NTNU – Trondheim
Norwegian University of
Science and Technology

Mining, Analysis and Targeted Activation of Secondary Metabolite Gene Clusters in Streptomyces bambergiensis

Jon Petur Joelsson

Biotechnology
Submission date: May 2015
Supervisor: Sergey Zotchev, IBT
Co-supervisor: Olga Sekurova, IBT

Norwegian University of Science and Technology
Department of Biotechnology

Preface

I started this academic journey five years ago when I signed up for the biotechnology program at the University of Akureyri. I have not regretted that decision for one moment since. This journey brought me to Trondheim two years ago when I decided to study for my master's degree there. This has been an incredibly rewarding experience.

When I started to look for a topic for my master's thesis, I thought back to my classes in Akureyri and about how fascinating I found secondary metabolism in microorganisms. This led me to Professor Sergey Zotchev and Dr. Olga Sekurova. Professor Zotchev offered me the chance to work with *Streptomyces bambergiensis* and to genetically manipulate them into producing novel antibiotics. I was instantly fascinated.

This thesis is the result of countless hours in the lab, where I had many successes and failures. It is also the result of countless hours spent reading papers and getting more informed about the subject matter and finally writing everything down.

I would like to thank Professor Zotchev for his guidance in this work. He has been very generous with his time and has a way of making you really challenge yourself. I would like to thank Dr. Olga Sekurova, my lab supervisor. She made me feel welcome the first day I met her and it has been a genuine pleasure to work with and learn from them both.

I want to thank my mother, Þóra Ólafsdóttir who has always been my number one supporter. I would not have been able to take up my studies to begin with without the support of my father in-law, Sigurður Steingrímsson and mother in-law Kristjana Friðriksdóttir, who generously opened up their home for me and have supported me throughout, I wish to thank them especially.

My deepest thanks to Ásdís Sