predicted to have a chemical formula of $C_{28}H_{46}N_9O_7$ (Figure 3.19). This formula does not correspond to a known discovered compound and the two clusters in CO-121 that have increased expression of

the core biosynthetic gene (Figure 3.4 and 3.10) are unlikely to be the cause due to expecting much larger compounds so it is more probable that this is an intermediate. CO-124 and CO-154 actually shared the same new peak when compared to M145 (Figure 3.20 and 3.21), with a predicted formula of $C_{36}H_{23}N_2O_3$ (Figure 3.22). This formula is again not one that corresponds to a known compound. As none of the core biosynthetic genes in any of the clusters have over expression in both CO-124 and CO-154, it does not seem that the new peak is part of secondary metabolism but another intermediate. The YEME cultures metabolomics had very little alteration in the profiles. CO-250 was compared to the M145 parent strain and as no changes were detected in the LCMS trace it would seem to indicate that the integration of pOSV556 does not alter the metabolome of the cell (Figure 3.23), reinforcing the RNA sequencing data. CO-121 was compared against CO-250 and one of the replicates showed a large peak but this was not conserved in the other two replicates so it seems likely that this was an anomaly (Figure 3.24). The CO-124 and CO-154 mutants were compared against CO-250 and both showed no changes (Figure 3.25). The lack of change in the metabolome in the YEME cultures would seem to indicate that the stress caused by the minimal media was an important aspect in the peaks being identified in those cultures.

The overexpression of a the MmyB homologs in the mutants CO-121, CO-124 and CO-154 seems to have a widespread affect in gene expression across the genome although the type of media has a significant influence on how many genes are being targeted. This coupled with the linked metabolomics of those same cultures having little to no changes in the metabolic profile, there doesn't seem to be a link for all of this differential expression and the production of secreted metabolites.

4.2 *Rhodococcus* investigations

From the genome mining analysis, the *Rhodococcus* genus looks to be a severely underutilised bacteria for natural product discovery with over 519 clusters predicted across 28 genomes. While there will be a large amount of redundancy of shared clusters between the different

genomes, even if each genome had only one novel cluster that would result in 28 compounds that could have antimicrobial activity to be developed for commercialisation.

As can be seen in Figure 4.1, there is a huge number of MmyB homologs located in the *Rhodococcus* genus that seem to group into three distinct clades. This grouping is likely due to there being different variations in the DNA and ligand binding domains of the proteins. There are some genes that are found in multiple species. There is a group of three genes; *RER_RS29060, XU06_RS25955* and *C1M55_RS27190* found in *R. erythropolis* PR4, *R. erythropolis* BG43 and *R. qingshengii* djl-6-2 respectively. This conservation could mean that this gene is beneficial in some way to the bacteria, potentially involved in secondary metabolism. There are two other pairs that are found in different strains of the same species which is unsurprising and unlikely to be involved in secondary metabolism. There's one final pair found in two species, the *A3L23_RS07750* gene from *R. fascians* D188 and the *A3Q41_RS08810* found in *R. sp.* PBTS2. This again could be an indication that this gene is somehow involved in secondary metabolism.

There are many other sets of genes with very small phylogenetic differences indicated by values lower than 0.05 that likely correspond to genes that are shared throughout the evolutionary tree of bacteria. The four genes found on plasmids in the various bacteria are all grouped together as well in the tree as Rhodococcus are prone to having large linear plasmids that are conserved through the genus, that contain a variety of useful genes in catabolism[25,26].

As can be seen in Table 4.2, there is a huge number of MmyB homologs but only a few are located anywhere close to a antiSMASH cluster prediction. There are ten homologs within five kilobases of a cluster and there's even four that are found within the prediction. The three conserved distances for the homologs would suggest that of the ten, it is probable that there is some redundancy and that there are three distinct clusters. Regardless these clusters are just the tip of what Rhodococcus may be hiding under the surface.

Both of the trio of genes (*RER_29060, XU06_RS25955* and *C1M55_RS27190*) and the pair (*A3L23_RS07750* and *A3Q41_RS08810*) are identified to be very close to a cluster in the former's case and actually present in the cluster for the latter. This would seem to suggest that the clusters they are associated to are conserved between the species and that these genes are somehow involved with the cluster in some way.

The construction of *Rhodococcus* mutants is really the first step in the investigation of this genus. The sixteen homologs were selected for varying reasons. The homologs from *R. erythropolis* PR4 and *R. jostii* RHA1 were chosen as these two strains are the most well studied and having the native gene being overexpressed gives more confidence that it will have some influence compared to other MmyB homologs that may not be able to function in a heterologous host. The *R. opacus* PD630 homologs were chosen due to the prevalence of NRPS clusters in that strain with seventeen different NRPS clusters being predicted. The mutants constructed did not show any changes in their metabolome when compared to the parent strain so it seems that more work is needed and potentially *R. opacus* PD630 needs to be the focus due to amount of NRPS clusters.

While the over expression of the MmyB homologs in CO-121, CO-124 and CO-154 did not result in the identification of a novel BGC compound, the previous evidence of homologs to methylenomycin regulation being a target to identify novel compounds, the prevalence of the 76 MmyB homologs across the *Rhodococcus* genus still make the deregulation of these homologs an attractive area for natural product discovery.

4.3 MmyB homolog bioinformatic analysis

So little is currently understood about the MmyB-like transcriptional family that the investigation of it is quite messy. Without having the knowledge of where the activators may want to target in the DNA, experiments to identify if they are involved in cryptic biosynthetic clusters are quite scattergun in their approach. While this tactic could be successful in bacteria

that have a majority of uncharacterised BGCs, in a species as well studied as *S. coelicolor* in the area of secondary metabolism, more background is required to streamline the point of attack. Bioinformatic analysis is a strong starting point to find commonalities across the family.

The phylogenetic analysis of the *S. coelicolor* M145 MmyB homologs groups the sixteen homologs into three distinct clades (Figure 3.28). The top clade could be a future starting point for experiments as it contains *mmyB, sco0110* and *sco6926* which are all located within BGCs indicating they might be involved in the regulation of secondary metabolism. The phylogenetic distribution when comparing the DNA binding domain and ligand binding module does vary (Figure 3.29). The consistency of the three distinct clades in the tree being present suggests that the maintenance of certain features that may be involved in protein function. The alteration as to how closely related some genes are to one another depending on which domain is being compared might mean that some homologs have a similar DNA binding target but require different ligands to be bound to the protein for function and vice versa. For example, the *sco0307* and *sco4944* are quite closely related in the ligand binding module, indicating they probably have a common ligand they bind to but the DNA binding domains are much further apart. There are some genes that stay linked together in both trees such as *mmyB* and *sco2537,* which are closely related in both domains. This could indicate that *sco2537* although not situated in a cluster in the antiSMASH prediction is involved someway in secondary metabolism.

As the target sequence of MmyB is unknown, there is the potential that the homologs share a consensus sequence or that the different clades of DNA binding domains target different DNA sequences. The MEME enrichment of the promoter sequences of the MmyB homologs identify a motif of twenty-nine bases (Figure 3.30). As this is a similar size to most promoter binding sequences, that is a good indication that this motif is a possible consensus for MmyB homologs. If this was a consensus sequence, the large number of return hits identified by FIMO in the *S. coelicolor* M145 genome could also help explain the large swathe of changes in transcription that was identified by the RNA sequencing.

The top ten most over expressed genes in each of the mutants CO-121, CO-124 and CO-154 compared to the control in the SMM and YEME cultures were identified. Each of their promoter sequences were compared against the identified motif using FIMO to identify matches (Table 3.10 – 3.12). In total there are six genes out of sixty across the three mutants that do not have a positive match for the motif. However, the number of hits for each of the genes varies from as few as one, up to six when overlapping hits are discounted. This multiplicity of the sequence could indicate that multiple copies of the protein are required for activation or even allows multiple different homologs to interact together to promote expression.

The motif was compared against the promoter sequences of the MmyB homologs identified in *Rhodococcus* bacteria (Table 3.13). As the expected mode of regulation is a positive feedback loop for MmyB, it would be expected that if this motif was a consensus sequence for all MmyB homologs that the promoter sequence for all of them would contain at least one instance of this sequence. Out of the seventy-six MmyB homologs in *Rhodococcus* only four did not contain a single match for the motif in their promoter sequence. This conservation of the motif across different bacteria genus gives more confidence to the likelihood that this motif could be involved in MmyB regulation and is a potential point for further research.

5. Summary and Future Work

5.1 Summary

Due to the rise in AMR across the bacterial kingdom, novel antibiotics that aren't modified derivatives of existing drugs are an urgent need. With the vast majority of medically used antibiotics originating from *Streptomyces* perhaps the time has come to move away from the drug libraries and high-throughput screening that pharmaceuticals has gravitated towards and return to trying to exploit bacteria themselves. Advancements in techniques such as next generation sequencing and bioinformatic analysis has made this the perfect time to enjoy a new golden age of discovery centred around natural products.

5.1.1 *Streptomyces*

As have been proven for years, *Streptomyces* is a scientific wonder when it comes to secondary metabolism. The sheer number of drugs, not only antibiotics, that originated from this genus of bacteria is mind-boggling. Even now, with the model species *S. coelicolor* being researched into the ground, it still offers up new potential points of attack to eke out even more novel natural products from its genome.

The MmyB-like transcriptional activators are one such area of research. The prevalence of this family of proteins throughout all different actinomycete bacteria makes it hugely attractive, especially if they are all like *mmyB* itself and involved in the regulation of a biosynthetic gene cluster. Being able to understand how these proteins work, then makes it so much simpler to exploit them not only in *Streptomyces* but in all the bacteria that contain them. To date not much is known other than the general structure. Much is predicated on this and the activity of *mmyB* in regulation of the methylenomycin cluster.

Previous experiments have identified bioactivity triggered by deregulation of one of the MmyB homologs present in *S. coelicolor.* Three of these genes were successfully overexpressed to attempt to determine how they influenced transcription in the bacteria and whether it was

localised to any biosynthetic gene clusters. In both sets of RNA sequencing, all three of the mutants proved that they are somehow influencing not only the gene expression in BGCs such as the actinorhodin cluster through under expression of one of the key regulators of the cluster activator, but also having much wider spread alterations in transcription across the genome. How this is possible is not yet understood but it can be confidently said that gene expression of biosynthetic gene clusters can be influenced by MmyB-like transcriptional activators.

The RNA sequencing was coupled with comparative metabolic profiling that has identified across the three mutants, three novel metabolites, who's production has somehow been triggered by overexpression of a MmyB homolog. It would seem a distinct possibility now that in those few cryptic clusters still left in *S. coelicolor,* the MmyB homologs could be involved in the regulation of one, likely a homolog that can be found within a cluster prediction. This should be one of the next steps in attempting to exploit these genes alongside purifying those three metabolites and determining what cluster they might originate from. This would further inform the next steps as identification of the cluster being influenced would help understand not only its regulation but also the targeting of the MmyB homologs.

5.1.2 *Rhodococcus*

The *Rhodococcus* genus is one with great potential to be a source of novel natural products. The abundance of NRPS type clusters across the genus makes it an extremely attractive prospect considering how many pharmaceuticals originated in that same sort of cluster. That is the both an advantage and downside of a bacteria that hasn't previously been particularly focused in regards to secondary metabolism. The prospect of any secondary metabolism being something new is so high but the knowledge as to how to activate these clusters is lacking.

Perhaps the answer, at least for some of the clusters is the MmyB-like transcriptional activators. Almost eighty were identified across the genus and all could switch on a gene cluster. Obviously lots is still not understood about these genes so the phylogenetic analysis of

those present in *Rhodococcus* is a great indicator for those that may be involved in secondary metabolism. Coupling this with genome mining allows the identification of genes that are conserved between species and if this is due to their likely involvement with a cluster. The bacteria naturally want to keep beneficial genes and those a part of secondary metabolism are definitely that. This can prevent redundancy in experiments, ensuring that the same gene cluster is being accidently targeted in multiple species.

Two of the MmyB-like activators were successfully integrated into both *R. jostii* RHA1 and *R. erythropolis* PR4. Now only one of these is an extra copy of a gene found in either strain but it's an interesting experiment to investigate if the homologs are able to interact in other species. Nearly all the MmyB-like transcriptional activators before this analysis were identified in *S. coelicolor,* obviously being the bacteria that *mmyB* itself was found in, this isn't surprising. Now however, future experiments could be done using *Streptomyces* MmyB-like activators in *Rhodococcus* and vice versa.

The four mutants total had their metabolomic profile compared to their wildtype to investigate if the overexpression was influencing secondary metabolism in any way. There were no distinct changes between the mutants and wildtype bacteria but this could have been for a variety of reasons, most probable of which is the nutrient rich media. This metabolic comparison needs to be repeated using minimal media to stress the bacteria enough to encourage secondary metabolism. Ideally more mutants with a variety of overexpressed MmyB-like activators can be tested at the same time in the search of a novel natural product to develop onwards.

5.1.3 Bioinformatic investigation

The regulation of MmyB-like transcriptional activators isn't particular well understood. The hypothesis is that similar to *mmyB,* each gene is able to undergo autoregulation. This is obviously dependent on ligand availability as the protein structure contains a module for such a function. MEME enrichment of the *S. coelicolor* MmyB-like activators identified a conserved

sequence of 29 nucleotides present in the promoters of the genes. This could be the site that is targeted by the genes, allowing their autoregulation. These motif was further run against the *S. coelicolor* M145 genome as well as the promoter sequences of the *Rhodococcus* MmyB-like activators and identified a large number of hits throughout the genome and was found in nearly all of the *Rhodococcus* genes. This could go some way in explaining the results produced in the RNA sequencing data in the *S. coelicolor* mutants. This motif can be investigated further through mutagenesis experiments altering nucleotides and identifying how expression of the genes are altered. Probably the best experiment to ascertain the binding sites of the MmyB-like activators is a ChIP-seq analysis of the *S. coelicolor* genes.

6. References

1. Murray, C. J. L., Ikuta, K. S., Sharara, F., Swetzchinski, L., Aguilar, G. R., Gray, A., Han, C., Bisignano, C., Rao, P., Wool, E., Johnson, S. C., Browne, A. J., Chipeta, M. G., Fell, F., Hackett, S., Haines-Woodhouse, G., Hamadani, B. H. K., Kumaran, E. A. P., McManigal, B., Agarwal, R., Akech, S., Albertson, S., Amuasi, J., Andrews, J., Aravkin, A., Ashley, E., Bailey, F., Baker, S., Basnyat, B., Bekker, A., Bender, R., Bethou, A., Bielicki, J., Boonkasidecha, S., Bukosia, J., Carvalheiro, C., Castaneda-Orjuela, C., Chansamouth, V., Chaurasia, S., Chiurchiu, S., Chowdhury, F., Cook, A. J., Cooper, B., Cressey, T. R., Criollo-Mora, E., Cunningham, M., Darboe, S., Day, N. P. J., De Luca, M., Dokova, K., Dramowski, A., Dunachie, S. J., Eckmanns, T., Eibach, D., Emami, A., Feasey, N., Fisher-Pearson, N., Forrest, K., Garrett, D., Gastmeier, P., Giref, A. Z., Greer, R. C., Gupta, V., Haller, S., Haselbeck, A., Hay, S. I., Holm, M., Hopkins, S., Iregbu, K. C., Jacobs, J., Jarovsky, D., Javanmardi, F., Khorana, M., Kissoon, N., Kobeissi, E., Kostyanev, T., Krapp, F., Krumkamp, R., Kumar, A., Kyu, H. H., Lim, C., Limmathurotsakul, D., Loftus, M. J., Lunn, M., Ma, J., Mturi, N., Munera-Huertas, T., Musicha, P., Mussi-Pinhata, M. M., Nakamura, T., Nanavati, R., Nangia, S., Newton, P., Ngoun, C., Novotney, A., Nwakanma, D., Obiero, C. W., Olivas-Martinez, A., Olliaro, P., Ooko, E., Ortiz-Brizuela, E., Peleg, A. Y., Perrone, C., Plakkal, N., Ponce-de-Leon, A., Raad, M., Ramdin, T., Riddell, A., Roberts, T., Robotham, J. V., Roca, A., Rudd, K. E., Russell, N., Schnall, J., Scott, J. A. G., Shivamallappa, M., Sifuentes-Osornio, J., Steenkeste, N., Stewardson, A. J., Stoeva, T., Tasak, N., Thaiprakong, A., Thwaites, G., Turner, C., Turner, P., van Doorn, H. R., Velaphi, S., Vongpradith, A., Vu, H., Walsh, T., Waner, S., Wangrangsimakul, T., Wozniak, T., Zheng, P., Sartorius, B., Lopez, A. D., Stergachis, A., Moore, C., Dolecek, C., Naghavi, M. 2022. Global burden of bacterial antimicrobial resistance in 2019: a systemic analysis. *The Lancet*. 399 (10325): 629-655

2. Davies, J. & Davies, D. 2010. Origins and Evolution of Antibiotic Resistance. *Microbiology and Molecular Biology Reviews.* 74 (3): 417-433

3. Fleming, A. 1929. On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to their Use in the Isolation of B. influenzae. *The British Journal of Experimental Pathology*. 10 (3): 226-236

4. Shen, B. 2015. A New Golden Age of Natural Products Drug Discovery. *Cell.* 163 (6): 1297-1300

5. Ribiero da Cunha, B., Fonesca, L. P. & Calado, C. R. C. 2019. Antibiotic Discovery: Where Have We Come from, Where Do We Go? *Antibiotics (Basel).* 8 (2): 45

6. Rodriguez, E. & McDaniel, R. 2001. Combinatorial biosynthesis of antimicrobials and other natural products. *Current Opinion in Microbiology.* 4 (5): 526-534

7. Walsh, C. T. & Fischbach, M. A. 2010. Natural Products Version 2.0: Connecting Genes to Molecules. *Journal of the American Chemical Society.* 132 (8): 2469-2493
8. Fischbach, M. A. & Walsh, C. T. 2006. Assembly-Line Enzymology for Polyketide and Nonribosomal Peptide Antibiotics: Logic, Machinery and Mechanisms. *Chemical Reviews.* 106 (8): 3468-3496

9. Baltz, R. 2021. Genome mining for drug discovery: progress at the front end. *Journal of Industrial Microbiology and Biotechnology.* 48 (9-10)

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

11. Hutchings, M. I., Truman, A. W. & Wilkinson, B. 2019. Antibiotics: past, present and future. *Current Opinion in Microbiology.* 51: 72-80

12. Xu, M. & Wright, G. D. 2019. Heterologous expression-facilitated natural product discovery in actinomycetes. *Journal of Industrial Microbiology and Biotechnology.* 46: 415-431

13. Aigle, B. & Corre, C. 2012. Waking up Streptomyces secondary metabolism by constitutive expression of activators or genetic disruption of repressors. *Methods in Enzymology.* 517: 343-366

14. Tong, Y., Weber, T. & Lee, S. Y. 2019. CRISPR/Cas-based genome engineering in natural product discovery. *Natural Product Reports*. 36 (9): 1262-1280

15. Martens, E. & Demain, A. L. 2017. The antibiotic resistance crisis, with a focus on the United States. The *Journal of Antibiotics.* 70 (5): 520-526

16. Flardh, K. & Buttner, M. J. 2009. Streptomyces morphogenetics: dissecting differentiation in a filamentous bacterium. *Nature Reviews Microbiology.* 7 (1): 36-49

17. Procopio, R. E., Silva, I. R., Martins, M. K., Azevedo, J. L. & Araujo, J. M. 2012. Antibiotics produced by Streptomyces. *The Brazilian Journal of Infectious Diseases.* 16 (5): 466-471

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

19. Blin, K., Shaw, S., Kloosterman, A. M., Charlop-Powers, Z., van Weezel, G. P., Medema, M. H. & Weber, T. 2021. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Research*. 49 (W1): W29-W35

20. Hayes, A., Hobbs, G., Smith, C. P., Oliver, S. G. & Butler, P. R. 1997. Environmental signals triggering methylenomycin production by Streptomyces coelicolor A3(2). *Journal of Bacteriology.* 179 (17): 5511-5515

21. O'Rourke, S., Wietzorrek, A., Fowler, K., Corre, C., Challis, G. L. & Chater, K. F. 2009. Extracellular signalling, translational control, two repressors and an activator all contribute to the regulation of methylenomycin production in Streptomyces coelicolor. *Molecular Microbiology.* 71 (3): 763-778

22. Adeli, J. 2018. *Unpublished PhD thesis.* University of Warwick.

23. Xu, Q., van Wezel, G. P., Chiu, H.-J., Jaroszewski, L., Klock, H. E., Knuth, M. W., Miller, M. D., Lesley, S. A., Godzik, A., Elsliger, M.-A., Deacon, A. M. & Wilson, I. A. 2012. Structure of an MmyB-Like Regulator from *C. aurantiacus*, Member of a New Transcription Factor Family Linked to Antibiotic Metabolism in Actinomycetes. *PLoS One.* 7 (7): e41359

24. Sidda, J. D., Song, L., Poon, V., Al-Bassam, M., Lazos, O., Buttner, M. J., Challis, G. L. & Corre, C. 2014. Discovery of a family of γ-aminobutyrate ureas via rational derepression of a silent bacterial gene cluster. *Chemical Science*. 5: 86-89

25. van der Geize, R. & Dijkhuizen, L. 2004. Harnessing the catabolic diversity of Rhodococci for environmental and biotechnological applications. *Current Opinion in Microbiology*. 7 (3): 255-261

26. Larkin, M. J., Kulakov, L. A. & Allen, C. C. R. 2005. Biodegradation and Rhodococcus – masters of catabolic versatility. *Current Opinion in Biotechnology*. 16 (3): 282-290

27. Yoshimoto, T., Nagai, F., Fujimoto, J., Watanabe, K., Mizukoshi, H., Makino, T., Kimura, K., Saino, H., Sawada, H. & Omura, H. 2004. Degradation of estrogens by Rhodococcus zopfii and Rhodococcus equi isolates from activated sludge in wastewater treatment plants. *Applied and Environmental Microbiology.* 70 (9): 5283-5289

28. Pi, Y., Chen, B., Bao, M., Fan, F., Cai, Q., Ze, L. & Zhang, B. 2017. Microbial degradation of four crude oil by biosurfactant producing strain *Rhodococcus* sp. *Bioresource Technology*. 232: 263-269

29. Zampolli, J., Zeaiter, Z., Di Canito, A. & Di Gennaro, P. 2019. Genome analysis and -omics approaches provide insights into the biodegradation potential of *Rhodococcus*. *Applied Microbiology and Biotechnology*. 103: 1069-1080

30. Chiba, H., Agematu, H., Kaneto, R., Terasawa, T., Sakai, K., Dobashi, K. & Yoshioka, T. 1999. Rhodopeptins (Mer-N1033), Novel Cyclic Tetrapeptides with Antifungal Activity from *Rhodococcus* sp. *The Journal of Antibiotics.* 52 (8): 695-699

31. Iwatsuki, M., Uchida, R., Takakusagi, Y., Matsumoto, A., Jiang, C-L., Takahashi, Y., Arai, M., Kobayashi, S., Matsumoto, M., Inokoshi, J., Tomoda, H. & Omura, S. 2007. Lariatins, Novel Anti-mycobacterial Peptides with a Lasso Structure, Produced by *Rhodococcus jostii* KO1-BO171. *The Journal of Antibiotics*. 60: 357-363

32. Kitagawa, W. & Tamura, T. 2008. Three Types of Antibiotics Produced from *Rhodococcus erythropolis* Strains. *Microbes and Environments.* 23 (2): 167-171

33. Bosello, M., Robbel, L., Linne, U., Xie, X. & Marahiel, M. A. 2011. Biosynthesis of the siderophore rhodochelin requires the coordinated expression of three independent gene clusters in Rhodococcus jostii RHA1. *Journal of the American Chemical Society*. 133 (12): 4587-4595

34. Ward, A. L., Reddyvari, P., Borisova, R., Shilabin, A. G. & Lampson, B. C. 2018. An inhibitory compound produced by a soil isolate of *Rhodococcus* has strong activity against the veterinary pathogen *R. equi*. *PLOS ONE.* 13 (12): e0209275

35. Ceniceros, A., Dijkhuizen, L., Petrusma, M. & Medema, M. H. 2017. Genome-based exploration of the specialized metabolic capacities of the genus Rhodococcus. *BMC Genomics.* 18 (1): 593

36. McLeod, M. P., Warren, R. L., Hsiao, W. W. L., Araki, N., Myhre, M., Fernandes, C., Miyazawa, D., Wong, W., Lillquist, A. L., Wang, D., Dosanjh, M., Hara, H., Petrescu, A., Morin, R. D., Yang, G., Stott, J. M., Schein, J. E., Shin, H., Smailus, D., Siddiqui, A. S., Marra, M. A., Jones, S. J. M., Holt, R., Brinkman, F. S. L., Miyauchi, K., Fukuda, M., Davies, J. E., Mohn, W. W. & Eltis, L. D. 2006. The complete genome of Rhodococcus sp. RHA1 provides insights into a catabolic powerhouse. *Proceedings of the National Academy of Sciences of the United States of America*. 103 (42): 15582-15587

37. Sun, C., Yang, Z., Zhang, C., Liu, Z., He, J., Liu, Q., Zhang, T., Ju, J. & Ma, J. 2019. Genome Mining of Streptomyces atratus SCSIO ZH16: Discovery of Atratumycin and Identification of its Biosynthetic Gene Cluster. *Organic Letters.* 21 (5): 1453-1457

38. Robbel, L., Knappe, T. A., Linne, U., Xie, X. & Marahiel, M. A. 2010. Erythrochelin—a hydroxamate-type siderophore predicted from the genome of Saccharopolyspora erthraea. *The FEBS Journal.* 277 (3): 663-676

39. Szafran, M. J., Gongerowska, M., Malecki, T., Elliot, M. & Jakimowicz, D. 2019. Transcriptional Response of *Streptomyces coelicolor* to Rapid Chromosome Relaxation or Long-Term Supercoiling Imbalance. *Frontiers in Micrbiology*. 10: 1605

40. Becerril, A., Alvarez, S., Brana, A. F., Rico, S., Diaz, M., Santamaria, R. I., Salas, J. A. & Mendez, C. 2018. Uncovering production of specialized metabolites by Streptomyces argillaceus: Activation of cryptic biosynthesis gene clusters using nutritional and genetic approaches. *PLoS One.* 13 (5): e0198145

41. Challis, G. L. & Ravel, J. 2000. Coelichelin, a new peptide siderophore encoded by *Streptomyces coelicolor* genome: structure prediction from the sequence of its non-ribosomal peptide synthetase. *FEMS Microbiology Letters*. 187 (2): 111-114

42. Bursy, J., Kuhlmann, A. U., Pittelkow, M., Hartmann, H., Jebbar, M., Pierik, A. J. & Bremer, E. 2008. Synthesis and uptake of the compatible solutes ectoine and 5-hydroxyectoine by Streptomyces coelicolor A3(2) in response to salt and heat stresses. *Applied Environmental Microbiology*. 74 (23): 7286-7296

43. Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., Kikuchi, H., Shiba, T., Sakaki, Y. & Hattori, M. 2001. Genome sequence of an industrial microorganism Streptomyces avermitilis: deducing the ability of producing secondary metabolites. *Procceedings of the National Academy of Sciences of the United States of America.* 98 (21): 12215-12220

44. Barona-Gomez, F., Wong, U., Giannakopulos, A. E., Derrick, P. J. & Challis, G. L. 2004. Identification of a cluster of genes that directs desferrioxamine biosynthesis in Streptomyces coelicolor M145. *Journal of the American Chemical Society.* 126 (50): 16282-16283

45. Hojati, Z., Milne, C., Harvey, B., Gordon, L., Borg, M., Flett, F., Wilkinson, B., Sidebottom, P. J., Rudd, B. A. M., Hayes, M. A., Smith, C. P. & Micklefield, J. 2002. Structure, biosynthetic origin, and engineered biosynthesis of calcium-dependent antibiotics from Streptomyes coelicolor. *Chemistry & Biology.* 9 (11): 1175-1187

46. Itoh, T., Taguchi, T., Kimberley, M. R., Booker-Milburn, K. I., Stephenson, G. R., Ebizuka, Y. & Ichinose, K. 2007. Actinorhodin biosynthesis: structural requirements for post-PKS tailoring intermediates revealed by functional analysis of ActVI-ORF1 reductase. *Biochemistry*. 46 (27): 8181-8188

47. Uguru, G. C., Stephens, K. E., Stead, J. A., Towle, J. E., Baumberg, S. & McDowall, K. J. 2005. Transcriptional activation of the pathway-specific regulator of the actinorhodin biosynthetic genes in Streptomyces coelicolor. *Molecular Microbiology*. 58 (1): 131-150

48. Zhao, B., Lin, X., Lei, L., Lamb, D. C., Kelly, S. L., Waterman, M. R. & Cane, D. E. 2008. Biosynthesis of the sesquiterpene antibiotic albaflavenone in Streptomyces coelicolor A3(2). *The Journal of Biological Chemistry*. 283 (13): 8183-8189

49. Novakova, R., Bistakova, J. & Kormanec, J. 2004. Characterization of the polyketide sport pigment cluster whiESa in Streptomyces aureofaciens CCM3239. *Archives of Microbiology*. 182 (5): 388-395

50. Bergh, S. & Uhlen, M. 1992. Analysis of a polyketide synthesis-encoding gene cluster of Streptomyces curacoi. *Gene.* 117 (1): 131-136

51. Williamson, N. R., Fineran, P. C., Leeper, F. J. & Salmond, G. P. C. 2006. The biosynthesis and regulation of bacterial prodiginines. *Nature Reviews Microbiology.* 4 (12): 887-899

52. Jiang, J., He, X. & Cane, D. E. 2007. Biosynthesis of the earthy odorant geosmin by a bifunctional Streptomyces coelicolor enzyme. *Nature Chemical Biology.* 3 (11): 711-715

53. Gottelt, M., Kol, S., Gomez-Escribano, J. P., Bibb, M. & Takano, E. 2010. Deletion of a regulatory gene within the cpk gene cluster reveals novel antibacterial activity in Streptomyces coelicolor A3(2). *Microbiology (Reading)*. 158 (8): 2343-2353

54. Torkkell, S., Kunnari, T., Palmu, K., Mantsala, P., Hakala, J. & Ylihonko, K. 2001. The entire nogalamycin biosynthetic gene cluster of Streptomyces nogalater: characterization of a 20-kb DNA region and generation of hybrid structures. *Molecular Genetics and Genomics*. 266 (2): 276-288

55. Kodani, S., Hudson, M. E., Durrant, M. C., Buttner, M. J., Nodwell, J. R. & Willey, J. M. 2004. The SapB morphogen is a lantibiotic-like peptide derived from the product of the developmental gene ramS in Streptomyces coelicolor. *Proceedings of the National Academy of Sciences of the United States of America.* 101 (31): 11448-11453

56. Poralla, K., Muth, G. & Hartner, T. 2000. Hopanoids are formed during transition from substrate to aerial hyphae in *Streptomyces coelicolor. FEMS Microbiology Letters.* 189 (1): 93-95

57. Cruz-Morales, P., Kopp, J. F., Martinez-Guerrero, C., Yanez-Guerra, L. A., Selem-Mojica, N., Ramos-Aboites, H., Feldmann, J. & Barona-Gomez, F. 2016. Phylogenomic Analysis of Natural Products Biosynthetic Gene Clusters Allows Discovery of Arseno-Organic Metabolites in Model Streptomyces. *Genome Biology and Evolution.* 8 (6): 1906-1916

58. Wang, X., Reynolds, A. R., Elshahawi, S. I., Shaaban, K. A., Ponomareva, L V., Saunders, M. A., Elgumati, I. S., Zhang, Y., Copley, G. C., Hower, J. C., Sunkara, M., Morris, A. J., Kharel, M. K., Van Lanen, S. G., Prendergast, M. A. & Thorson, J. S. 2015. Terfestatins B

and C, New p-Terphenyl Glycosides Produced by Streptomyces sp. RM-5-8. *Organic Letters.* 17 (11): 2796-2799

59. Funabashi, M., Funa, N. & Horinouchi, S. 2008. Phenolic lipids synthesized by type III polyketide synthase confer penicillin resistance on Streptomyces griseus. *Journal of Biological Chemistry.* 283 (20): 13983-13991

60. Oliynyk, M., Samborskyy, M., Lester, J. B., Mironenko, T., Scott, N., Dickens, S., Haydock, S. F. & Leadley, P. F. 2007. Complete genome sequence of the erythromycin-producing bacterium Saccharopolyspora erythraea NRRL23338. *Nature Biotechnology*. 25 (4): 447-453

61. Escoria-Rodriguez, J. M., Tauch, A. & Freyre-Gonzalez, J. A. 2020. Abasy Atlas v2.2: The most comprehensive and up-to-date inventory of meta-curated, historical, bacterial regulatory networks, their completeness and system-level characterization

62. Onaka, H., Ando, N., Nihira, T., Yamada, Y., Beppu, T. & Horinouchi, S. 1995. Cloning and characterization of the A-factor receptor gene from Streptomyces griseus. *Journal of Bacteriology*. 177 (21): 6083-6092

63. Bailey, T. L & Elkan, C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology.* pg 28-36

64. Grant, C. E., Bailey, T. L. & Noble, W. S. 2011. FIMO: Scanning for occurrences of a given motif. *Bioinformatics*. 27 (7): 1017-1018

65. Aseev, L. V., Koledinskaya, L. S. & Boni, I. V. 2016. Regulation of Ribosomal Protein Operons *rplM-rplI, rpmB-rpmG,* and *rplU-rpmA* at the Transcriptional and Translational Levels. *Journal of Bacteriology.* 198 (18): 2494-2502

66. Bernier-Villamor, V., Camacho, A., Hidalgo-Zarco, F., Perez, J., Ruiz-Perez, L. M. & Gonzalez-Pacanowska, D. 2002. Characterization of deoxyuridine 5'-triphosphate nucleotidohydrolase from *Trypanosoma cruzi*. *FEBS Letters*. 526 (1-3): 147-150

67. Bruggink, A., Roos, E. C. & de Vroom, E. 1998. Penicillin Acylase in the Industrial Product of ß-Lactam Antibiotics. *Organic Process Research and Development*. 2 (2): 128-133

68. Rico, S., Santamaria, R. I., Yepes, A., Rodriguez, H., Laing, E., Bucca, G., Smith, C. P. & Diaz, M. 2014. Deciphering the Regulon of *Streptomyces coelicolor* AbrC3, a Positive Response Regulator of Antibiotic Production. *Applied Environmental Microbiology*. 80 (8): 2417-2428

69. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. 1990. Basic local alignment search tool. *Journal of Molecular Biology*. 215 (3): 403-410

70. Madeira, F., Pearce, M., Tivey, A. R. N., Basutkar, P., Lee, J., Edbali, O., Madhusoodanan, N., Kolesnikov, A. & Lopez, R. 2022. Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Research*. gkac240

71. Langmead, B. & Salzberg, S. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*. 9: 357-359

72. Love, M. I., Huber, W. & Anders, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*. 15: 550

73. Terlouw, B. R., Blin, K., Navarro-Munoz, J. C., Avalon, N. E., Chevrette, M. G., Egbert, S., Lee, S., Meijer, D., Recchia, M. J. J., Reitz, Z. L., van Santen, J. A., Selem-Mojica, N.,

Torring, T., Zaroubi, L., Alanjary, M., Aleti, G., Aguilar, C., Al-Salihi, S. A. A., Augustijn, H. E., Avelar-Rivas, J. A., Avitia-Dominguez, L. A., Barona-Gomez, F., Bernaldo-Aguero, J., Bielinski, V. A., Biermann, F., Booth, T. J., Carrion Bravo, V. J., Castelo-Branco, R., Chagas, F. O., Cruz-Morales, P., Du, C., Duncan, K. R., Gavriilidou, A., Gayrard, D., Gutierrez-Garcia, K., Haslinger, K., Helfrich, E. N., van der Hooft, J. J. J., Jati, A. P., Kalkreuter, E., Kalyvas, N., Bin Kang, K., Kautsar, S., Kim, W., Kunjapur, A. M., Li, Y.-X., Lin, G.-M., Loureiro, C., Louwen, J. J. R., Louwen, N. L. L., Lund, G., Parra, J., Philmus, B., Pourmohsenin, B., Pronk, L. J. U., Rego, A., Balaya Rex, D. A., Robinson, S., Rosas-Becerra, L. R., Roxborough, E. T., Schorn, M. A., Scobie, D. J., Singh, K. S., Sokolova, N., Tang, X., Udwary, D., Vigneshwari, A., Vind, K., Vromans, S. P. J. M., Waschulin, V., Williams, S. E., Winter, J. M., Witte, T. E., Xie, H., Yang, D., Yu, J., Zdouc, M., Zhong, Z., Collemare, J., Linington, R. G., Weber, T. & Medema, M. H. 2023. MIBIG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters. *Nucleic Acids Research*. 51 (D1): D603-610

74. Wang, M., Carver, J. J., Phelan, V. V., Sanchez, L. M., Garg, N., Peng, Y., Nguyen, D. D., Watrous, J., Kapono, C. A., Luzzatto-Knaan, T., Porto, C., Bouslimani, A., Melnik, A. V., Meehan, M. J., Liu, W.-T., Crusemann, M., Boudreau, P. D., Esquenazi, E., Sandoval-Calderon, M., Kersten, R. D., Pace, L. A., Quinn, R. A., Duncan, K. R., Hsu, C.-C., Floros, D. J., Gavilan, R. G., Kleigrewe, K., Northen, T., Dutton, R. J., Parrot, D., Carlson, E. E., Aigle, B., Michelsen, C. F., Jelsbak, L., Sohlenkamp, C., Pevzner, P., Edlund, A., McLean, J., Piel, J., Murphy, B. T., Gerwick, L., Liaw, C.-C., Yang, Y.-L., Humpf, H.-U., Maansson, M., Keyzers, R. A., Sims, A. C., Johnson, A. R., Sidebottom, A. M., Sedio, B. E., Klitgaard, A., Larson, C. B., Boya, C. A., Torres-Mendoza, D., Gonzalez, D. J., Silva, D. B., Marques, L. M., Demarque, D. P., Pociute, E., O'Neill, E. C., Briand, E., Helfrich, E. J. N., Granatosky, E. A., Glukhov, E., Ryffel, F., Houson, H., Mohimani, H., Kharbush, J. J., Zeng, Y., Vorholt, J. A., Kurita, K. L., Charusanti, P., McPhail, K. L., Nielsen, K. F., Vuong, L., Elfeki, M., Traxler, M. F., Engene, N., Koyama, N., Vining, O. B., Baric, R., Silva, R. R., Mascuch, S. J., Tomasi, S., Jenkins, S., Macherla, V., Hoffman, T., Agarwal, V., Williams, P. G., Dai, J., Neupane,

R., Gurr, J., Rodriguez, A. M. C., Lamsa, A., Zhang, C., Dorrestein, K., Duggan, B. M., Almaliti, J., Allard, P.-M., Phapale, P., Nothias, L.-F., Alexandrov, T., Litaudon, M., Wolfender, J.-L., Kyle, J. E., Metz, T. O., Peryea, T., Nguyen, D.-T., VanLeer, D., Shinn, P., Jadhav, A., Muller, R., Waters, K. M., Shi, W., Liu, X., Zhang, L., Kinght, R., Jensen, P. R., Palsson, B. O., Pogliano, K., Linington, R. G., Gutierrez, M., Lopes, N. P., Gerwick, W. H., Moore, B. S., Dorrestein, P. C. & Bandeira, N. 2016. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nature Biotechnology*. 34 (8): 828-837

75. Hofer, U. 2019. The cost of antimicrobial resistance. *Nature Reviews Microbiology*. 17 (3)

76. Holmes, A. H., Moore, L. S. P., Sundsfjord, A., Steinbakk, M., Regmi, S., Karkey, A., Guerin, P. J. & Piddock, L. J. V. 2016. Understanding the mechanisms and drivers of antimicrobial resistance. *The Lancet*. 387 (10014): 176-187

77. Prestinaci, F., Pezzotti, P. & Pantosti, A. 2015. Antimicrobial resistance: a global multifaceted phenomenon. *Pathogens and Global Health*. 109 (7): 309-318

78. Khan, I., Ibrar, A., Abbas, N. & Saeed, A. 2014. Recent advances in the structural library of functionalized quinazoline and quinazoline scaffolds: Synthetic approaches and multifarious applications. *European Journal of Medicinal Chemistry*. 76: 193-244

79. Gerry, C. J. & Schreiber, S. L. 2018. Chemical probes and drug leads from advances in synthetic planning and methodology. *Nature Reviews Drug Discovery*. 17: 333-352

80. Barnes, E. C., Kumar, R. & Davis, R. A. 2016. The use of isolated natural products as scaffolds for the generation of chemically diverse screening libraries for drug discovery. *Natural Product Reports*. 33: 372-381

81. Genilloud, O. 2017. Actinomycetes: still a source of novel antibiotics. *Natural Product Reports*. 34: 1203-1232

82. Atanasov, A. G., Zotchev, S. B., Dirsch, V. M., International Natural Product Sciences Taskforce & Supuran, C. T. 2021. Natural products in drug discovery: advances and opportunities. *Nature Reviews Drug Discovery*. 20 (3): 200-216

83. Albarano, L., Esposito, R., Ruocco, N. & Costantini, M. 2020. Genome Mining as New Challenge in Natural Products Discovery. *Marine Drugs*. 18 (4): 199

84. Challis, G. L. 2008. Genome mining for novel natural product discovery. *Journal of Medicinal Chemistry*. 51 (9): 2618-2628

85. Cimermancic, P., Medema, M. H., Claesen, J., Kurita, K., Wieland Brown, L. C., Mavrommatis, K., Pati, A., Godfrey, P. A., Koehrsen, M., Clardy, J., Birren, B. W., Takano, E., Sali, A., Linington, R. G. & Fischbach, M. A. 2014. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell*. 158 (2): 412-421

86. Medema, M. H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J. B., Blin, K., de Bruijn, I., Choi, Y. H., Claesen, J., Coates, R. C., Cruz-Morales, P., Duddela, S., Dusterhus, S., Edwards, D. J., Fewer, D. P., Garg, N., Geiger, C., Gomez-Escribano, J. P., Greule, A., Hadjithomas, M., Haines, A. S., Helfrich, E. J. N., Hillwig, M. L., Ishida, K., Jones, A. C., Jones, C. S., Jungmann, K., Kegler, C., Kim, H. U., Kotter, P., Krug, D., Masschelein, J., Melnik, A. V., Mantovani, S. M., Monroe, E. A., Moore, M., Moss, N., Nutzmann, H.-W., Pan, G., Pati, A., Petras, D., Reen, F. J., Rosconi, F., Rui, Z., Tian, Z., Tobias, N. J., Tsunematsu, Y., Wiemann, P., Wyckoff, E., Yan, X., Yim, G., Yu, F., Xie, Y., Aigle, B., Apel, A. K., Balibar, C. J., Balskus, E. P., Barona-Gomez, F., Bechthold, A., Bode, H. B., Borriss, R., Brady, S. F., Brakhage, A. A., Caffrey, P., Cheng, Y.-Q., Clardy, J., Cox, R. J., De Mot, R., Donadio, S., Donia, M. S., van der Donk, W. A., Dorrestein, P. C., Doyle, S., Driessen, A. J. M., Ehling-Schulz, M., Entian, K.-D., Fischbach, M. A., Gerwick, L., Gerwick, W. H., Gross, H., Gust, B., Hertweck, C., Hofte, M., Jensen, S. E., Ju, J., Katz, L., Kaysser, L., Klassen, J. L., Keller, N. P., Kormanec, J., Kuipers, O. P., Kuzuyama, T., Kyrpides, N. C., Kwon, H.-J., Lautru, S., Lavigne, R., Lee, C. Y., Linquan, B., Liu, X., Liu, W., Luzhetskyy, A., Mahmud, T.,

Mast, Y., Mendez, C., Metsa-Ketela, M., Micklefield, J., Mitchell, D. A., Moore, B. S., Moreira, L. M., Muller, R., Neilan, B. A., Nett, M., Nielsen, J., O'Gara, F., Oikawa, H., Osbourn, A., Osbourne, M. S., Ostash, B., Payne, S. M., Pernodet, J.-L., Petricek, M., Piel, J., Ploux, O., Raaijmakers, J. M., Salas, J. A., Schmitt, E. K., Scott, B., Seipke, R. F., Shen, B., Sherman, D. H., Sivonen, K., Smanski, M. J., Sosio, M., Stegmann, E., Sussmuth, R. D., Tahlan, K., Thomas, C. M., Tang, Y., Truman, A. W., Viaud, M., Walton, J. D., Walsh, C. T., Weber, T., van Wezel, G. P., Wilkinson, B., Willey, J., Wohlleben, W., Wright, G. D., Ziemert, N., Zhang, C., Zotchev, S. B., Breitling, R., Takano, E. & Glocker, F. O. 2015. Minimum Information about a Biosynthetic Gene cluster. *Nature Chemical Biology*. 11: 625-631

87. Kwon, M. J., Steiniger, C., Cairns, T. C., Wisecaver, J. H., Lind, A. L., Pohl, C., Regner, C., Rokas, A. & Meyer, V. 2021. Beyond the Biosynthetic Gene Cluster Paradigm: Genome-Wide Coexpression Networks Connect Clustered and Unclustered Transcription Factors to Secondary Metabolic Pathways. *Microbiology Spectrum*. 9 (2): e00898-21

88. Caesar, L. K., Montaser, R., Keller, N. P. & Kelleher, N. L. 2021. Metabolomics and genomics in natural products research: complementary tools for targeting new chemical entities. *Natural Product Reports*. 38: 2041-2065

89. Wu, C., Kim, H. Y., van Wezel, G. P. & Choi, Y. H. 2015. Metabolomics in the natural products field – a gateway to novel antibiotics. *Drug Discovery Today: Technologies.* 13: 11-17

90. Amos, G. C. A., Awakawa, T., Tuttle, R. N. & Jensen, P. R. 2017. Comparative transcriptomics as a guide to natural product discovery and biosynthetic gene cluster functionality.

91. Gao, W., Sun H.-X., Xiao, H., Cui, G., Hillwig, M. L., Jackson, A., Wang, X., Shen, Y., Zhao, N., Zhang, L., Wang, X.-J., Peters, R. J. & Huang, L. 2014. Combining metabolomics and

transcriptomics to characcterize tanshinone biosynthesis in *Salvia miltiorrhiza. BMC Genomics*. 15: 73

92. Selim, M. S. M., Abdulhamid, S. A. & Mohamed, S. S. 2021. Secondary metabolites and biodiversity of actinomycetes. *Journal of Genetic Engineering and Biotechnology*. 19: 72

7. Appendix

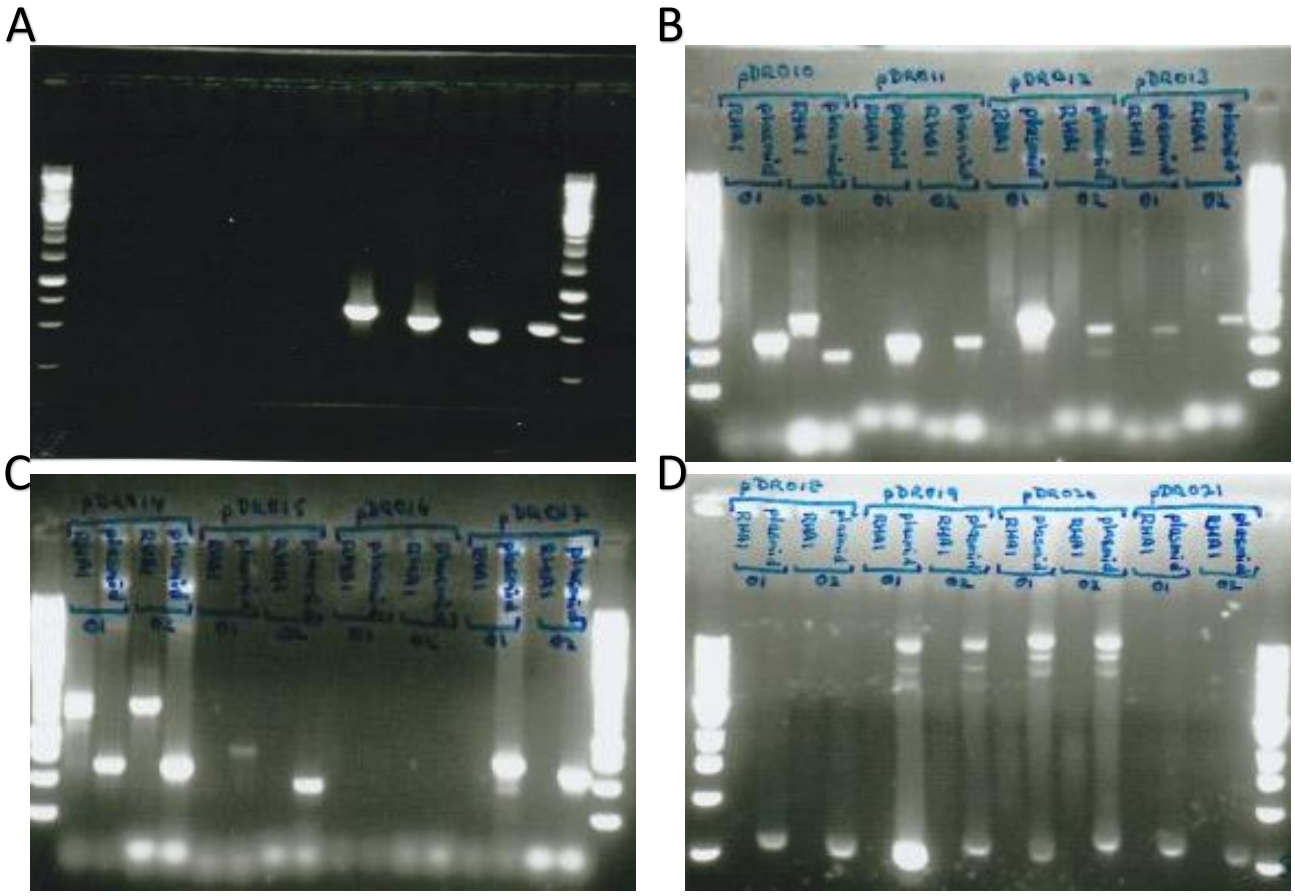


Figure A1: PCR testing integration of plasmids into *R. jostii* RHA1. Even lanes are the genomic DNA of potential mutants, odd lanes are the plasmid positive control for all gels.
A) Lanes correspond to **1.** Generuler 1kb ladder **2-3.** pDR005 01+pOSV02 **4.5.** pDR005 02+pOSV01 **6-7.** pDR007 01+pOSV02 **8-9.** pDR007 02+pOSV01 **10-11.** pDR008 01+pOSV02 **11-12.** pDR008 02+pOSV01 **13-14.** pDR009 01+pOSV02 **15-16.** pDR009 02+pOSV01 **17.** Generuler 1kb ladder
B) Lanes correspond to **1.** Generuler 1kb ladder **2-3.** pDR0010 01+pOSV02 **4.5.** pDR0010 02+pOSV01 **6-7.** pDR011 01+pOSV02 **8-9.** pDR011 02+pOSV01 **10-11.** pDR012 01+pOSV02 **11-12.** pDR012 02+pOSV01 **13-14.** pDR013 01+pOSV02 **15-16.** pDR013 02+pOSV01 **17.** Generuler 1kb ladder
C) Lanes correspond to **1.** Generuler 1kb ladder **2-3.** pDR014 01+pOSV02 **4.5.** pDR014 02+pOSV01 **6-7.** pDR0015 01+pOSV02 **8-9.** pDR015 02+pOSV01 **10-11.** pDR016 01+pOSV02 **11-12.** pDR016 02+pOSV01 **13-14.** pDR017 01+pOSV02 **15-16.** pDR017 02+pOSV01 **17.** Generuler 1kb ladder
D) Lanes correspond to **1.** Generuler 1kb ladder **2-3.** pDR018 01+pOSV02 **4.5.** pDR018 02+pOSV01 **6-7.** pDR019 01+pOSV02 **8-9.** pDR019 02+pOSV01 **10-11.** pDR020 01+pOSV02 **11-12.** pDR020 02+pOSV01 **13-14.** pDR021 01+pOSV02 **15-16.** pDR021 02+pOSV01 **17.** Generuler 1kb ladder

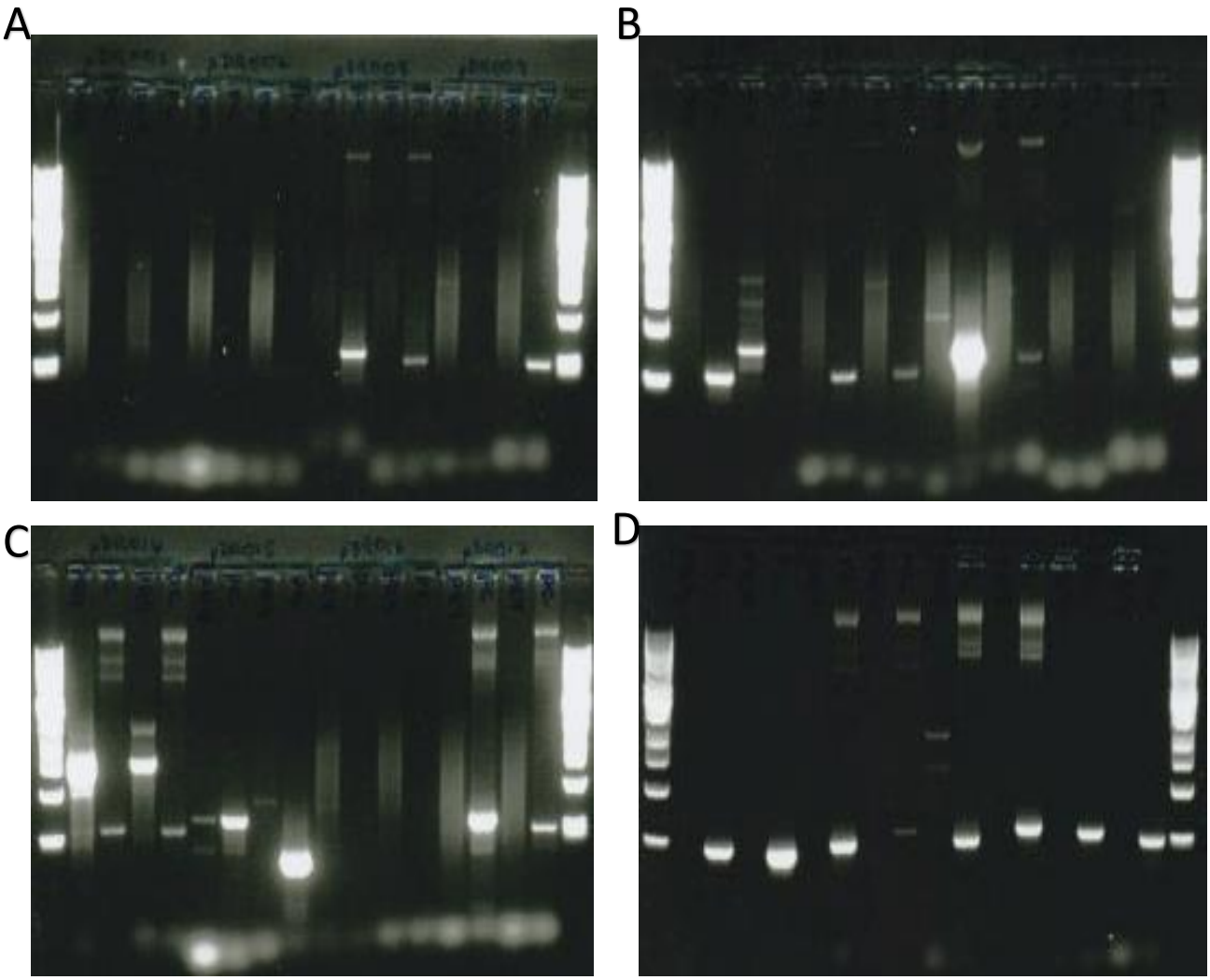


Figure A2: PCR testing integration of plasmids into *R. erythropolis* PR4. Even lanes are the genomic DNA of potential mutants, odd lanes are the plasmid positive control for all gels
A) Lanes correspond to **1.** Generuler 1kb ladder **2-3.** pDR005 01+pOSV02 **4.5.** pDR005 02+pOSV01 **6-7.** pDR007 01+pOSV02 **8-9.** pDR007 02+pOSV01 **10-11.** pDR008 01+pOSV02 **11-12.** pDR008 02+pOSV01 **13-14.** pDR009 01+pOSV02 **15-16.** pDR009 02+pOSV01 **17.** Generuler 1kb ladder
B) Lanes correspond to **1.** Generuler 1kb ladder **2-3.** pDR0010 01+pOSV02 **4.5.** pDR0010 02+pOSV01 **6-7.** pDR011 01+pOSV02 **8-9.** pDR011 02+pOSV01 **10-11.** pDR012 01+pOSV02 **11-12.** pDR012 02+pOSV01 **13-14.** pDR013 01+pOSV02 **15-16.** pDR013 02+pOSV01 **17.** Generuler 1kb ladder
C) Lanes correspond to **1.** Generuler 1kb ladder **2-3.** pDR014 01+pOSV02 **4.5.** pDR014 02+pOSV01 **6-7.** pDR0015 01+pOSV02 **8-9.** pDR015 02+pOSV01 **10-11.** pDR016 01+pOSV02 **11-12.** pDR016 02+pOSV01 **13-14.** pDR017 01+pOSV02 **15-16.** pDR017 02+pOSV01 **17.** Generuler 1kb ladder
D) Lanes correspond to **1.** Generuler 1kb ladder **2-3.** pDR018 01+pOSV02 **4.5.** pDR018 02+pOSV01 **6-7.** pDR019 01+pOSV02 **8-9.** pDR019 02+pOSV01 **10-11.** pDR020 01+pOSV02 **11-12.** pDR020 02+pOSV01 **13-14.** pDR021 01+pOSV02 **15-16.** pDR021 02+pOSV01 **17.** Generuler 1kb ladder

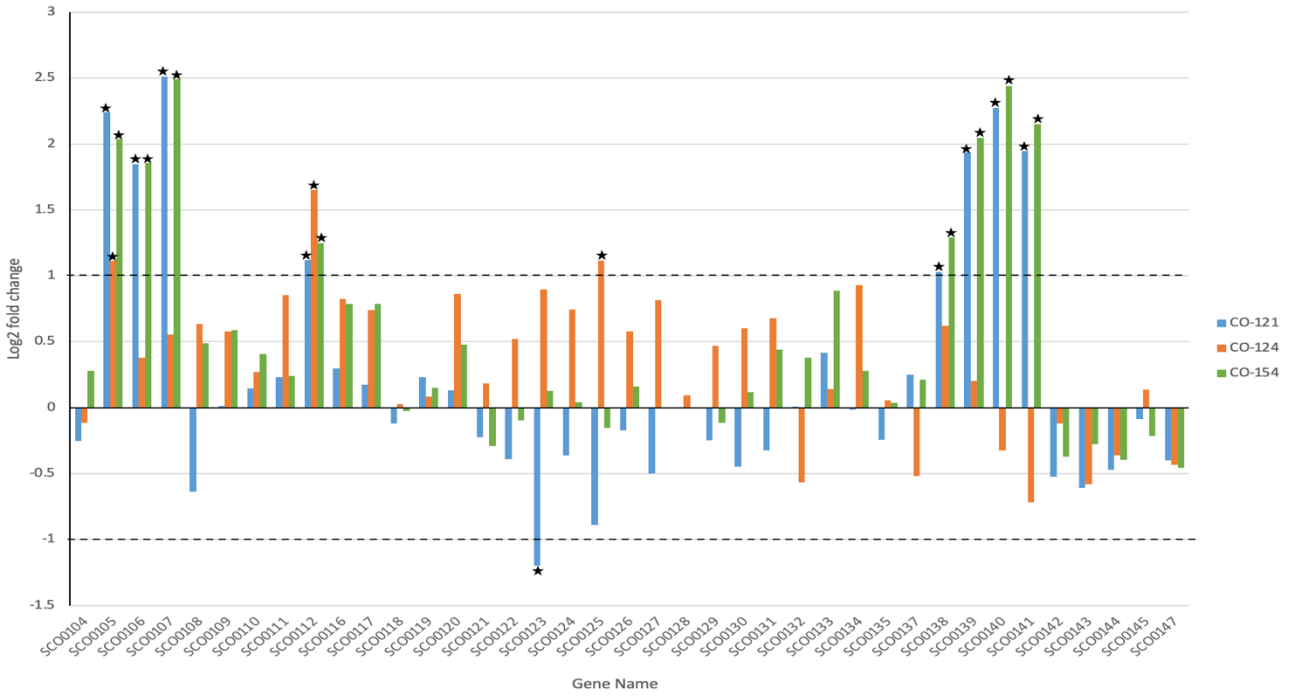


Figure A3: The log2 fold change of RNA transcription for the first antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to *S. coelicolor* M145 grown in SMM. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05.

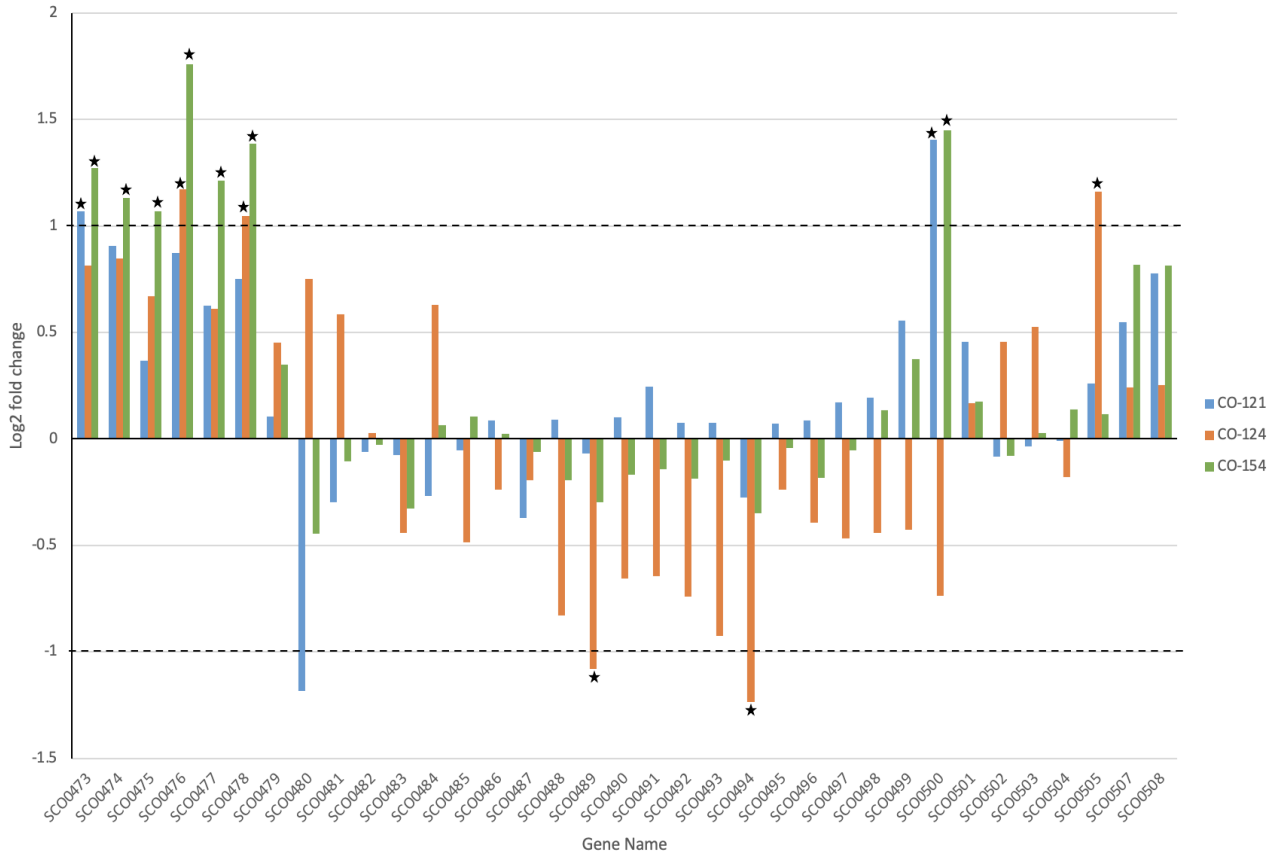


Figure A4: The log2 fold change of RNA transcription for the fourth antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to *S. coelicolor* M145 grown in SMM. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05.

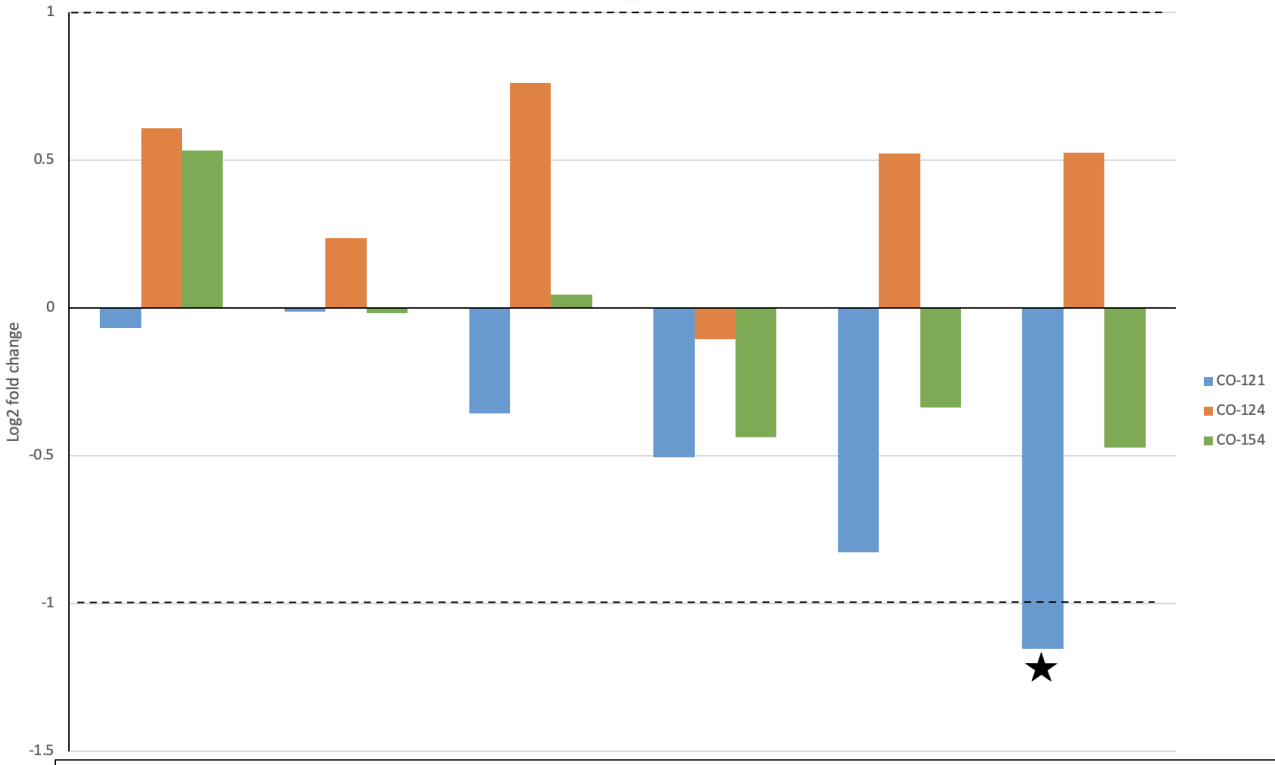


Figure A5: The log2 fold change of RNA transcription for the fifth antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to *S. coelicolor* M145 grown in SMM. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

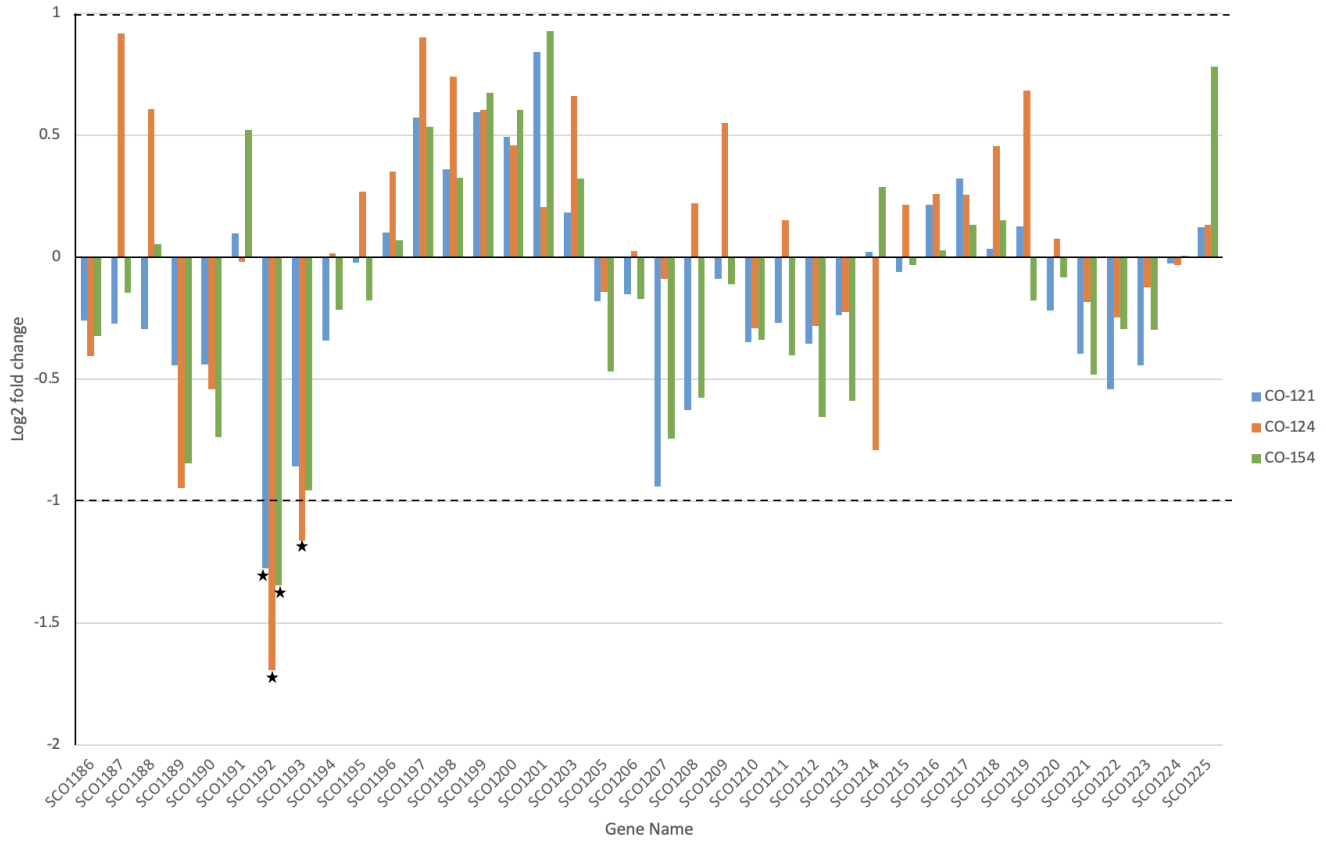


Figure A6: The log2 fold change of RNA transcription for the sixth antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to *S. coelicolor* grown in SMM. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05.

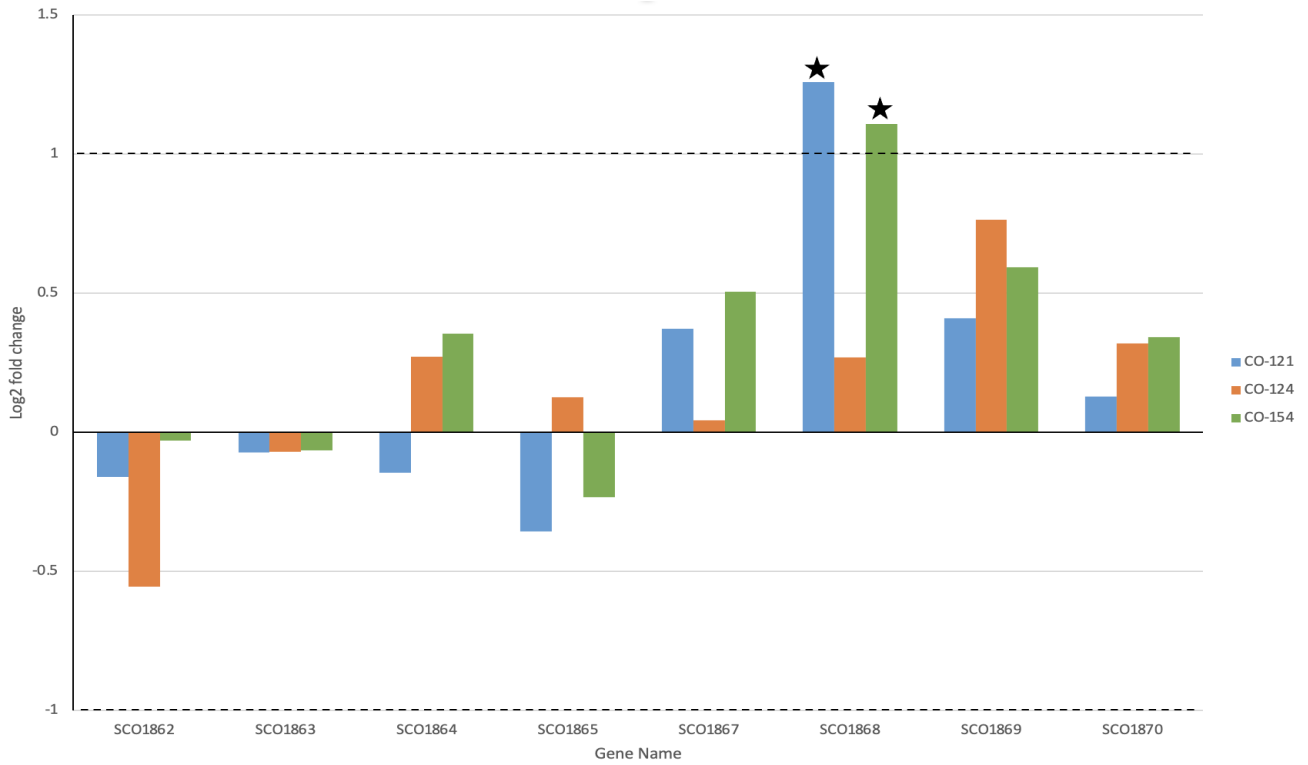


Figure A7: The log2 fold change of RNA transcription for the seventh antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to *S. coelicolor* M145 grown in SMM. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

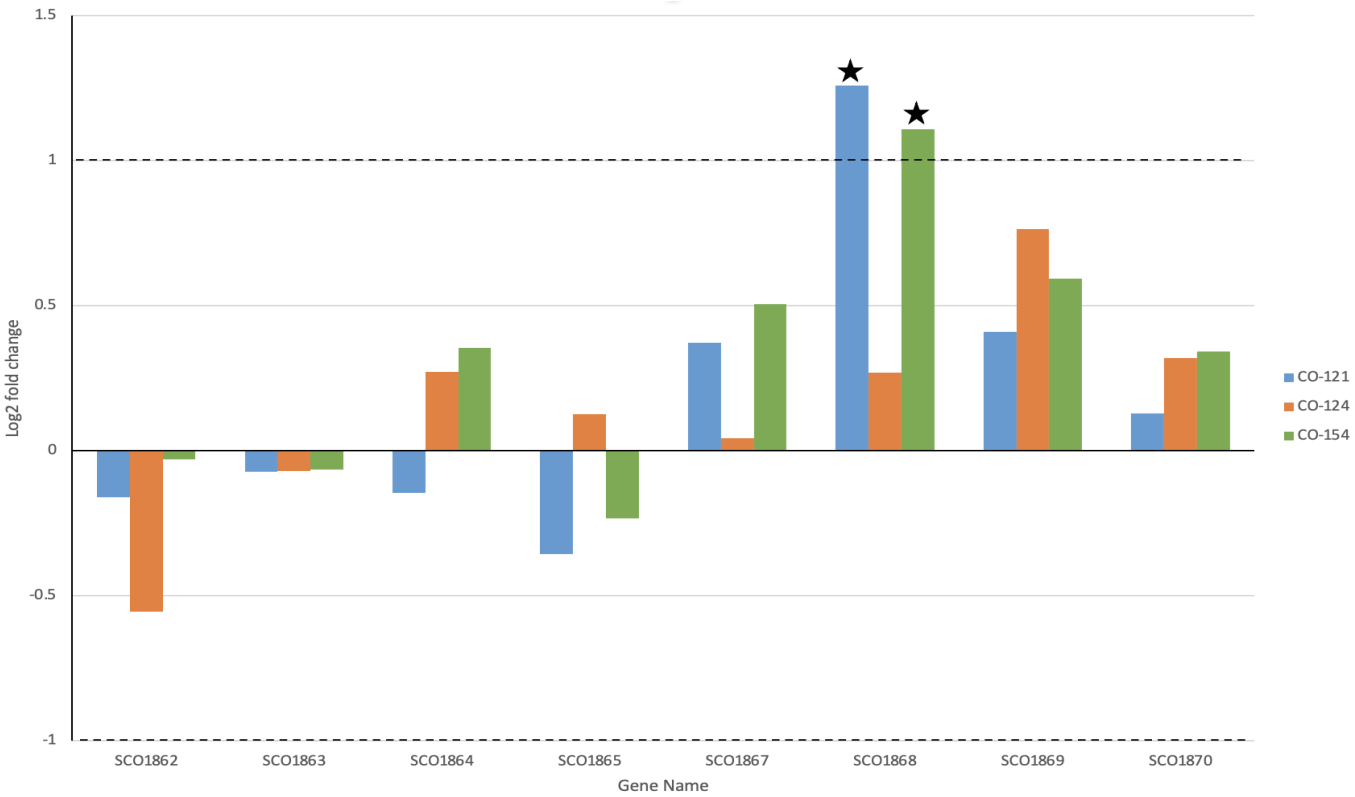


Figure A8: The log2 fold change of RNA transcription for the seventh antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to *S. coelicolor* M145 grown in SMM. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

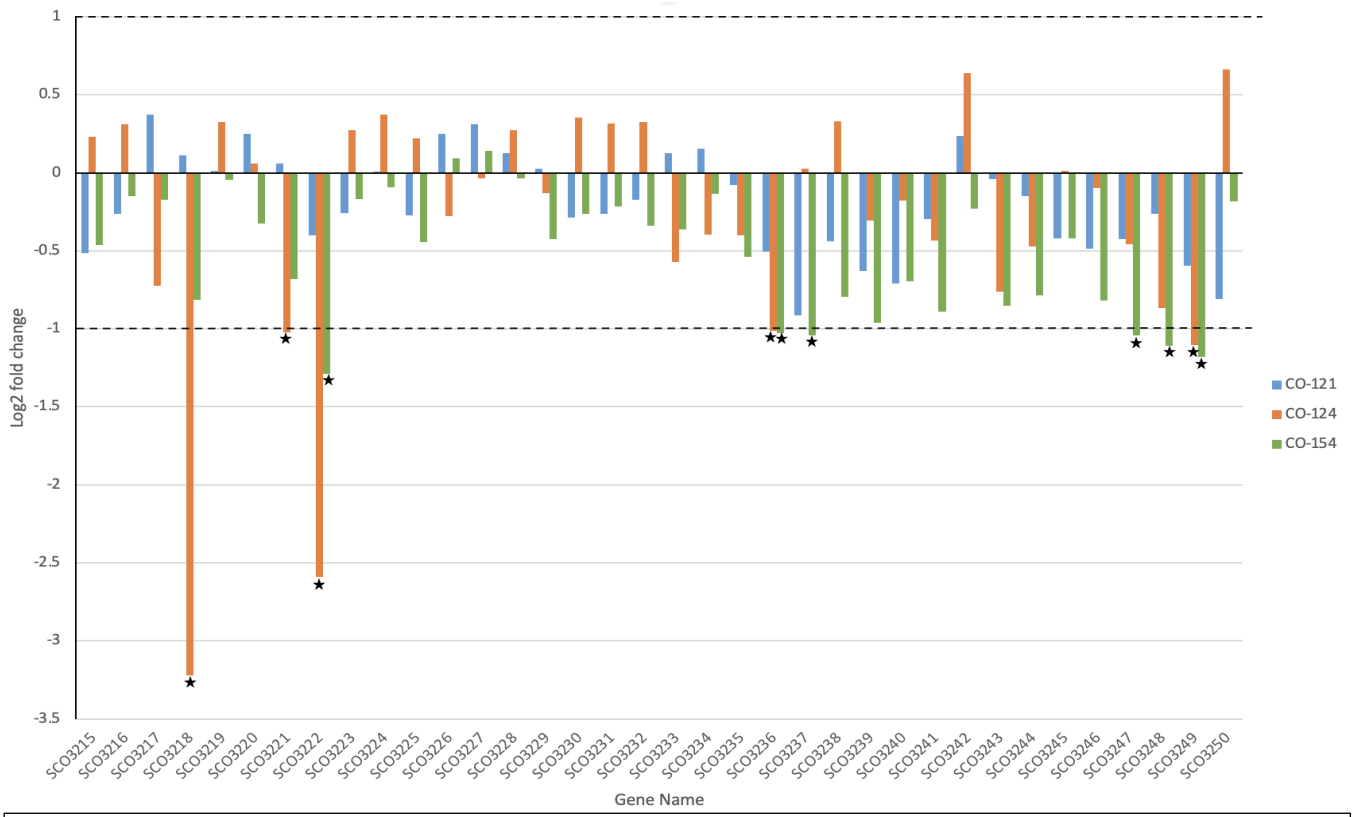


Figure A9: The log2 fold change of RNA transcription for the tenth antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to *S. coelicolor* M145 grown in SMM. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

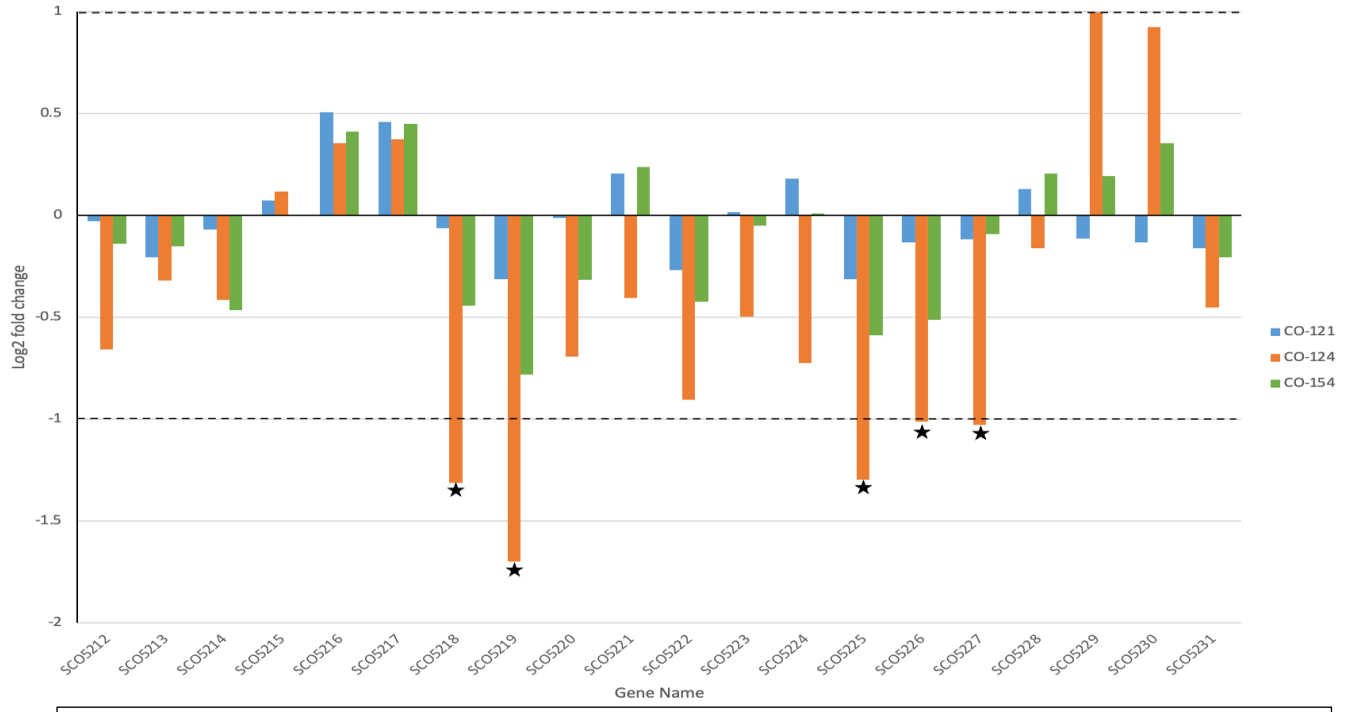


Figure A10: The log2 fold change of RNA transcription for the twelfth antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to *S. coelicolor* M145 grown in SMM. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

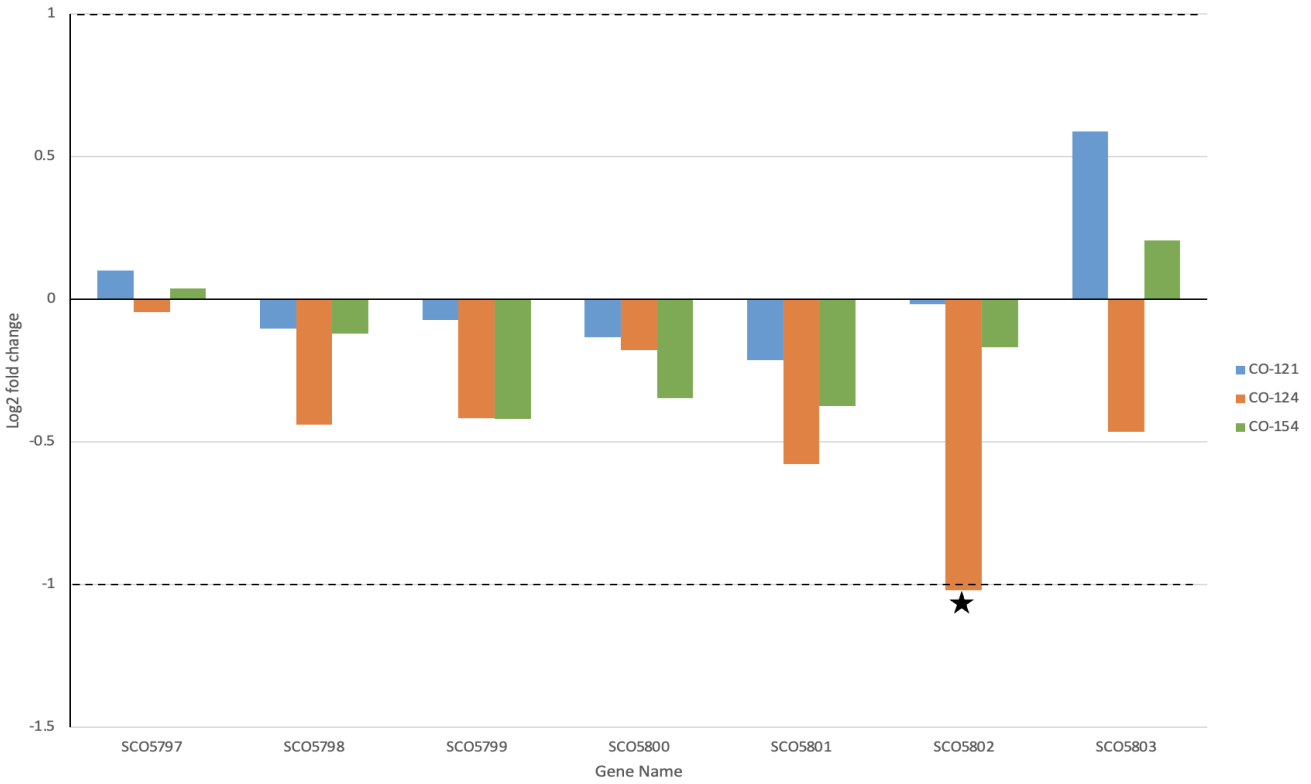


Figure A11: The log2 fold change of RNA transcription for the fourteenth antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to *S. coelicolor* M145 grown in SMM. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05.

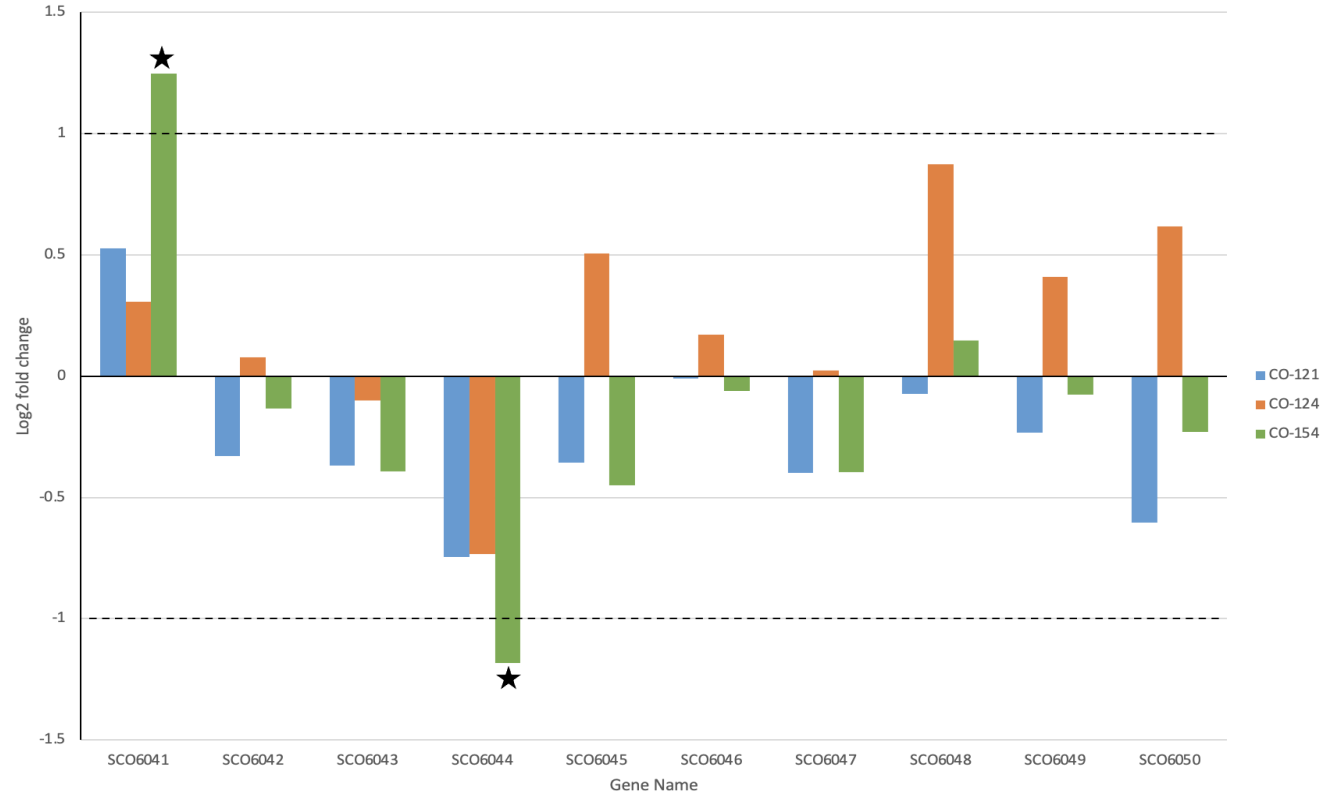


Figure A12: The log2 fold change of RNA transcription for the sixteenth antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to *S. coelicolor* M145 grown in SMM. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

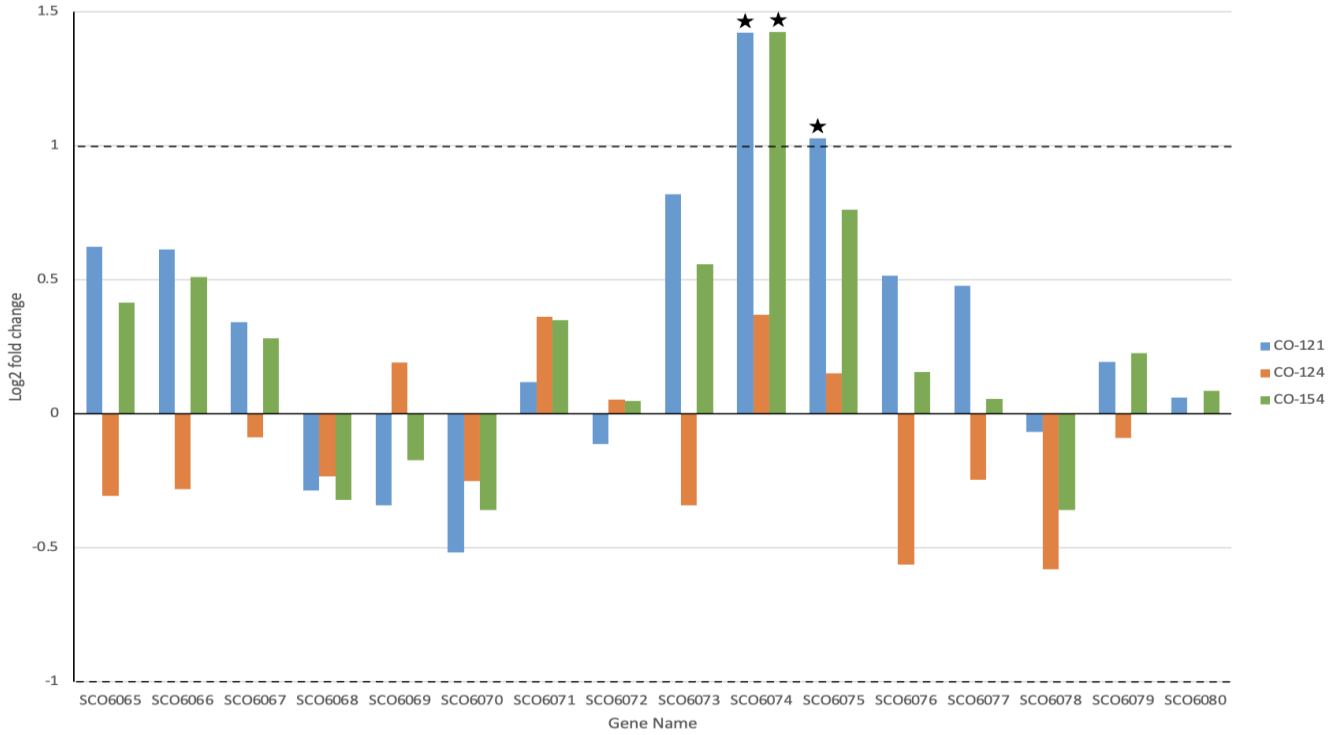


Figure A13: The log2 fold change of RNA transcription for the seventeenth antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to *S. coelicolor* M145 grown in SMM. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05.

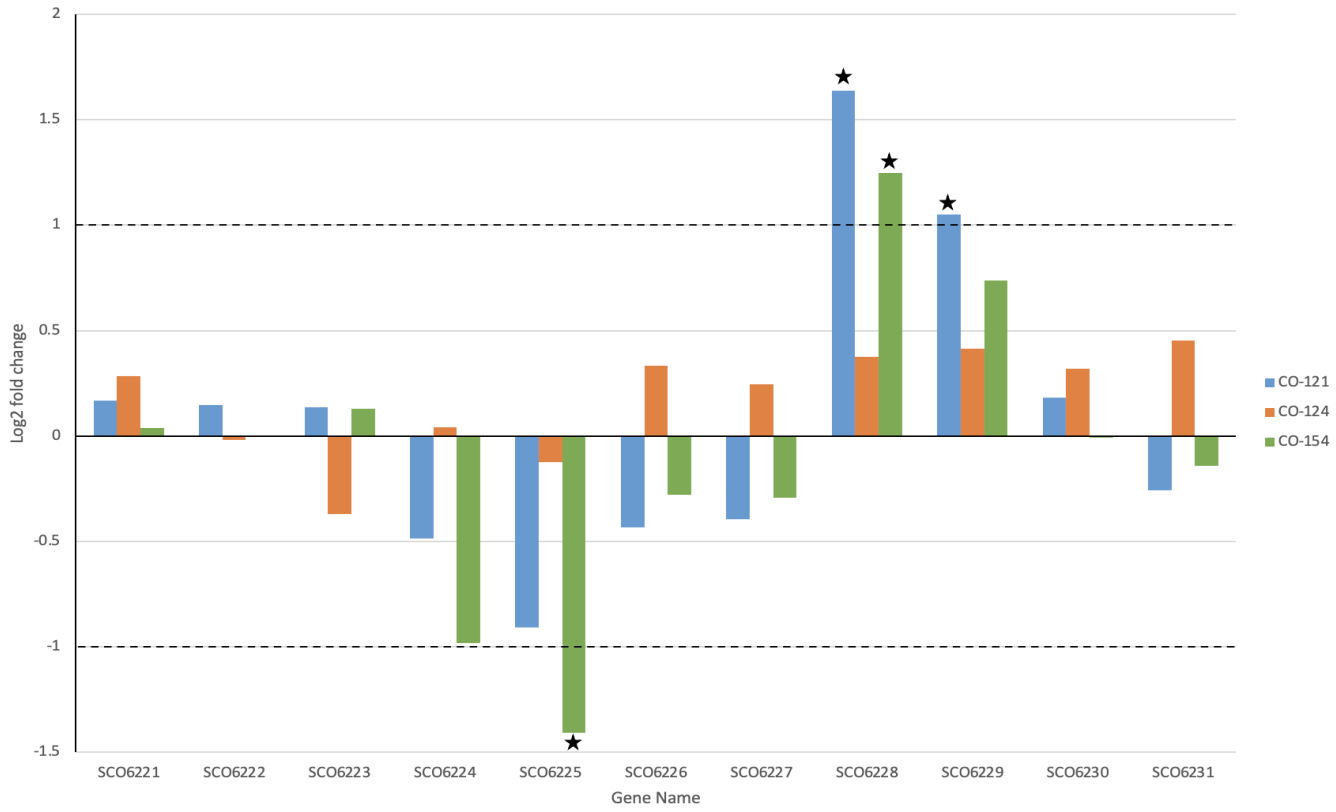


Figure A14: The log2 fold change of RNA transcription for the eighteenth antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to *S. coelicolor* M145 grown in SMM. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

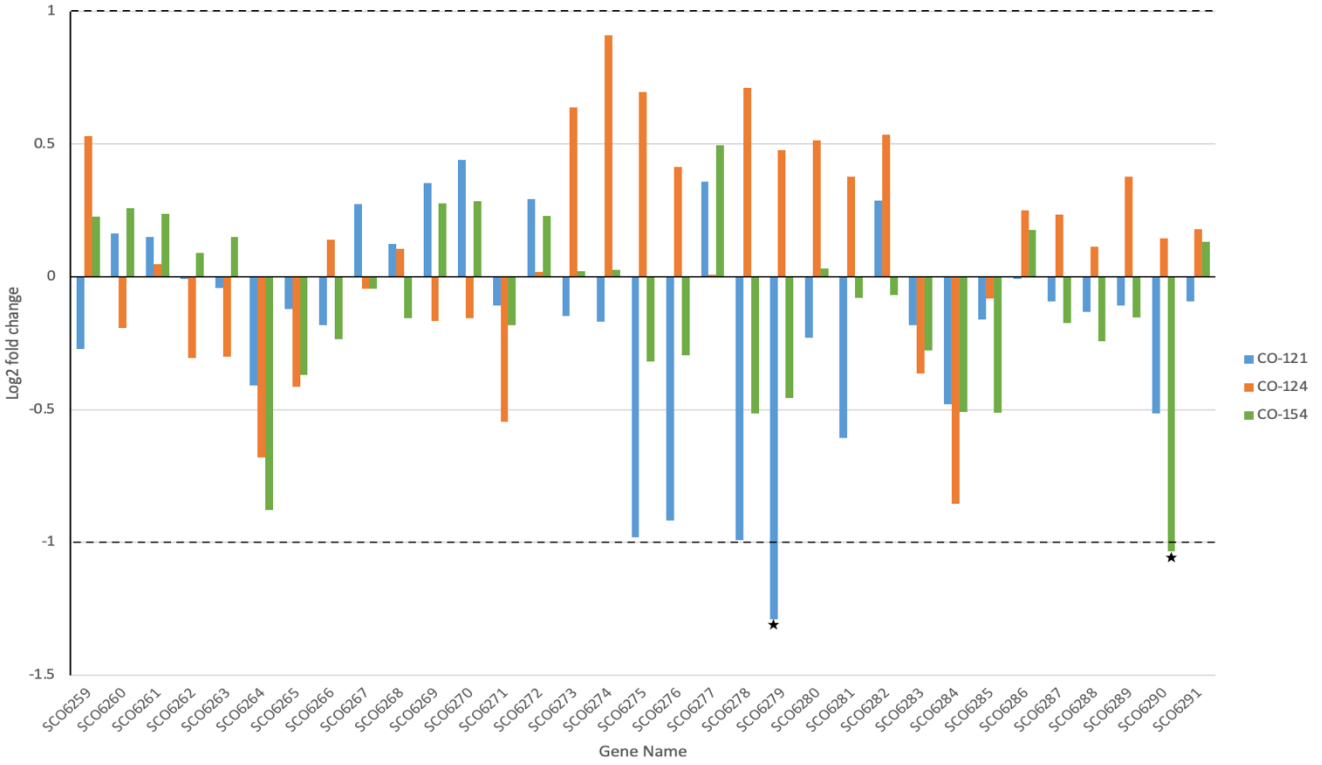


Figure A15: The log2 fold change of RNA transcription for the nineteenth antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to *S. coelicolor* M145 grown in SMM. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05.

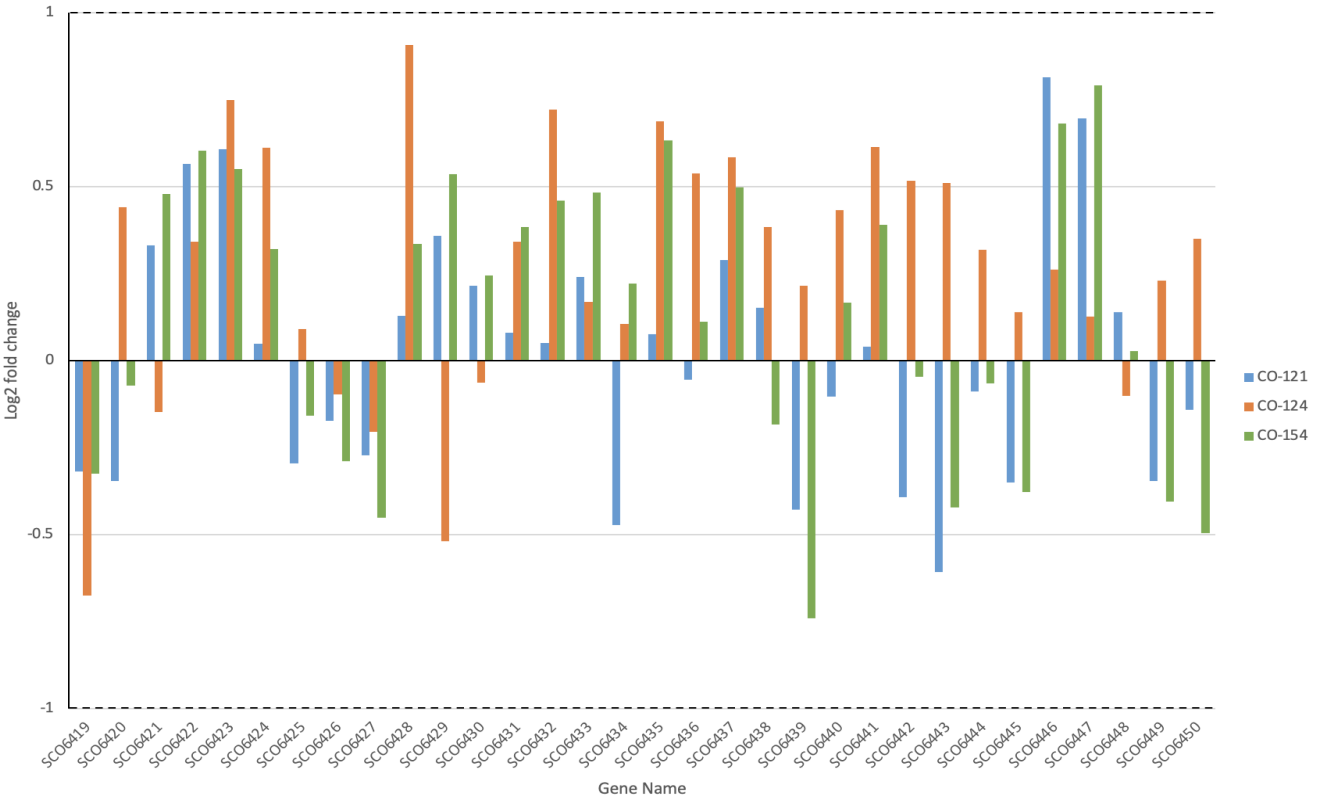


Figure A16: The log2 fold change of RNA transcription for the twentieth antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to *S. coelicolor* M145 grown in SMM. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

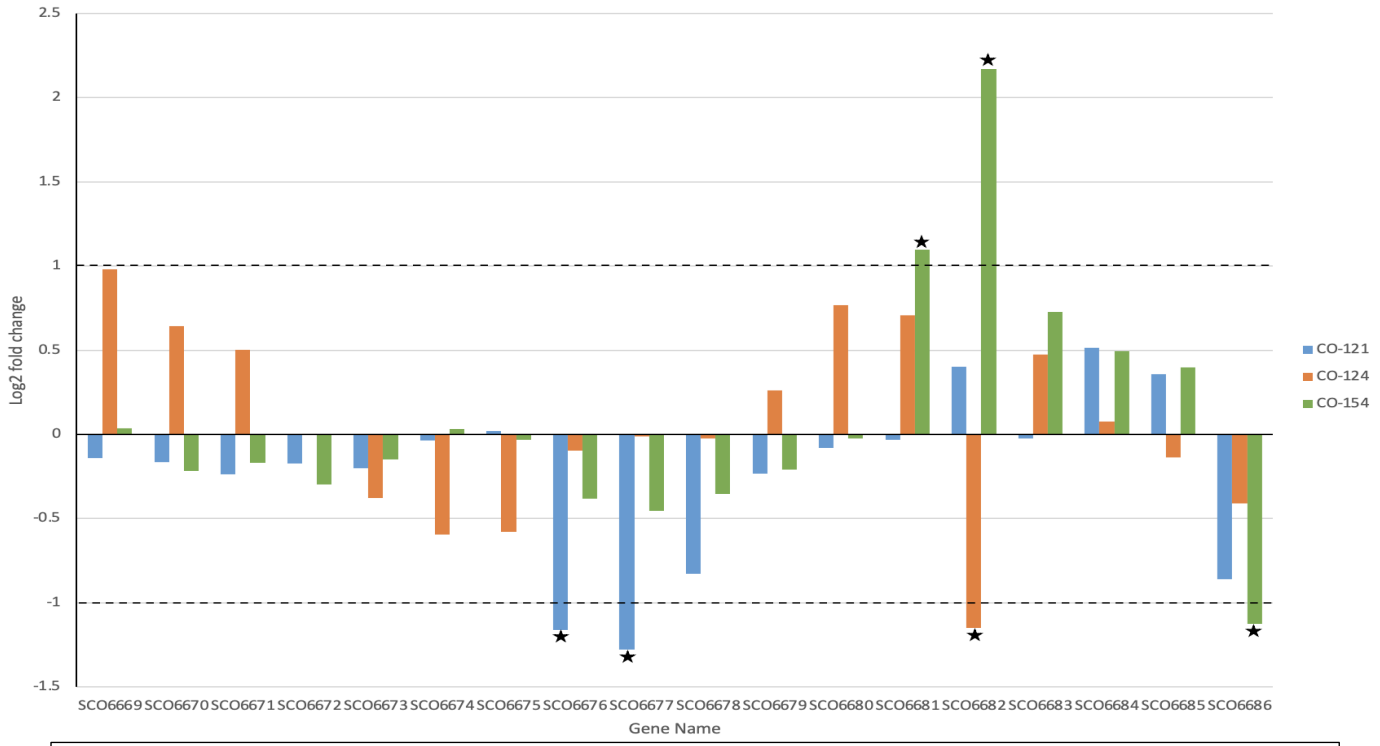


Figure A17: The log2 fold change of RNA transcription for the twenty-first antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to *S. coelicolor* M145 grown in SMM. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05.

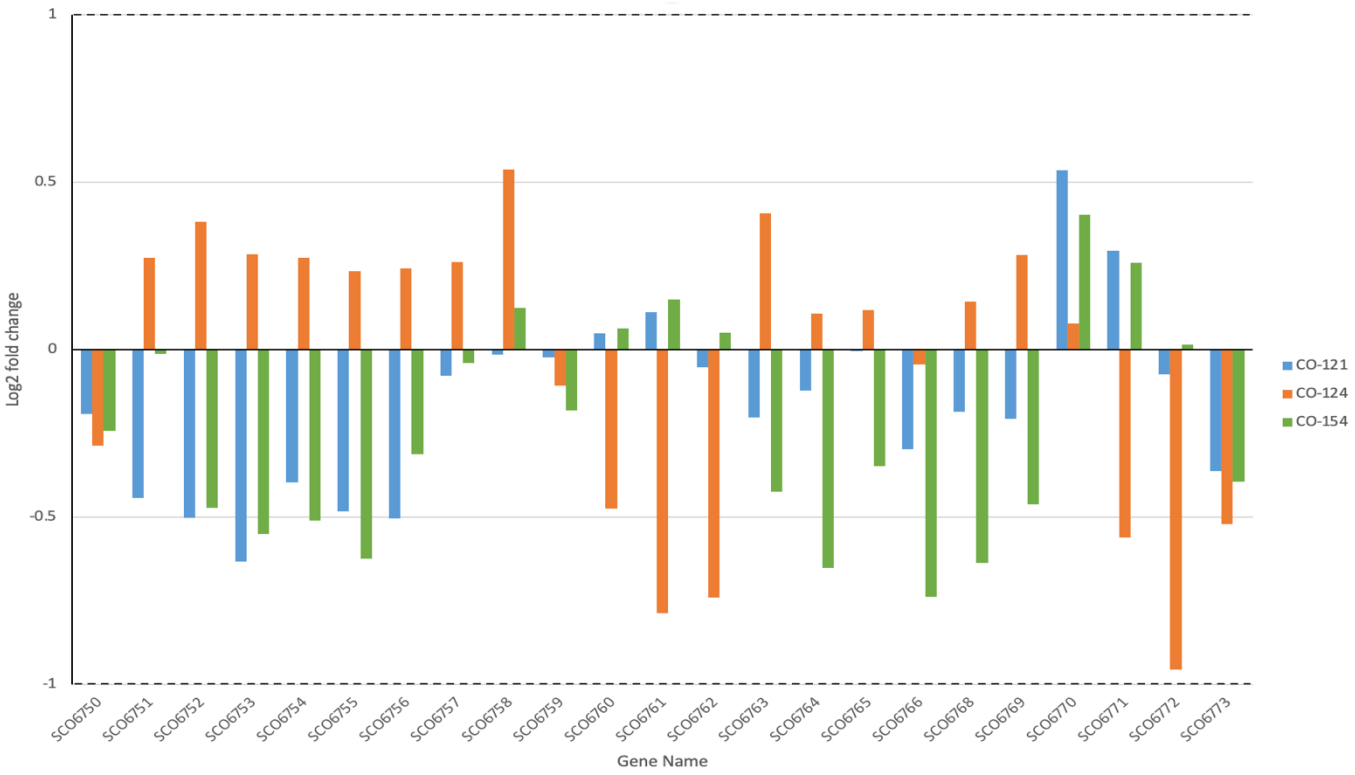


Figure A18: The log2 fold change of RNA transcription for the twenty-second antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to *S. coelicolor* M145 grown in SMM. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05.

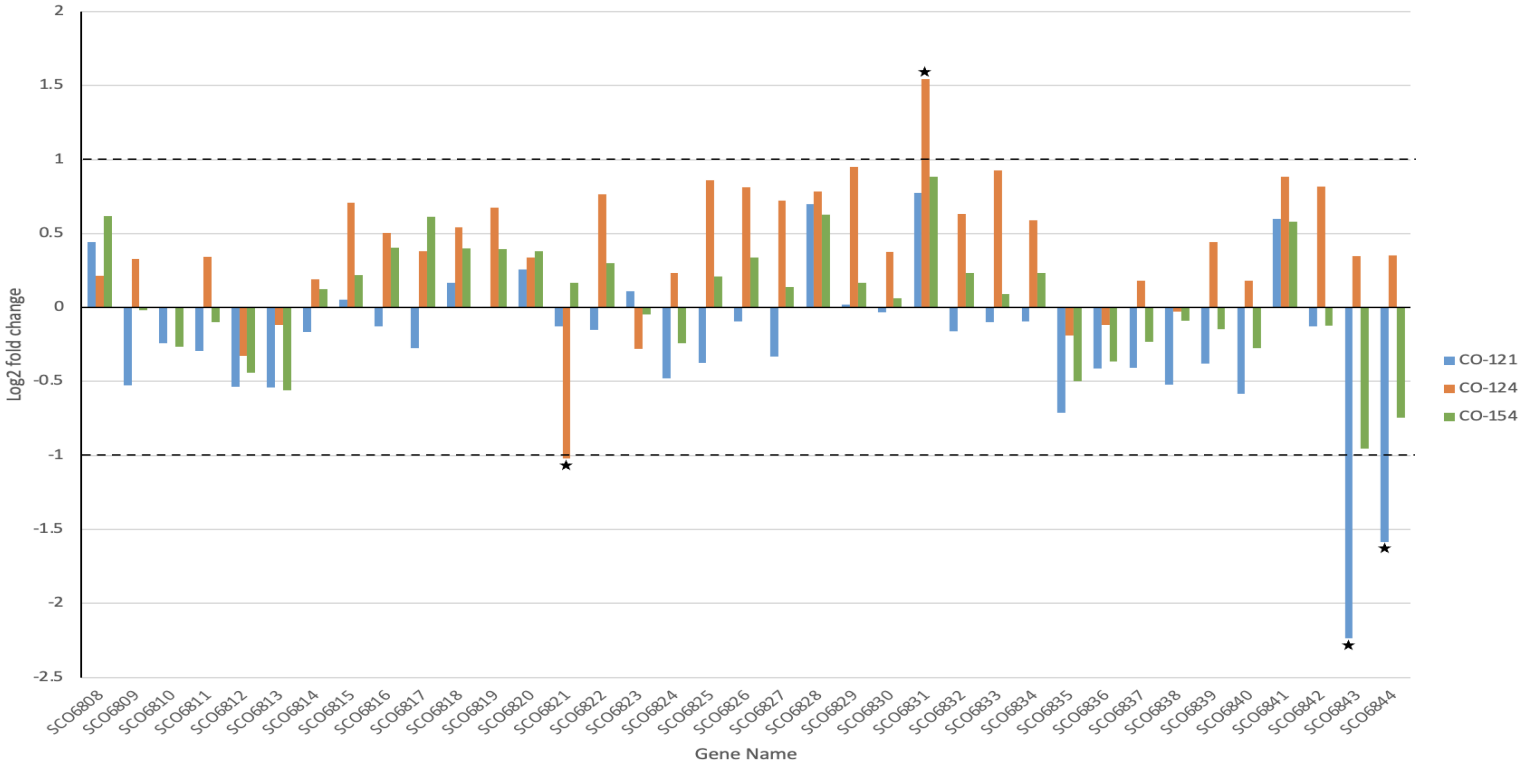


Figure A19: The log2 fold change of RNA transcription for the twenty-third antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to *S. coelicolor* M145 grown in SMM. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05.

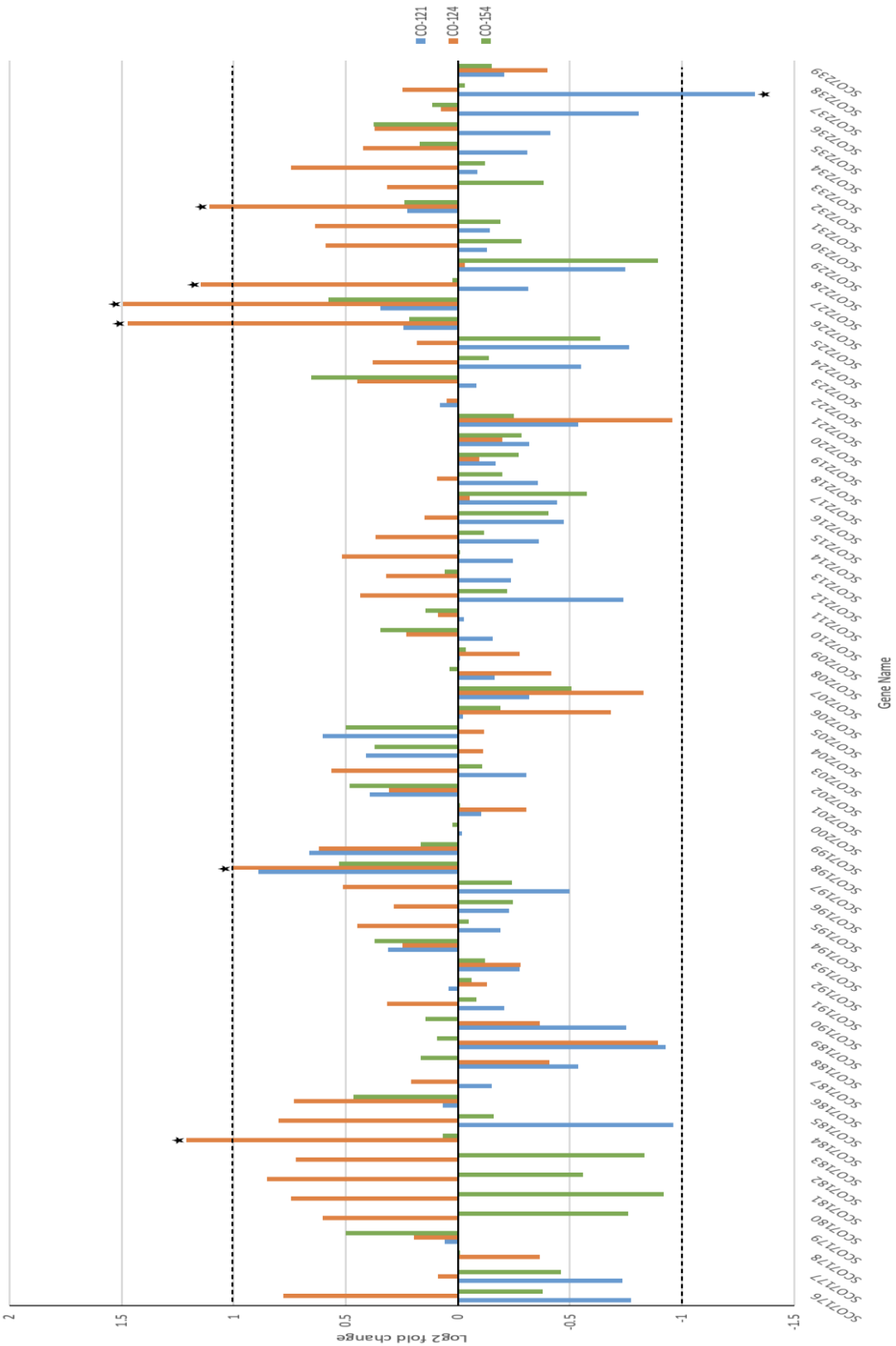


Figure A20: The log2 fold change of RNA transcription for the twenty-fifth antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to *S. coelicolor* M145 grown in SMM. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05.

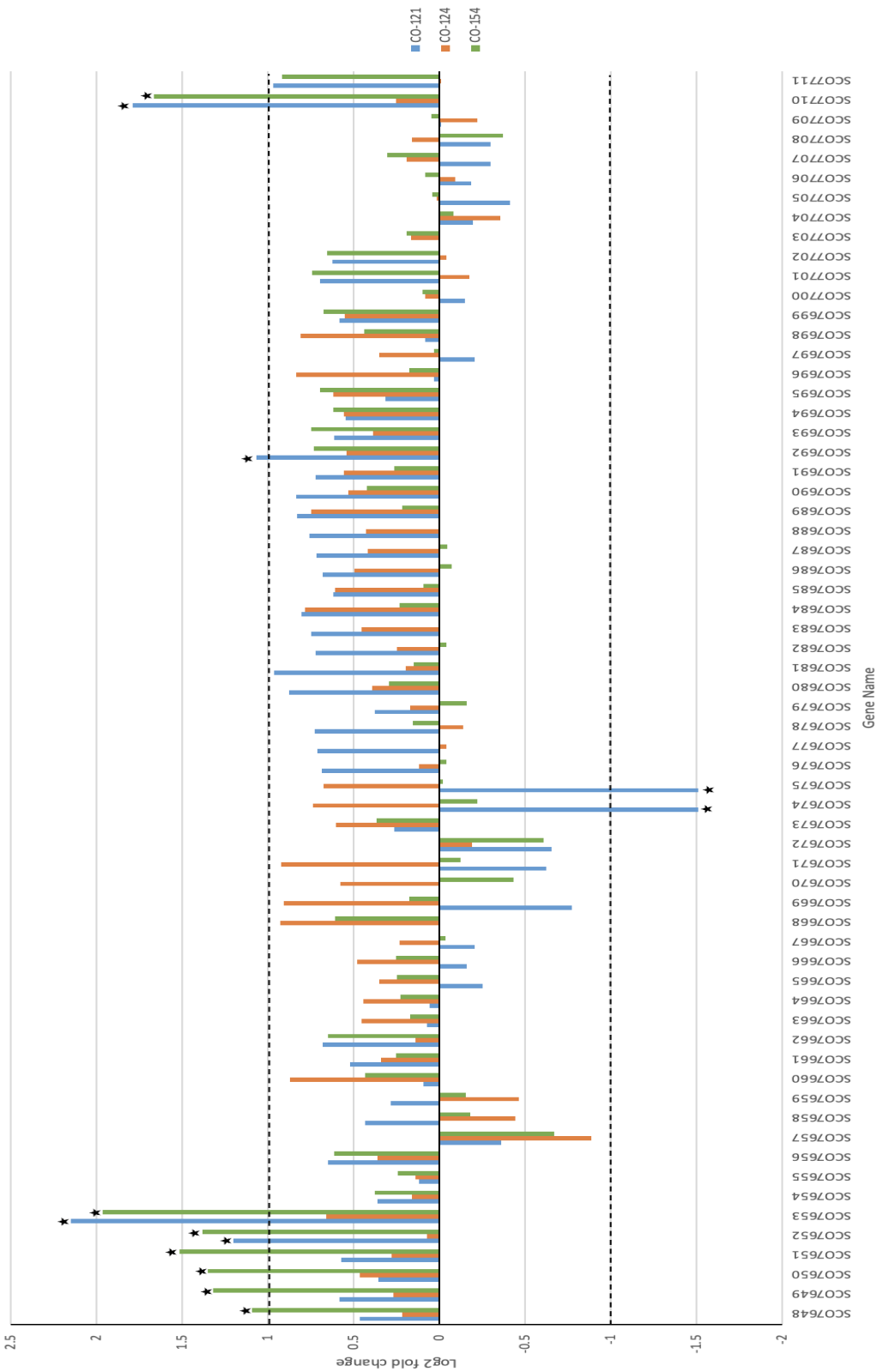


Figure A21: The log2 fold change of RNA transcription for the twenty-seventh antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to *S. coelicolor* M145 grown in SMM. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

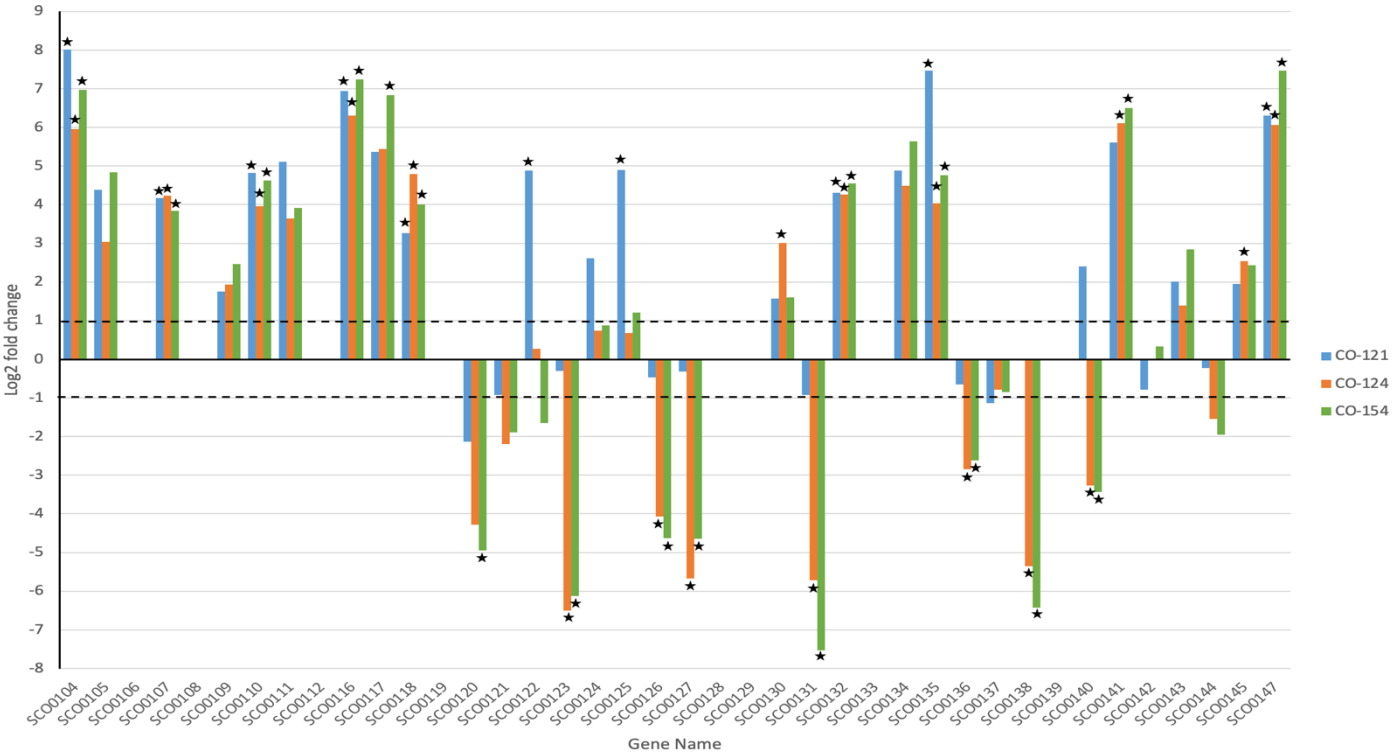


Figure A22: The log2 fold change of RNA transcription for the first antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to the empty vector control CO-250 grown in YEME media. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

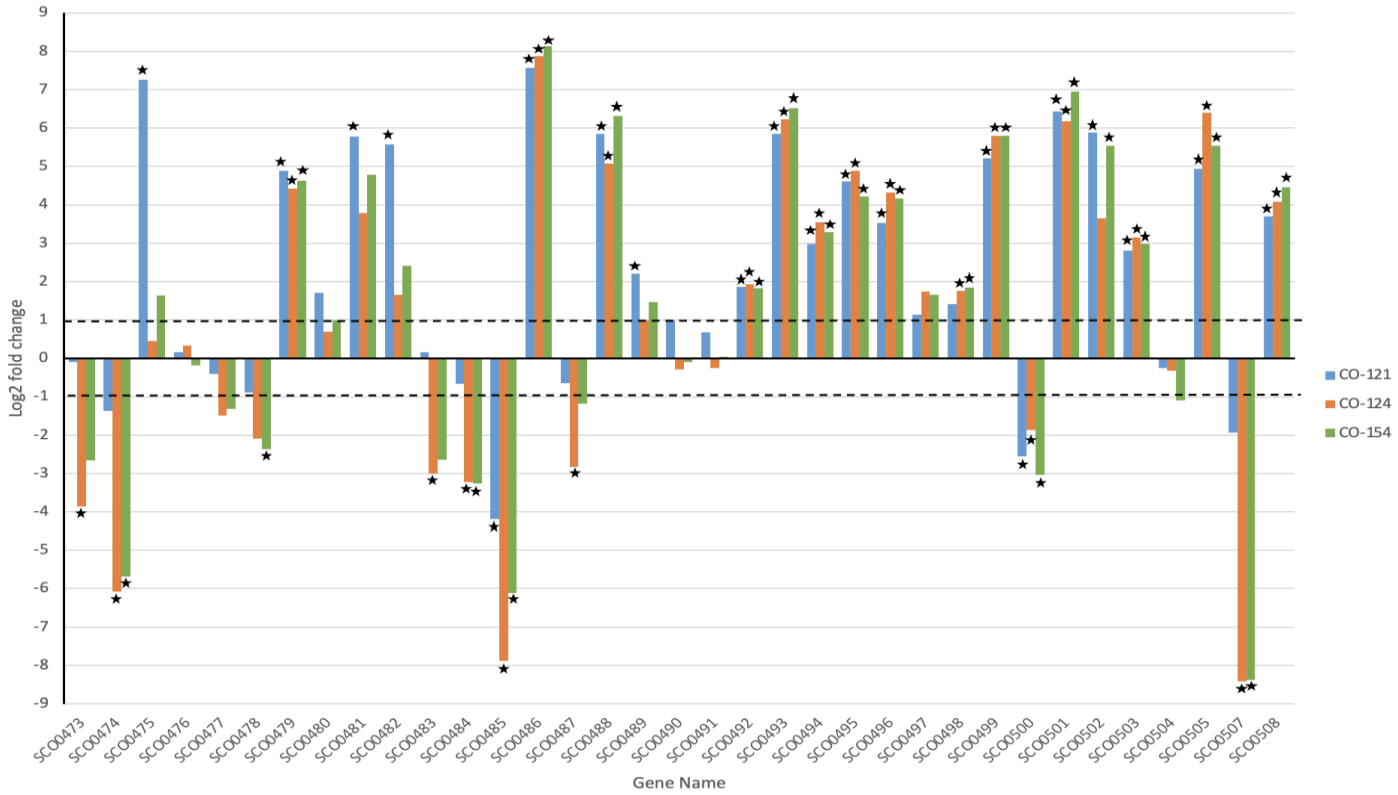


Figure A23: The log2 fold change of RNA transcription for the fourth antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to the empty vector control CO-250 grown in YEME media. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

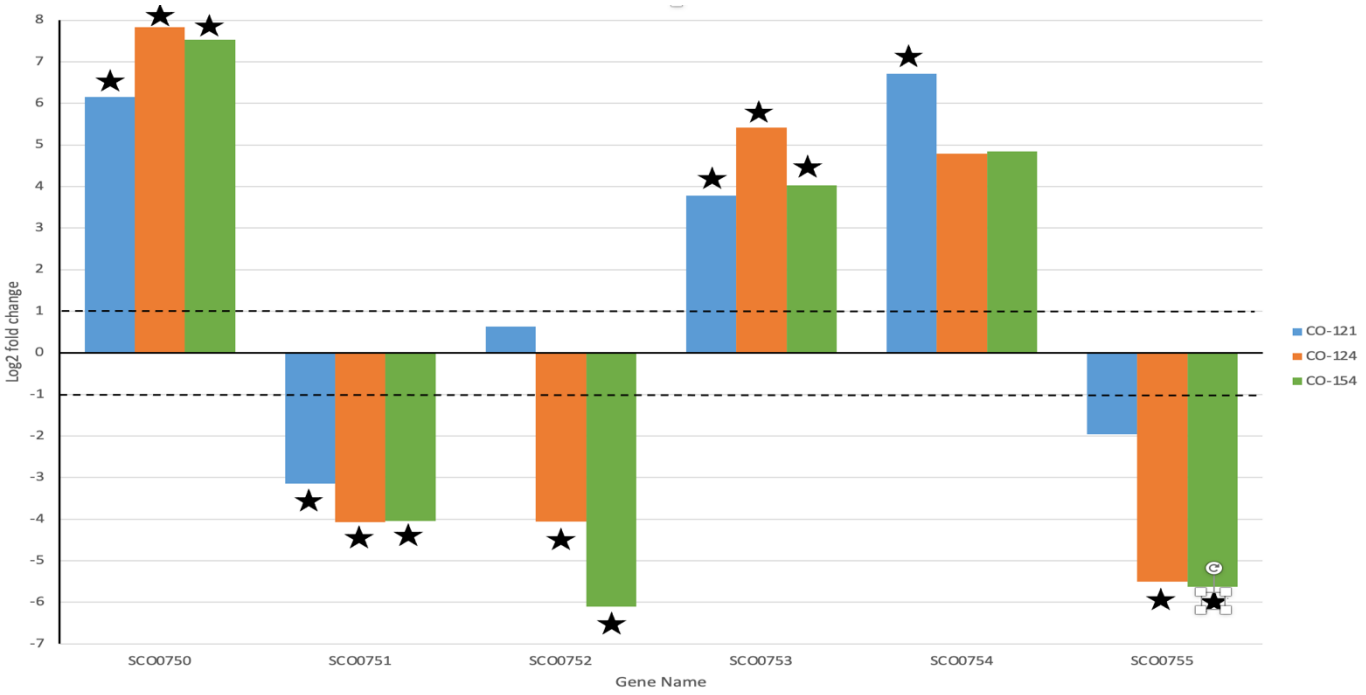


Figure A24: Figure 3.37: The log2 fold change of RNA transcription for the fifth antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to the empty vector control CO-250 grown in YEME media. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

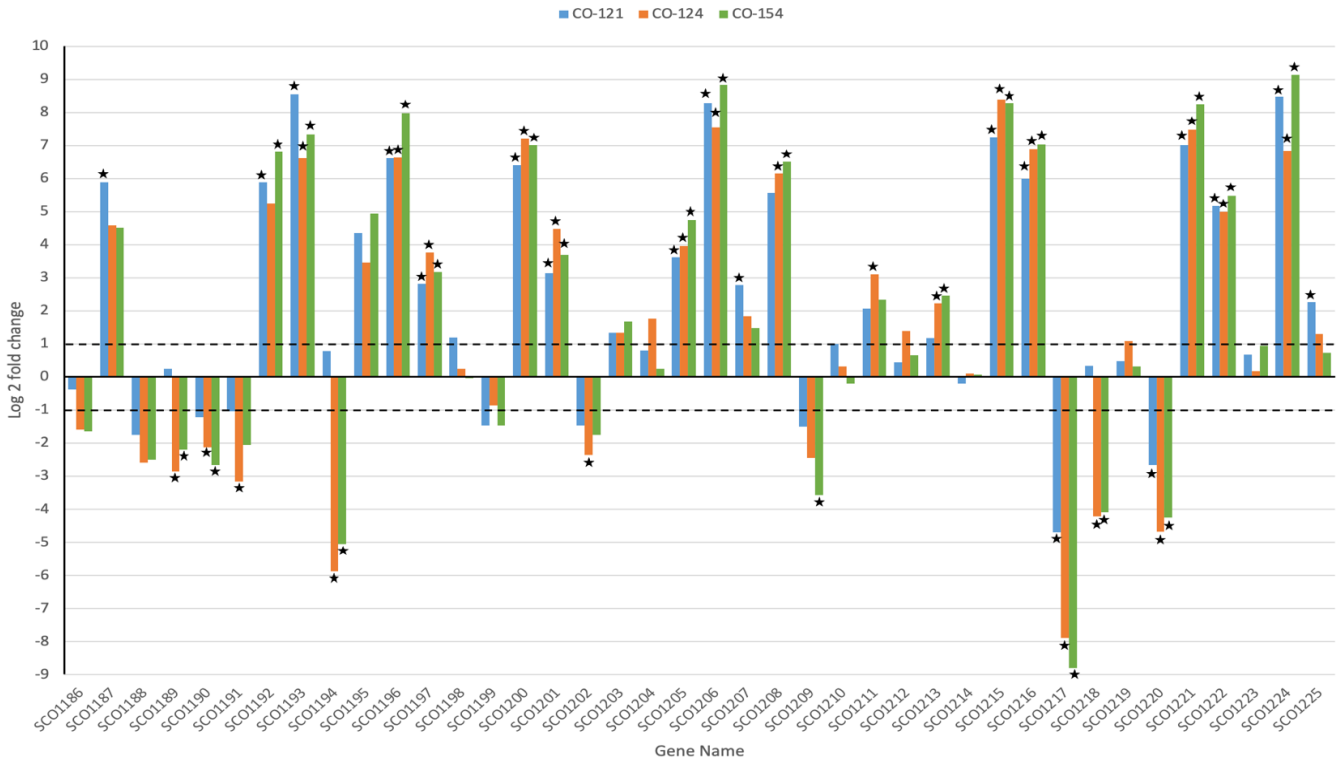


Figure A25: The log2 fold change of RNA transcription for the sixth antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to the empty vector control CO-250 grown in YEME media. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

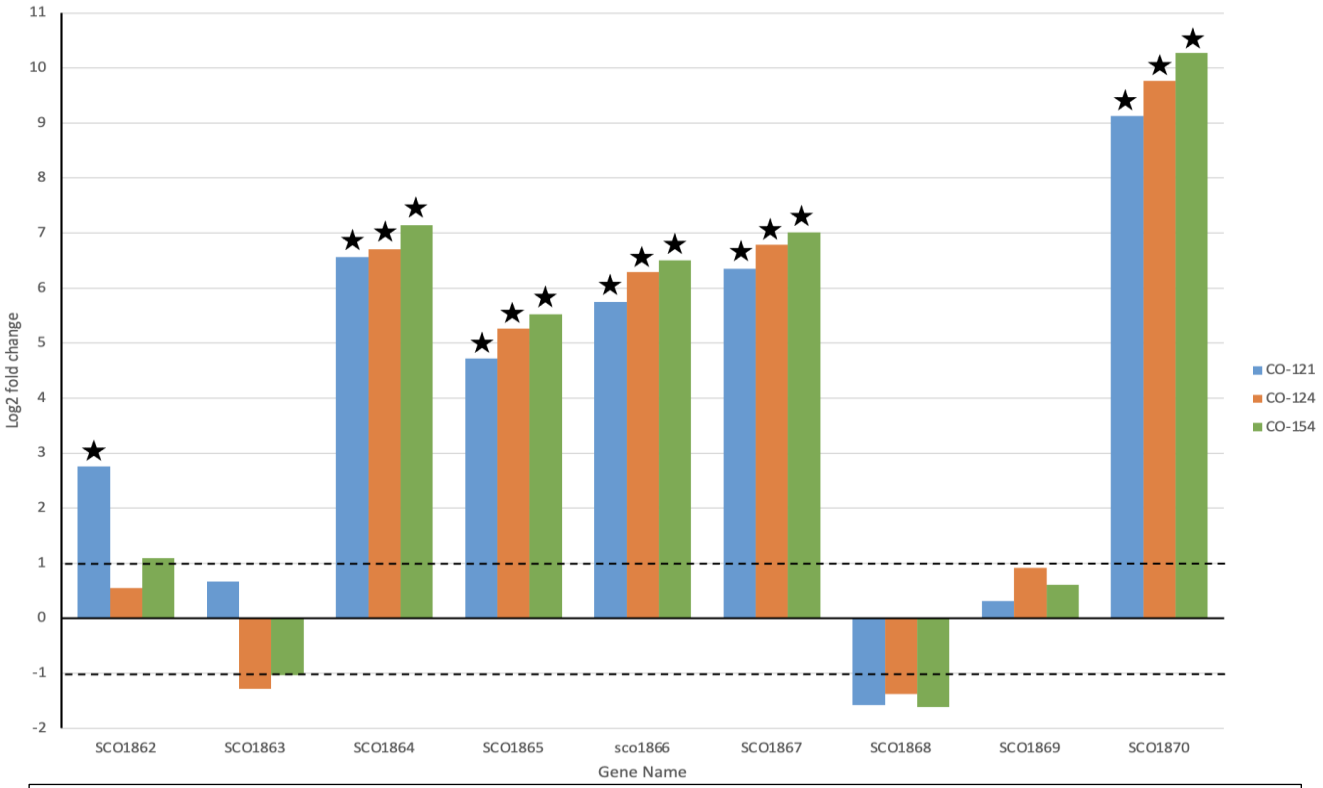


Figure A26: The log2 fold change of RNA transcription for the seventh antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to the empty vector control CO-250 grown in YEME media. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

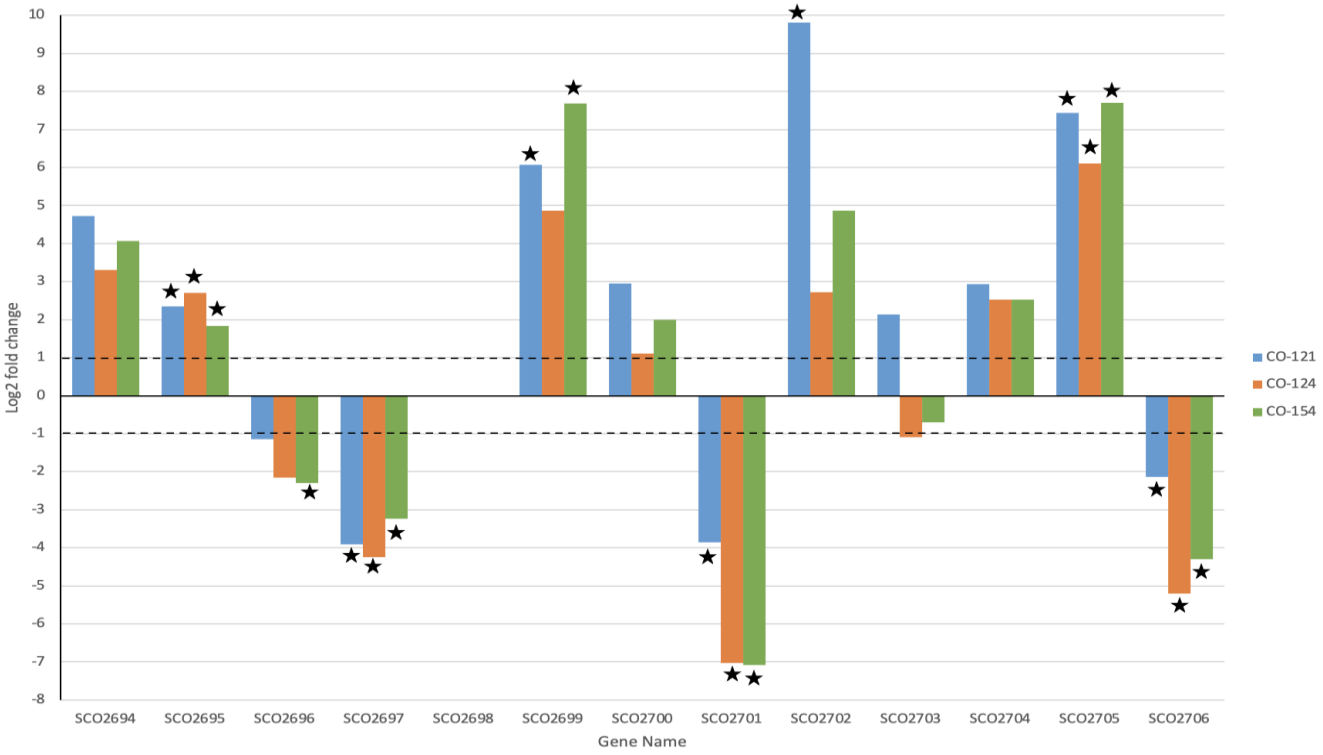


Figure A27: The log2 fold change of RNA transcription for the eighth antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to the empty vector control CO-250 grown in YEME media. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

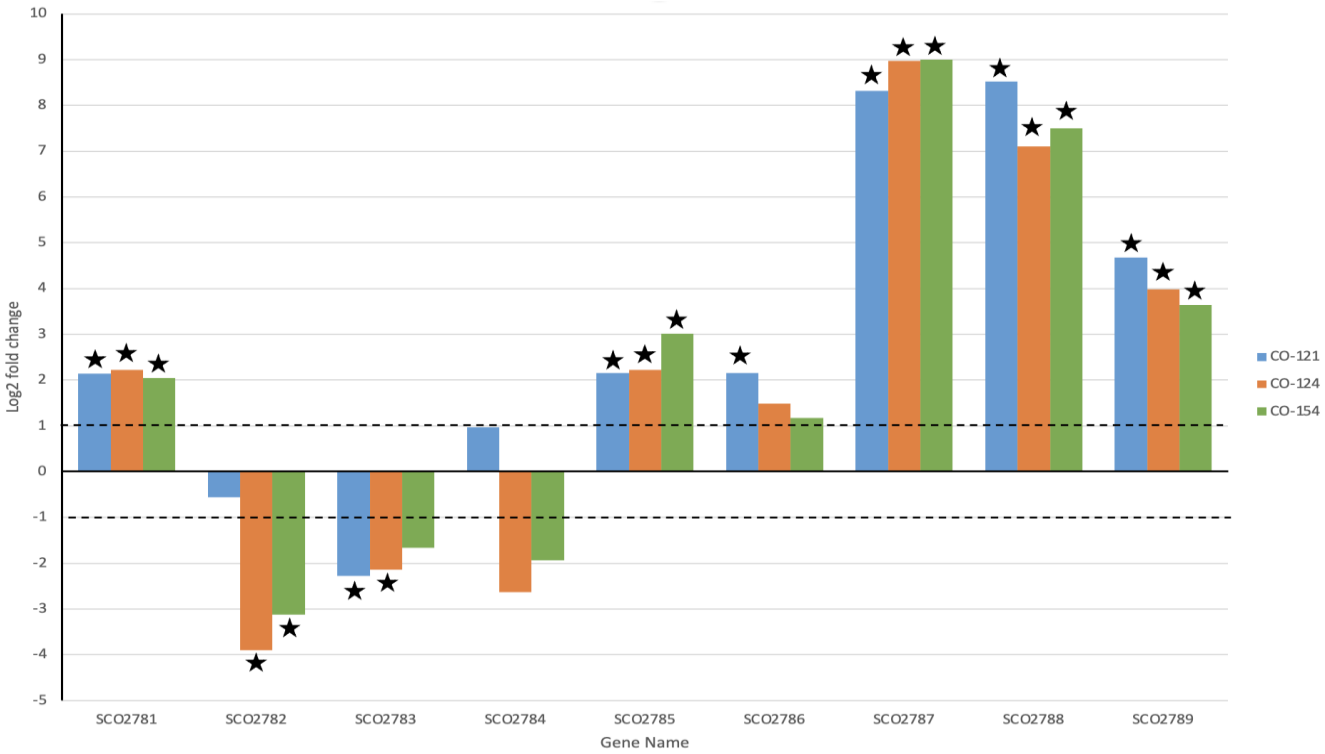


Figure A28: The log2 fold change of RNA transcription for the ninth antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to the empty vector control CO-250 grown in YEME media. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

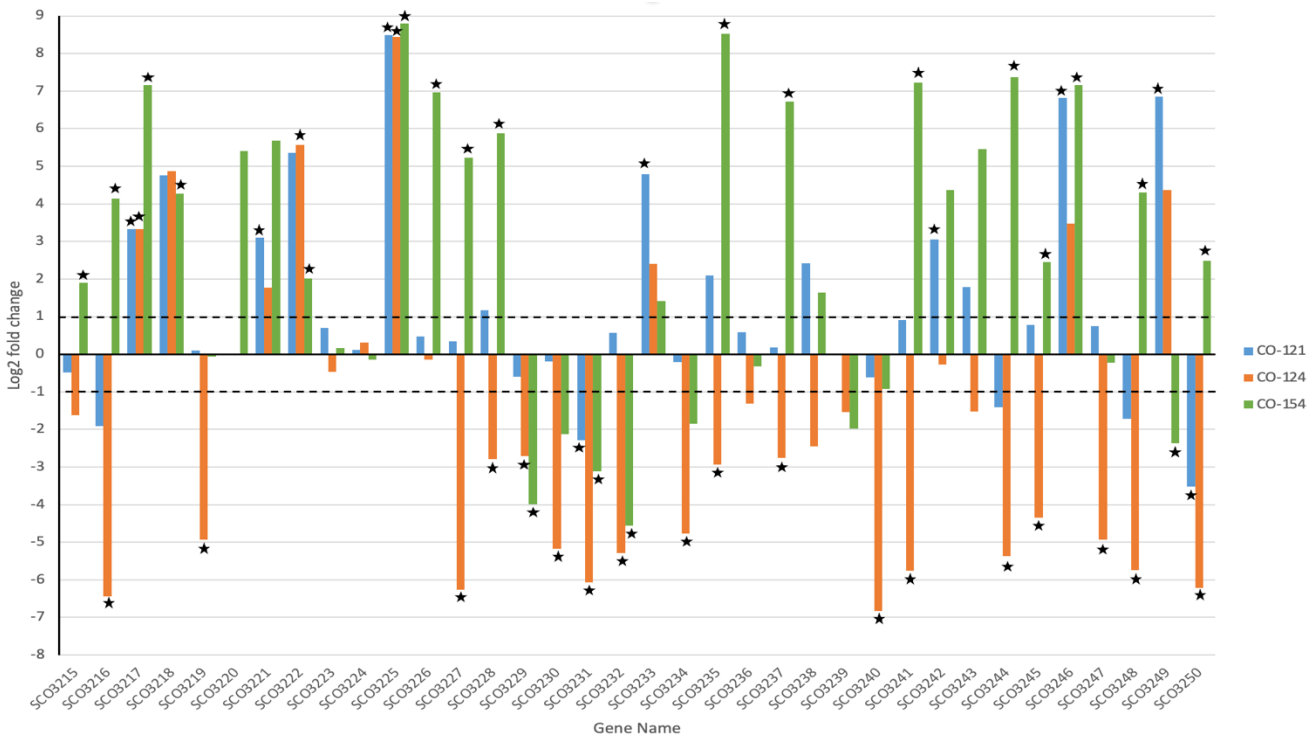


Figure A29: The log2 fold change of RNA transcription for the tenth antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to the empty vector control CO-250 grown in YEME media. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

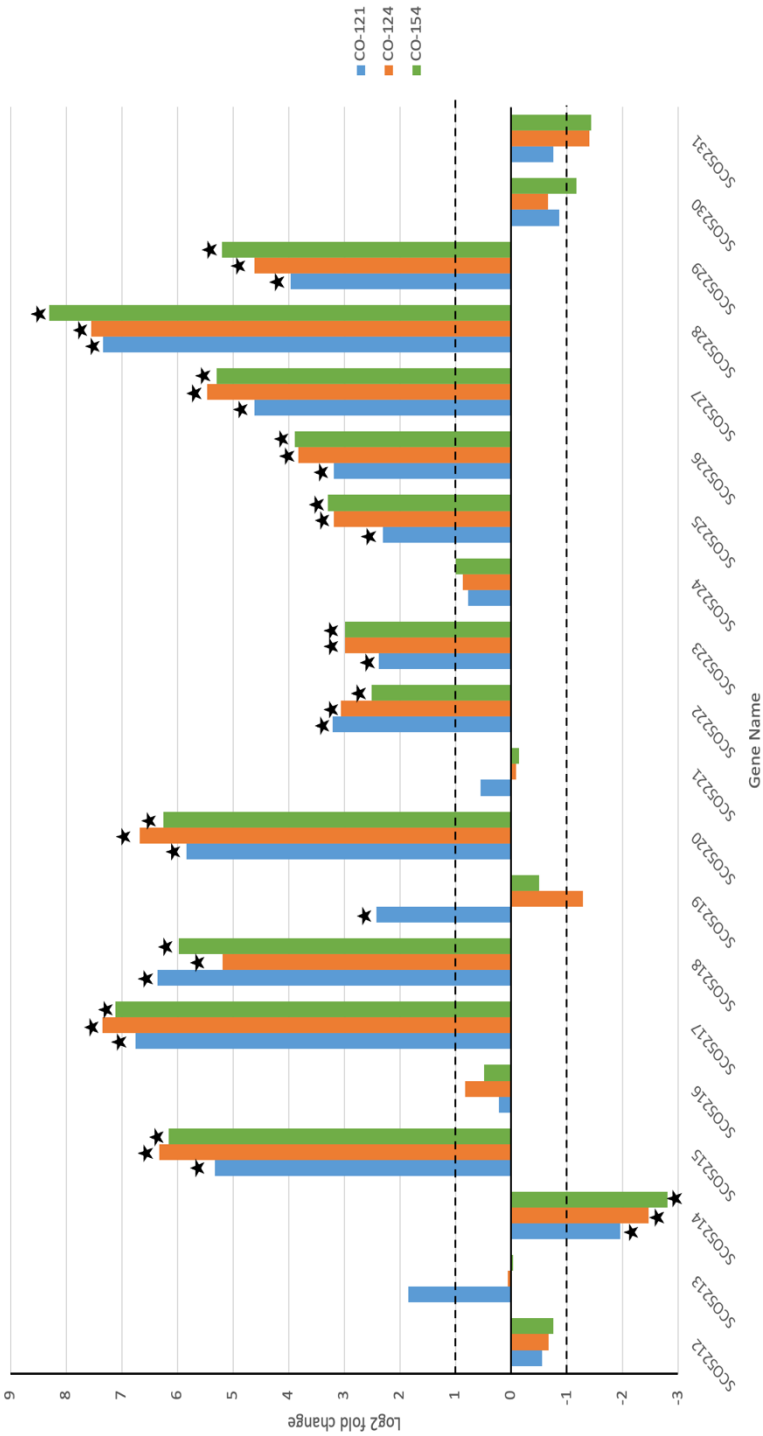


Figure A30: The log2 fold change of RNA transcription for the twelfth antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to the empty vector control CO-250 grown in YEME media. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

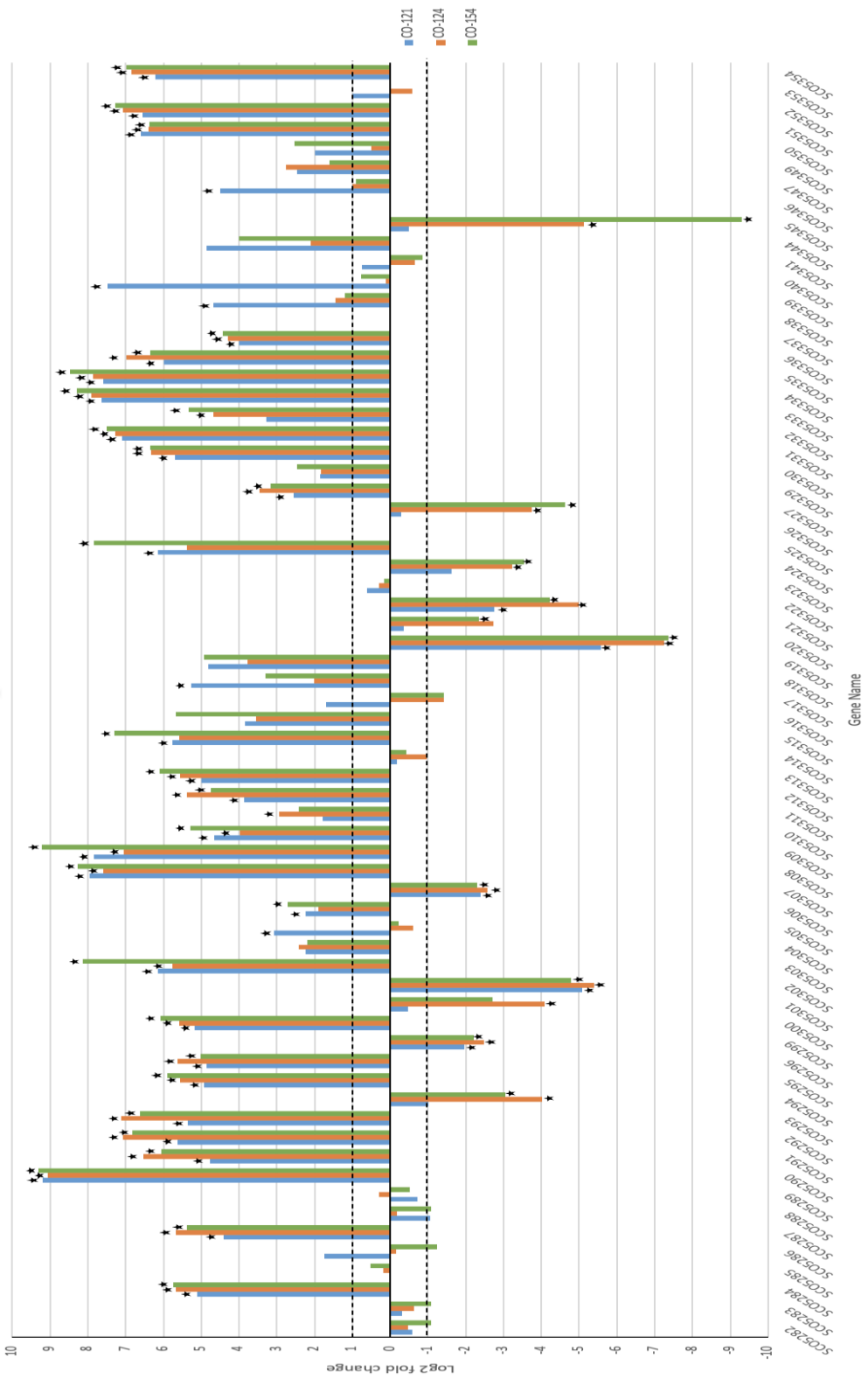


Figure A31: The log2 fold change of RNA transcription for the thirteenth antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to the empty vector control CO-250 grown in YEME media. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

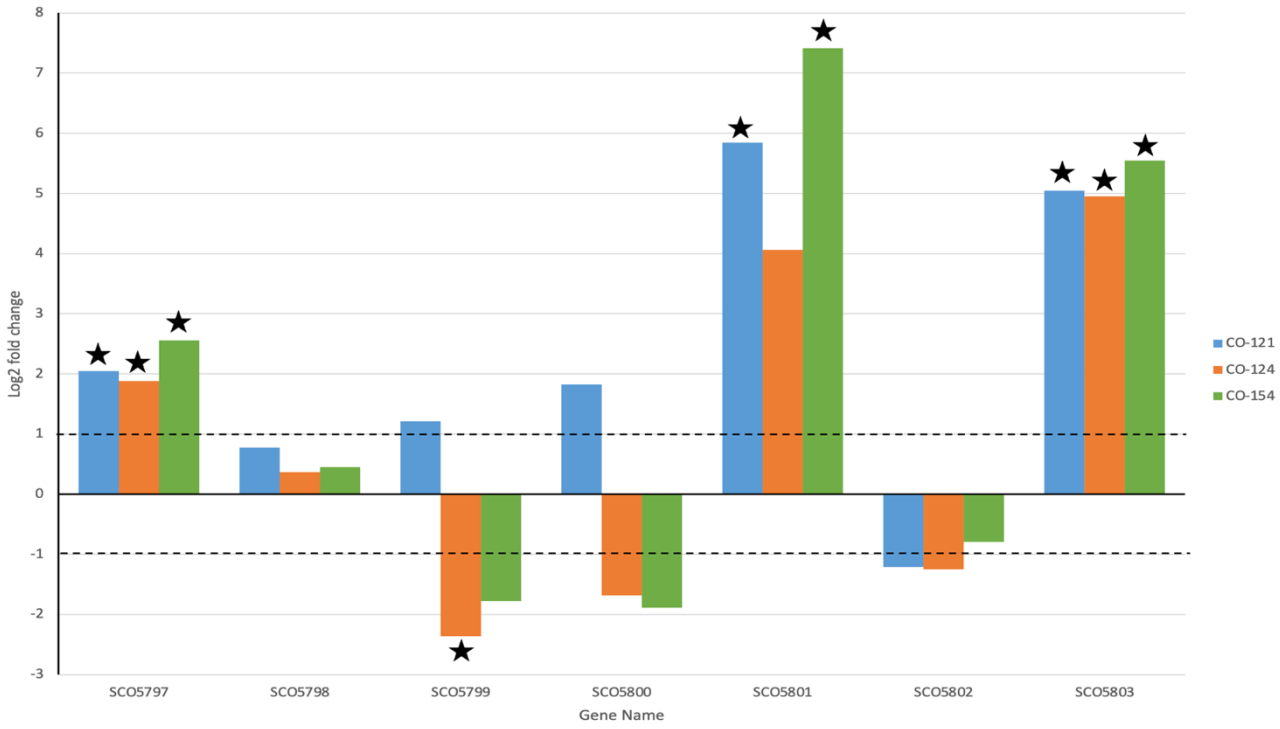


Figure A32: The log2 fold change of RNA transcription for the fourteenth antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to the empty vector control CO-250 grown in YEME media. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

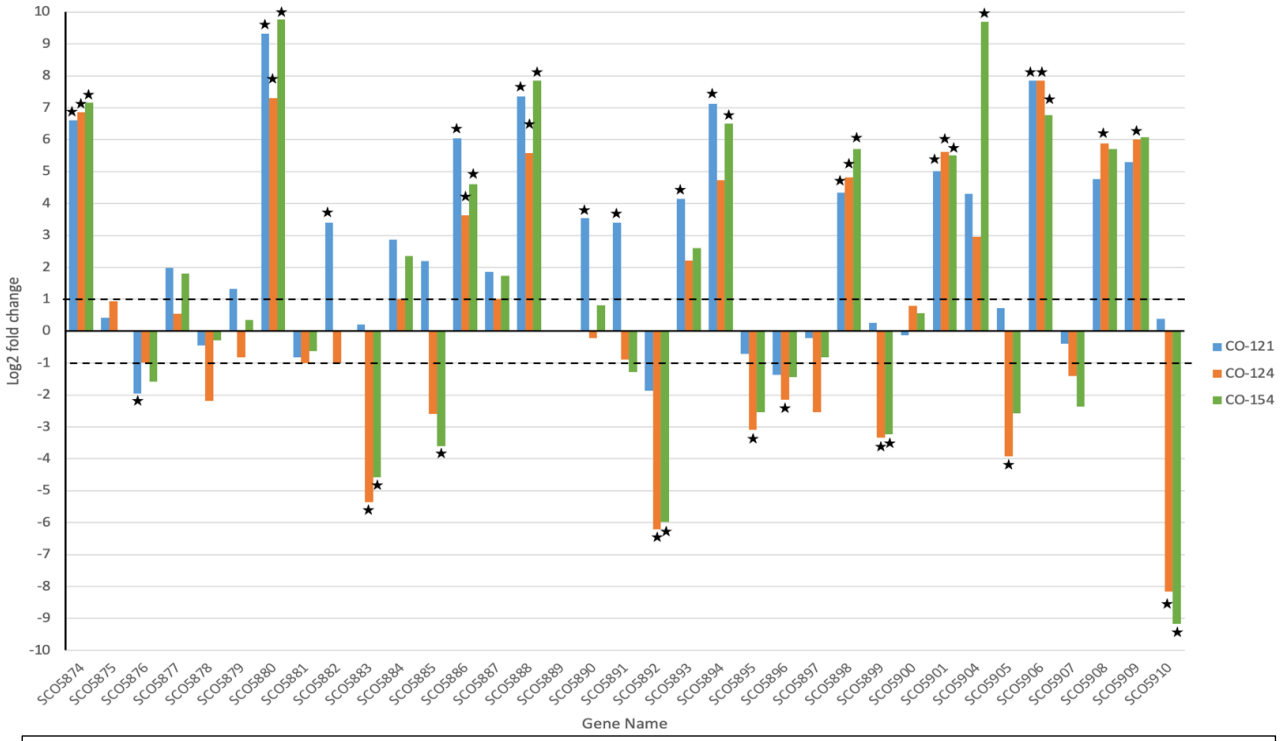


Figure A33: The log2 fold change of RNA transcription for the fifteenth antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to the empty vector control CO-250 grown in YEME media. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

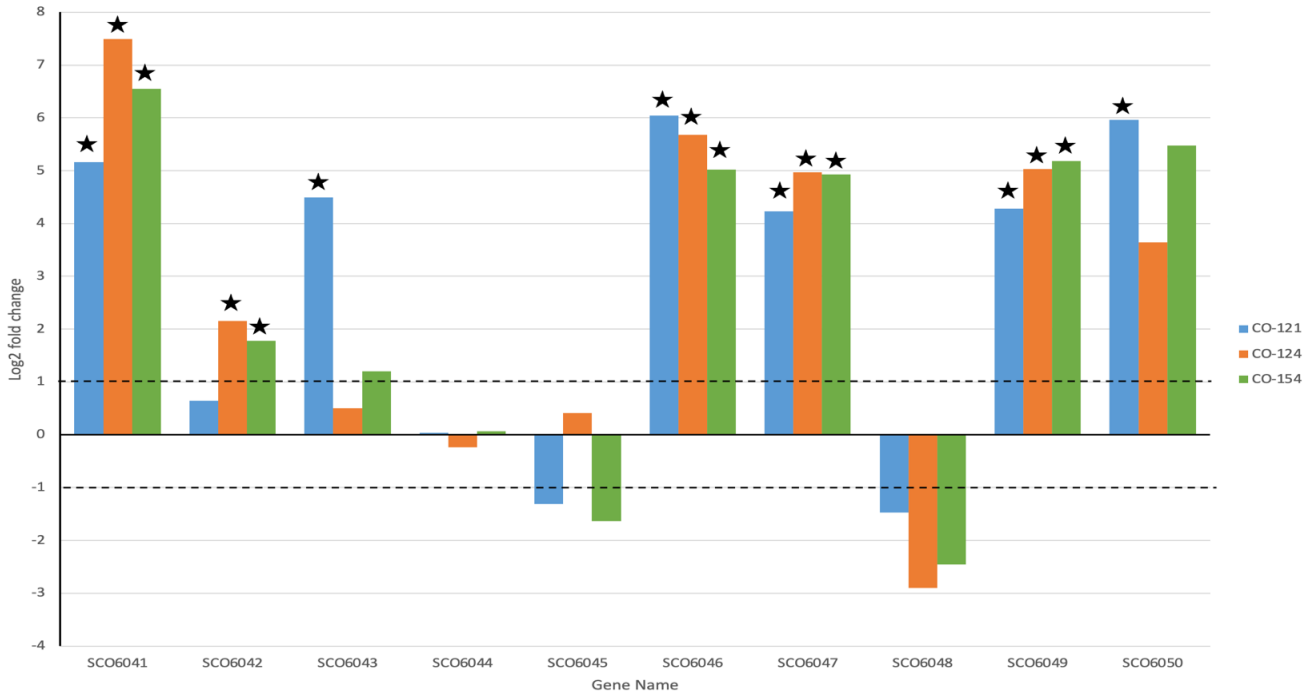


Figure A34: The log2 fold change of RNA transcription for the sixteenth antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to the empty vector control CO-250 grown in YEME media. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

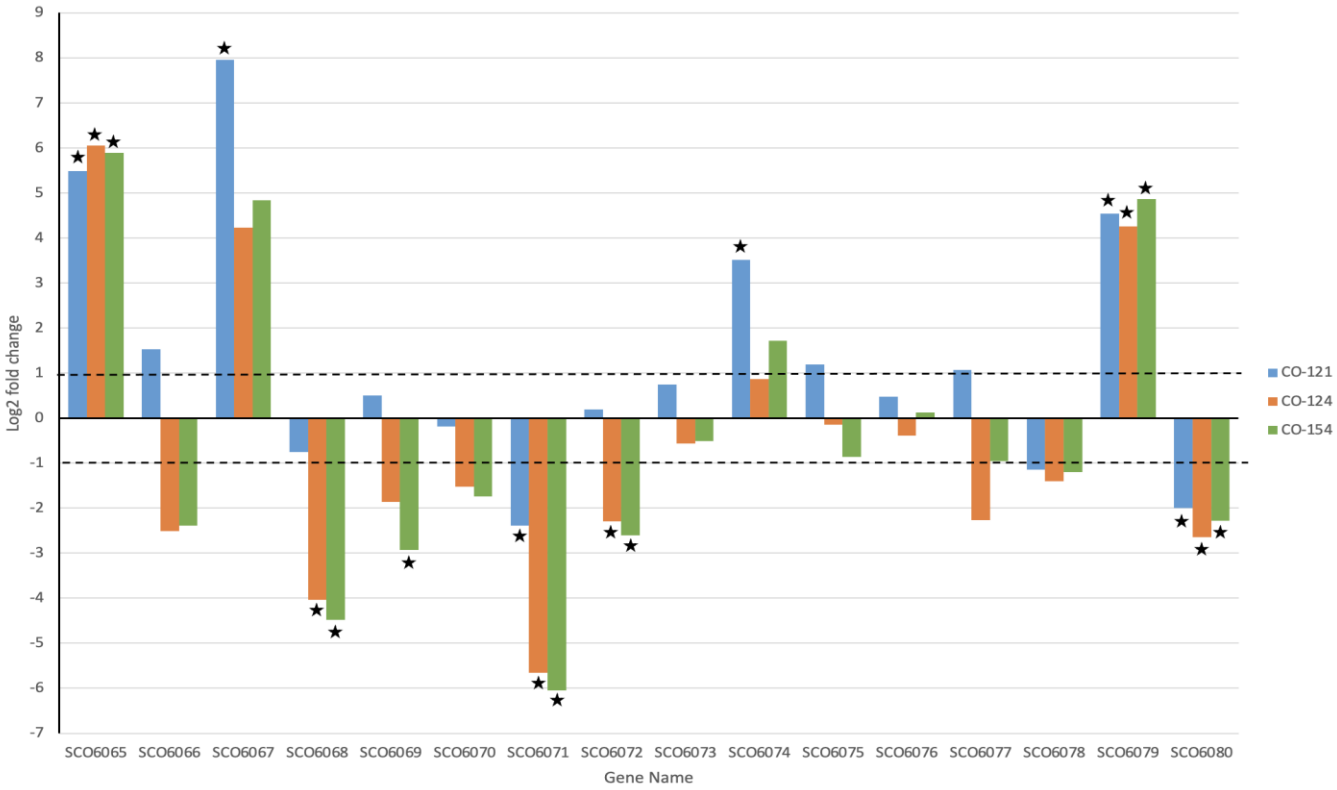


Figure A35: The log2 fold change of RNA transcription for the seventeenth antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to the empty vector control CO-250 grown in YEME media. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

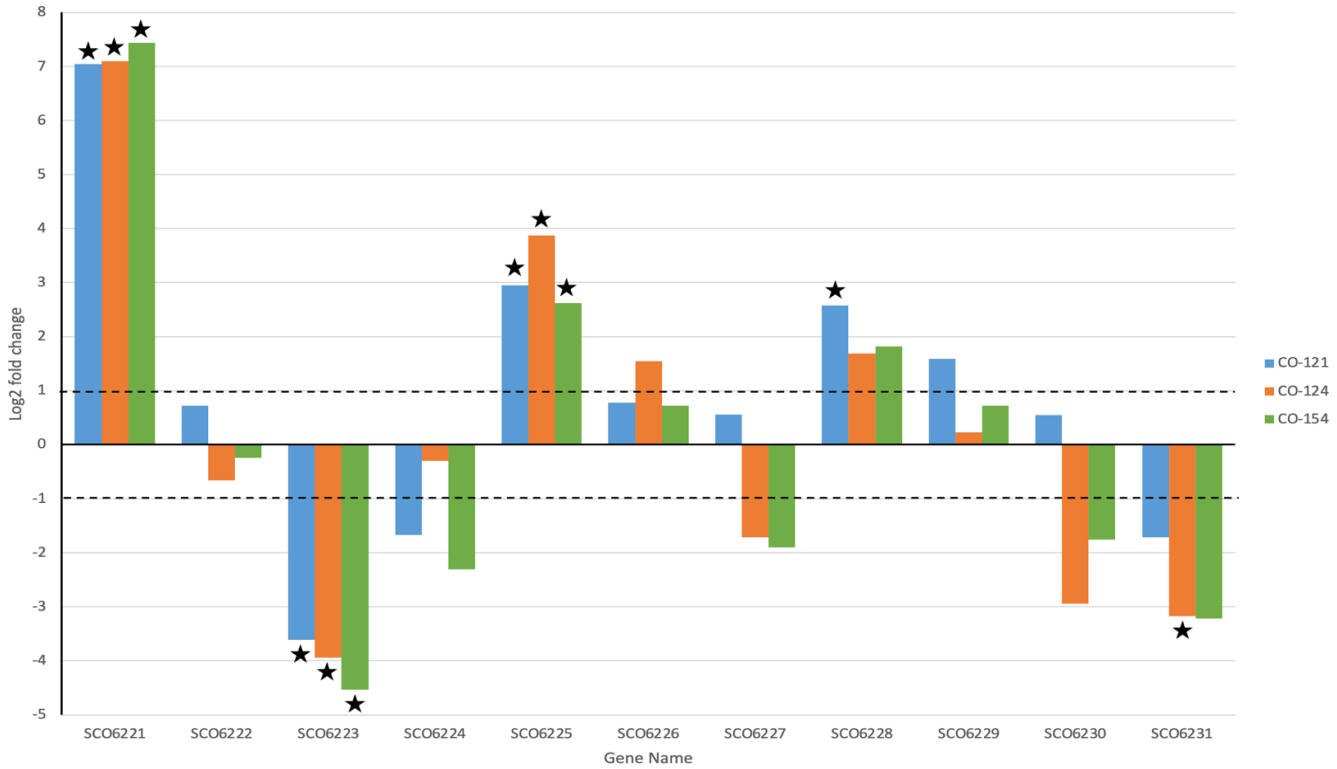


Figure A36: The log2 fold change of RNA transcription for the eighteenth antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to the empty vector control CO-250 grown in YEME media. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

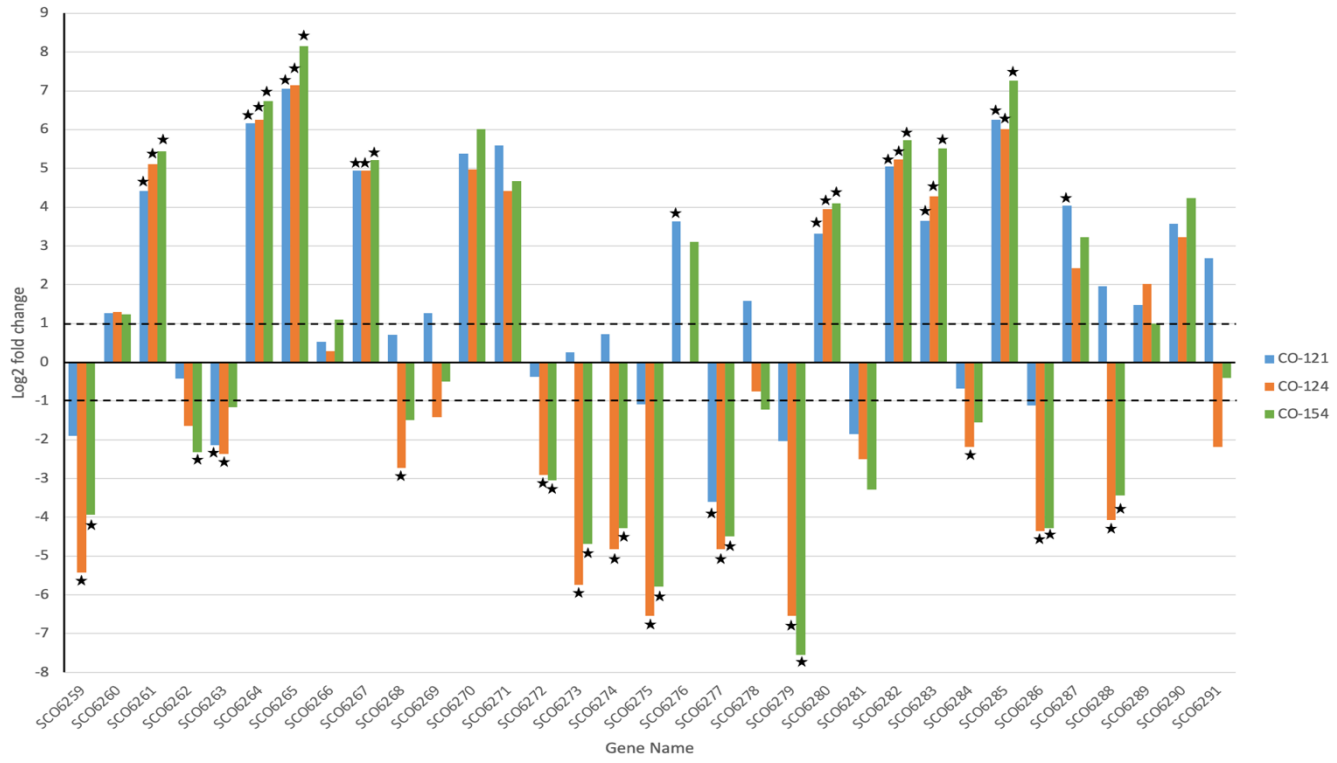


Figure A36: The log2 fold change of RNA transcription for the nineteenth antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to the empty vector control CO-250 grown in YEME media. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

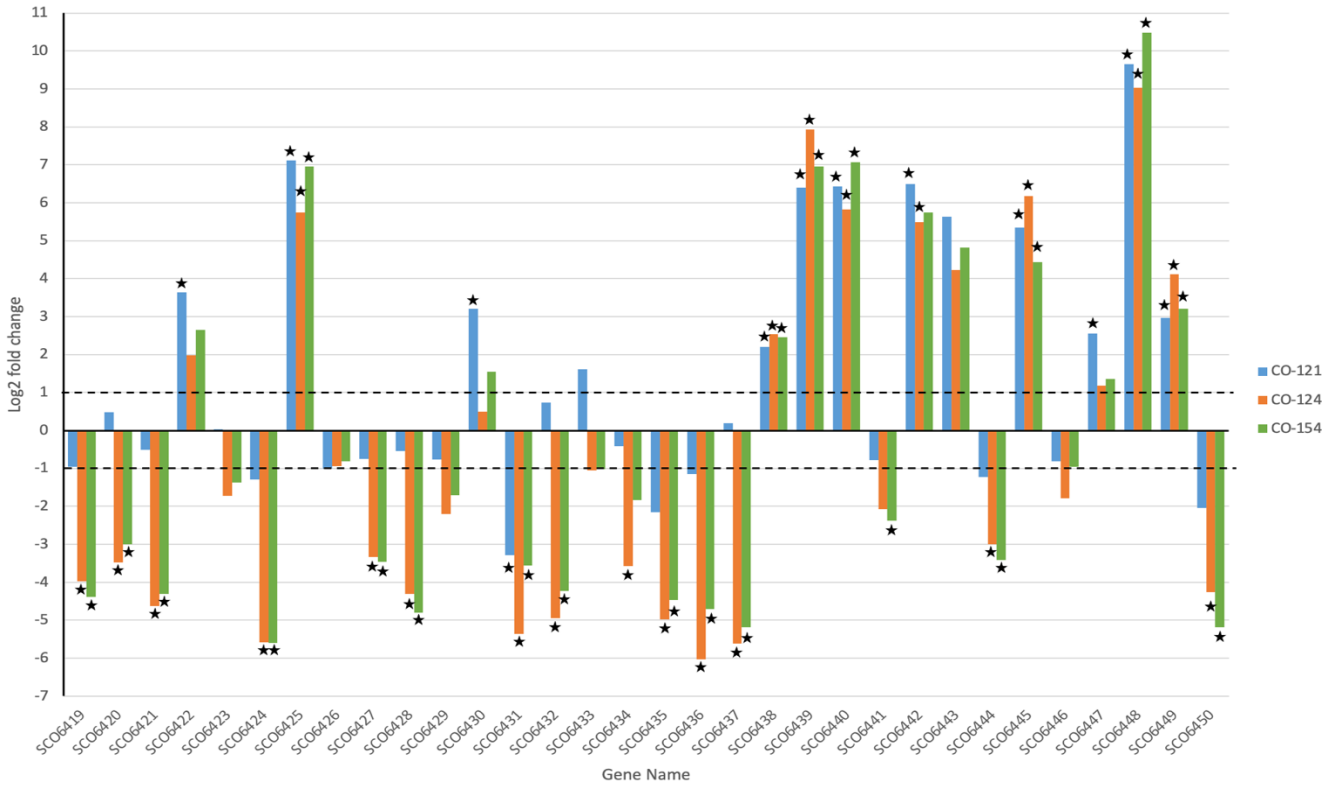


Figure A37: The log2 fold change of RNA transcription for the twentieth antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to the empty vector control CO-250 grown in YEME media. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

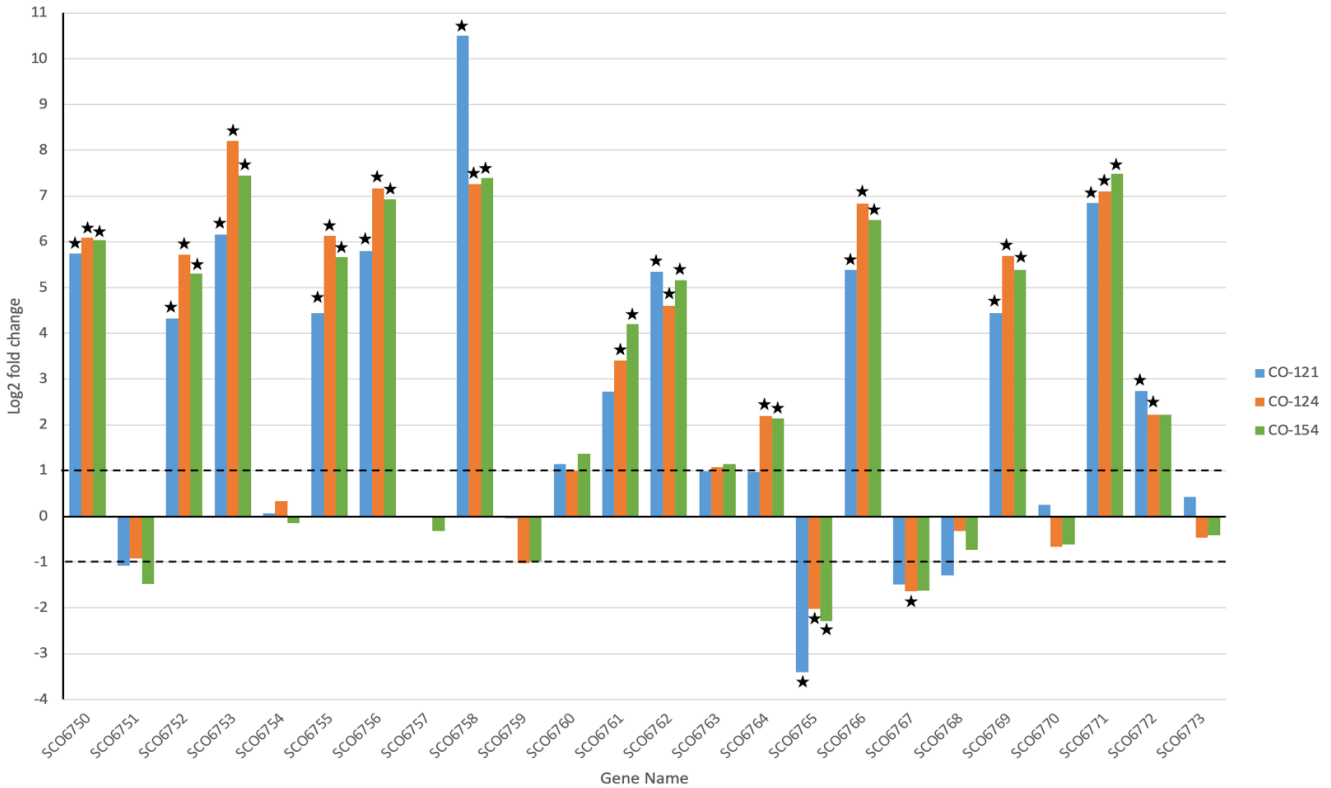


Figure A38: The log2 fold change of RNA transcription for the twenty-second antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to the empty vector control CO-250 grown in YEME media. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

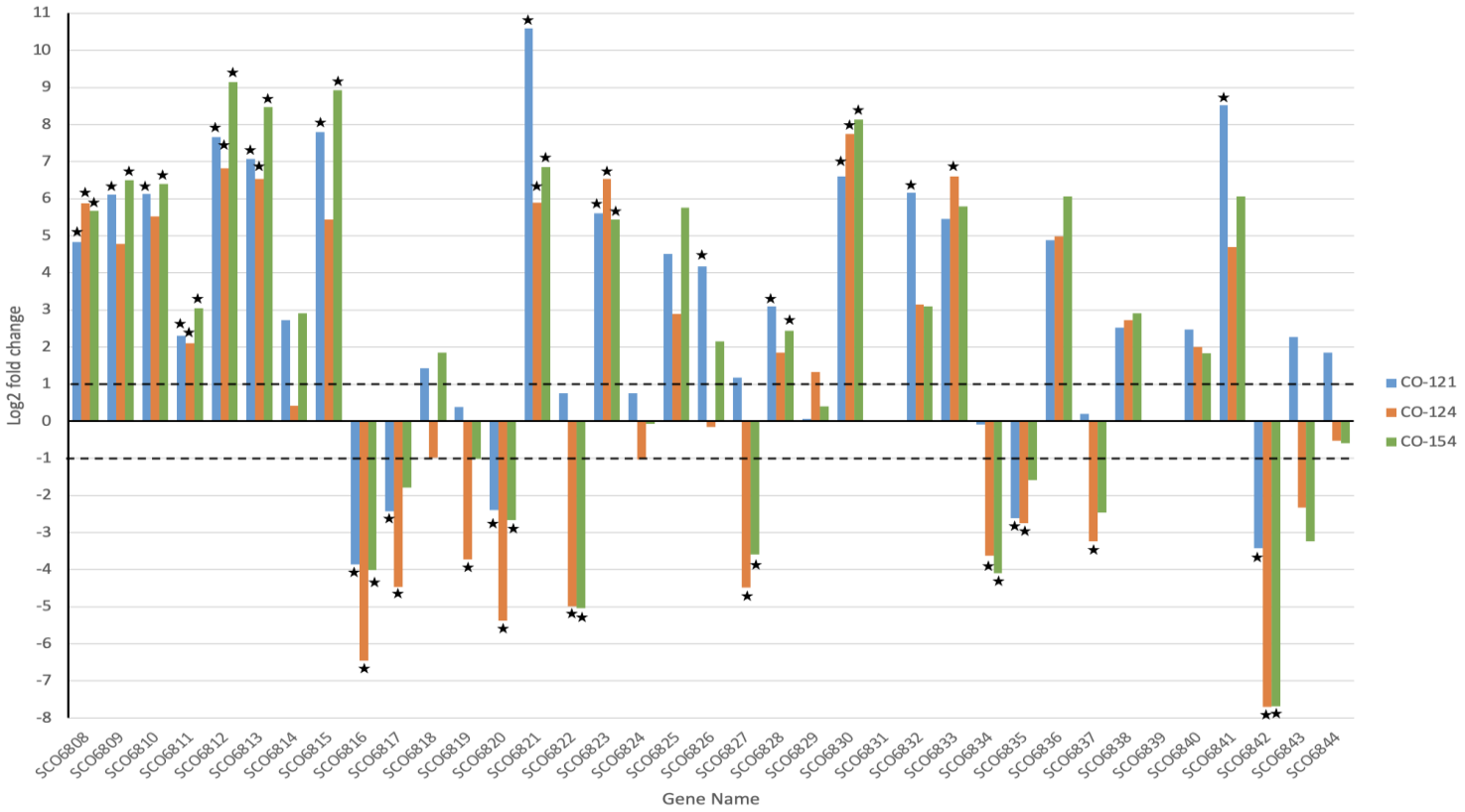


Figure A39: The log2 fold change of RNA transcription for the twenty-third antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to the empty vector control CO-250 grown in YEME media. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

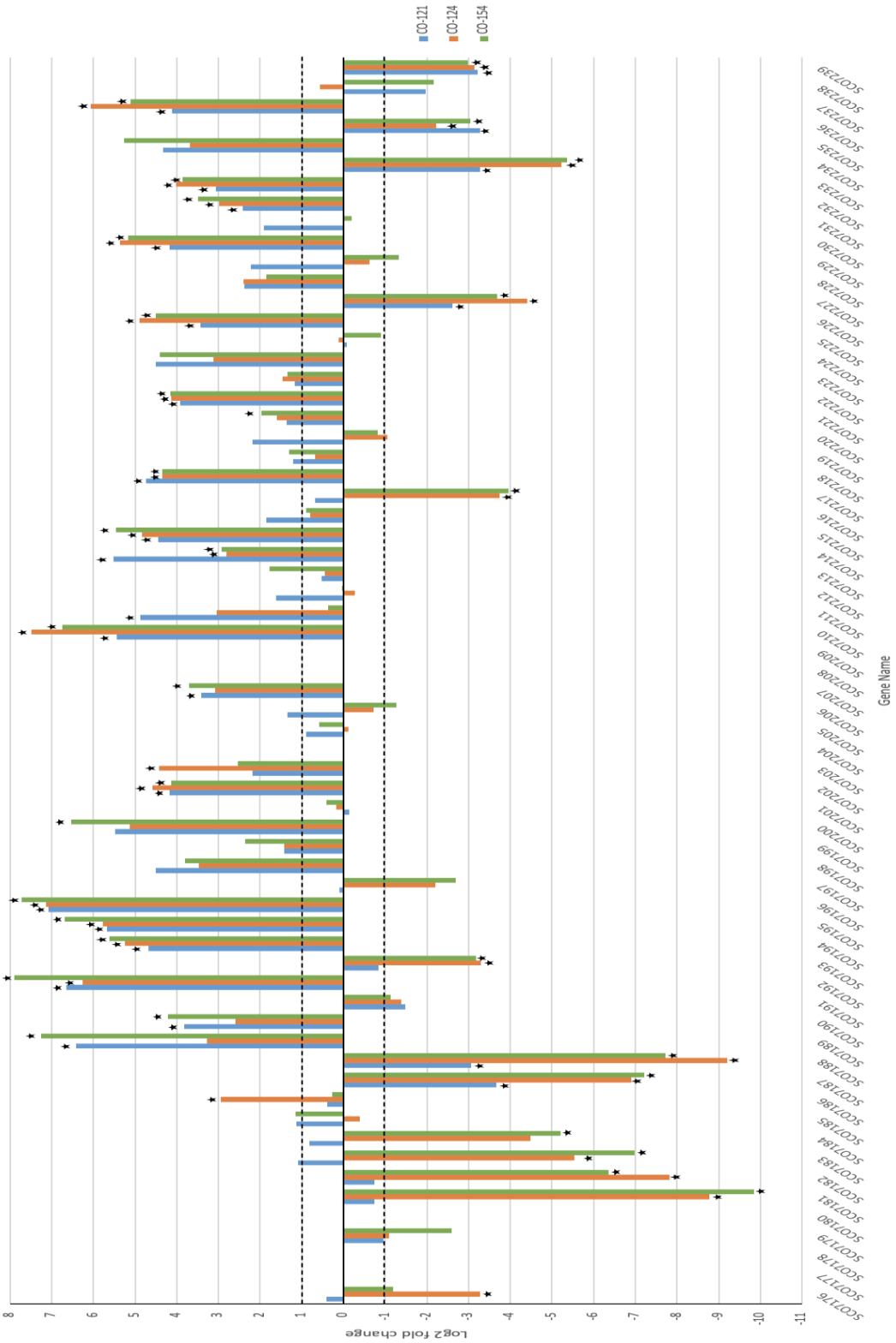


Figure A40: The log2 fold change of RNA transcription for the twenty-fifth antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to the empty vector control CO-250 grown in YEME media. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

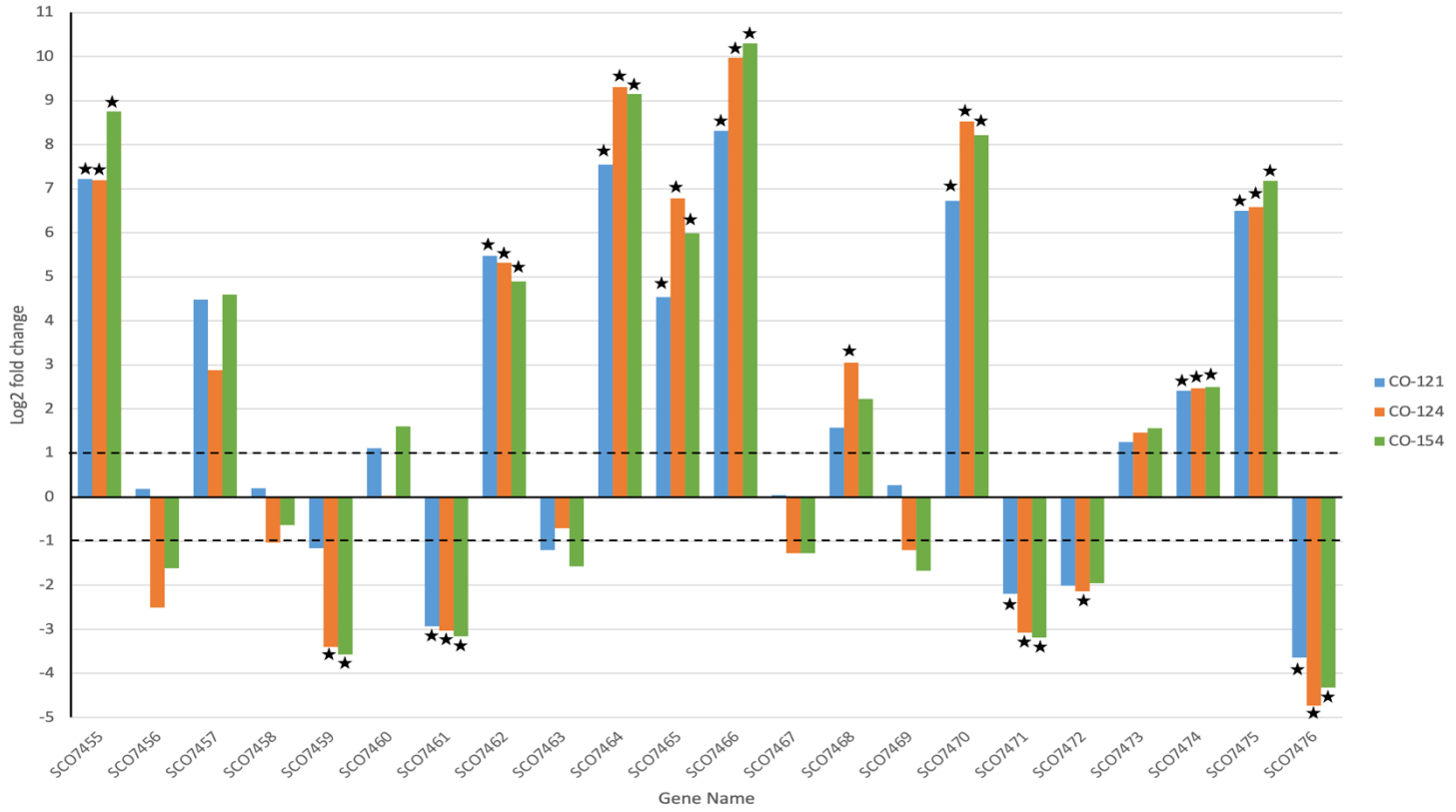


Figure A41: The log2 fold change of RNA transcription for the twenty-sixth antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to the empty vector control CO-250 grown in YEME media. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

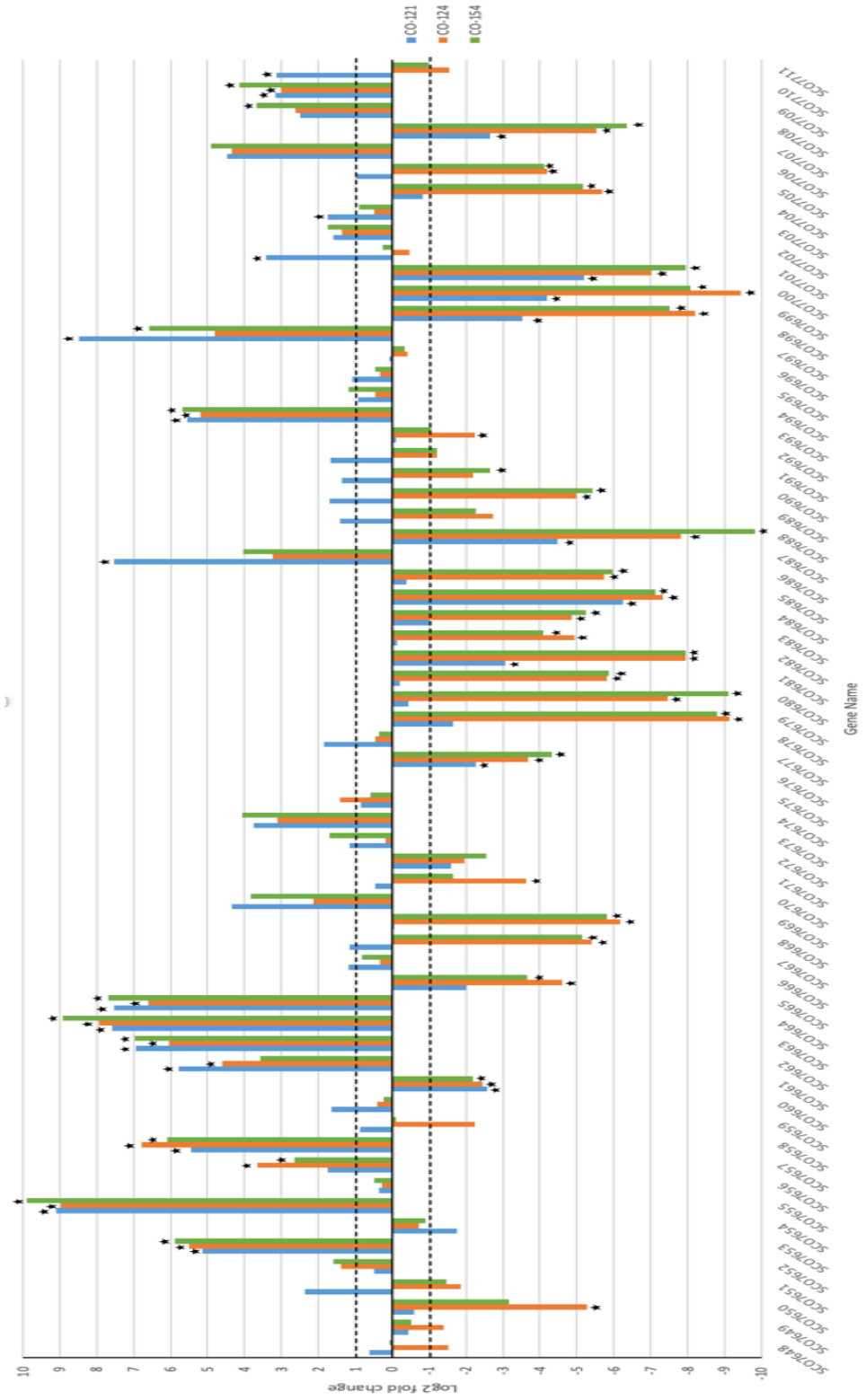


Figure A42: The log2 fold change of RNA transcription for the twenty-seventh antiSMASH predicted cluster in CO-121, CO-124 and CO-154 compared to the empty vector control CO-250 grown in YEME media. A log2 fold change greater than 1 or -1 indicates differential expression with the stars signifying significance due to a p value < 0.05

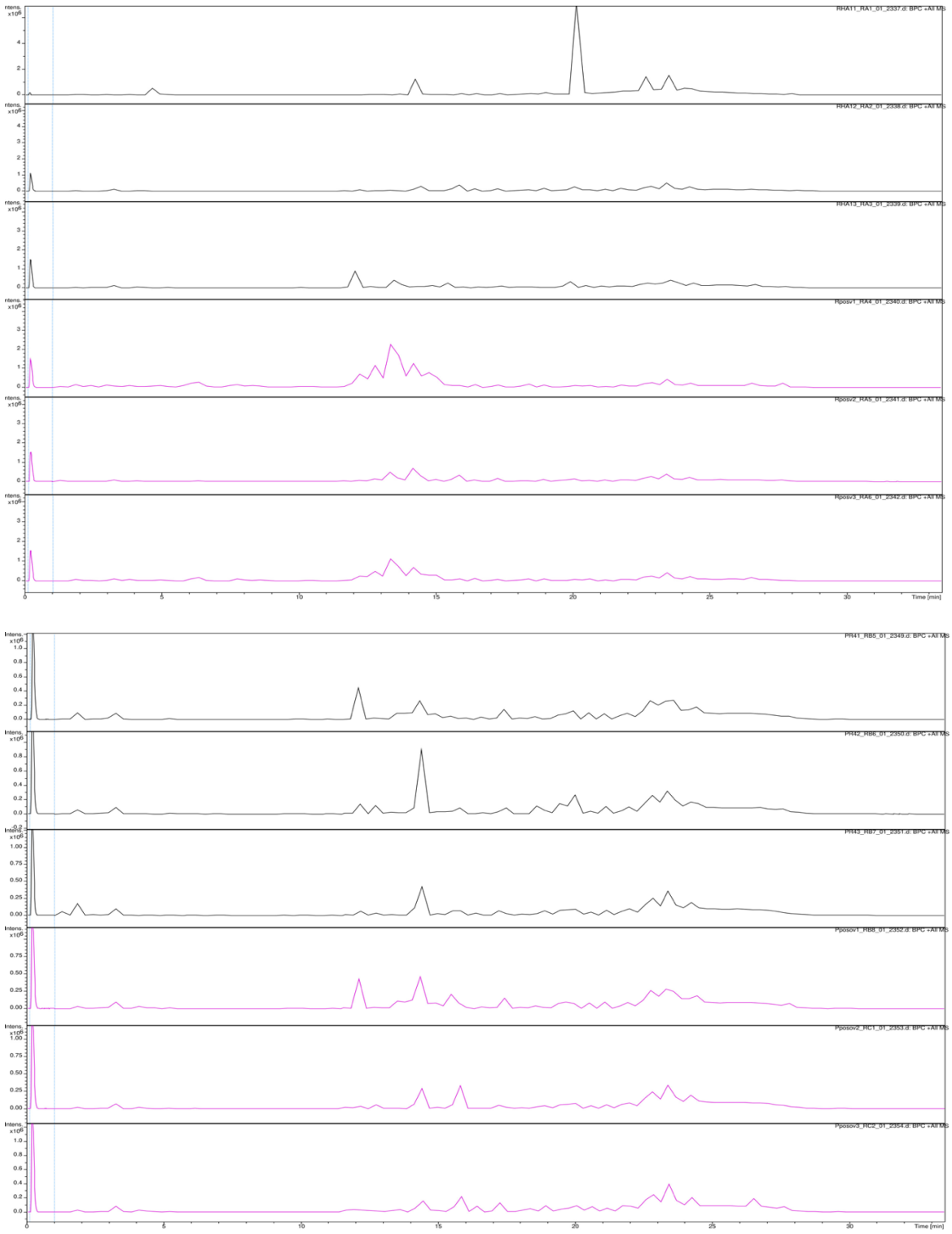


Figure A43: A) Base peak chromatogram of *R. jostii* RHA1 (black) v CO-213 (pink)
B) Base peak chromatogram of *R. erythropolis* PR4 (black) v CO-231 (pink)
The trace of the chromatograms in both mutants has no new peaks when compared against the wildtype bacteria.

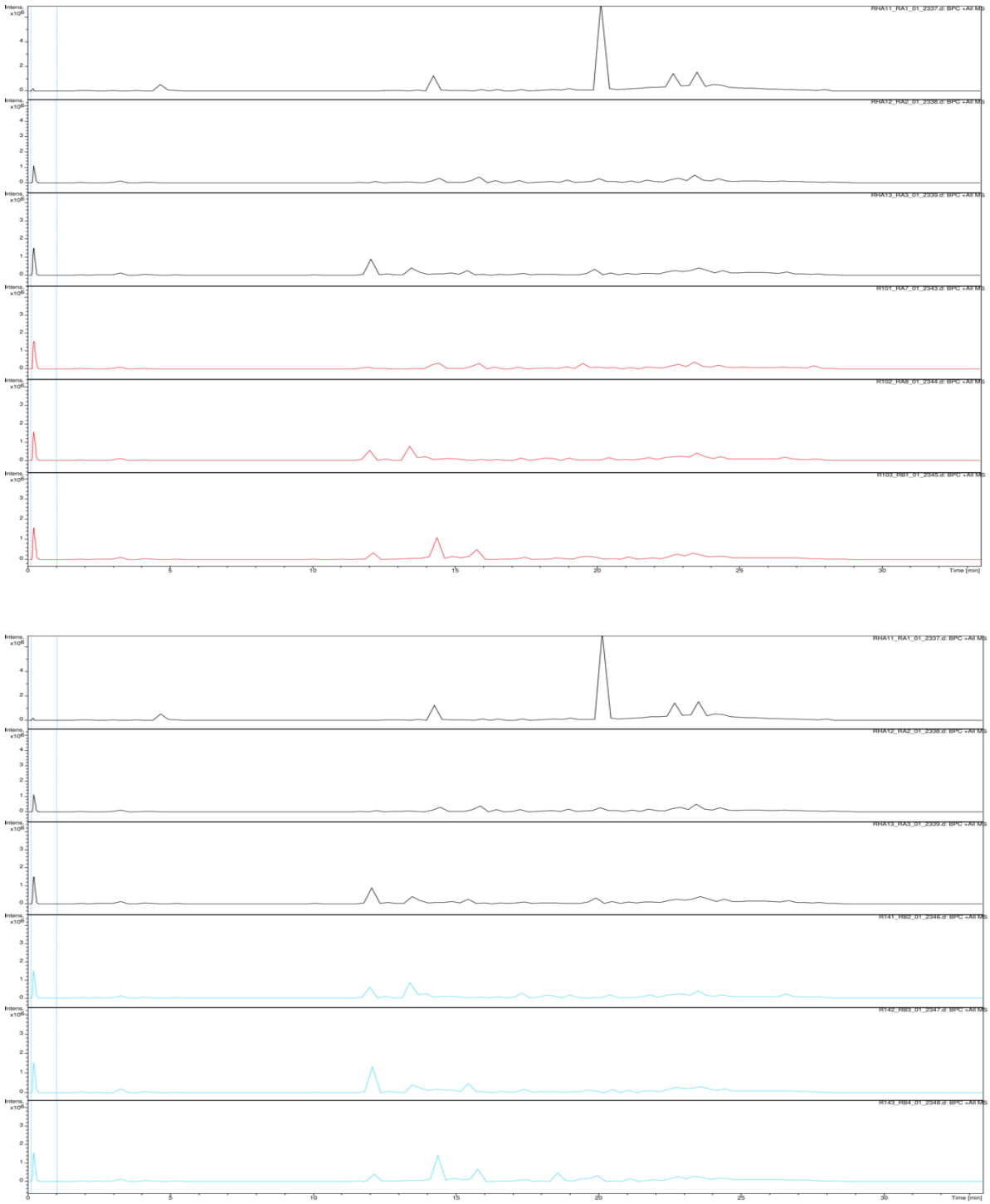


Figure A44: Base peak chromatogram of *R. jostii* RHA1 (black) v CO-218 (red) and CO-222 (blue). The trace of the chromatogram is similar across all replicates with no new peaks when compared against the wildtype bacteria.

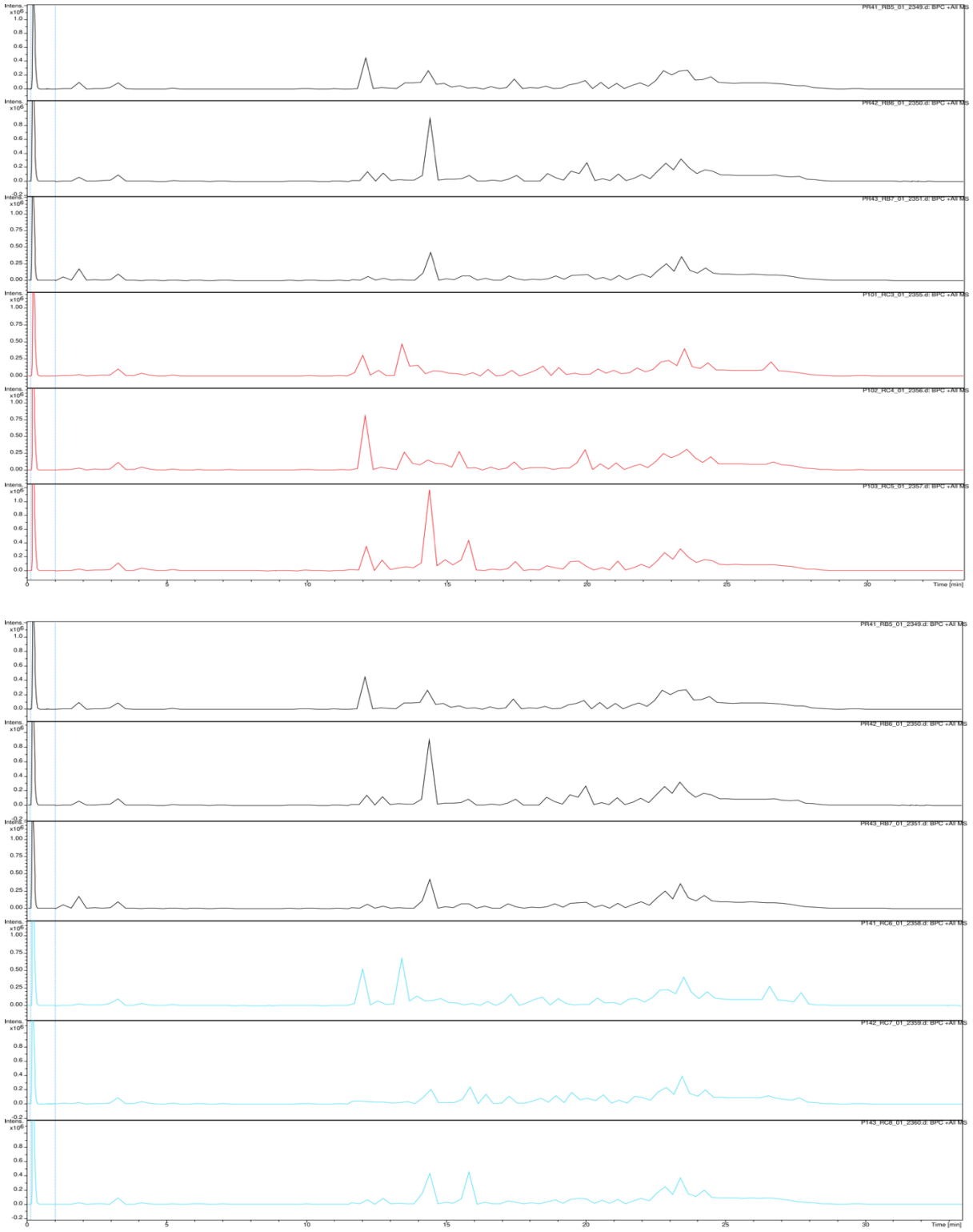


Figure A45: Base peak chromatogram of *R. erythropolis* PR4 (black) v CO-236 (red) and CO-240 (blue). The trace of the chromatogram is similar across all replicates with no new peaks when compared against the wildtype bacteria.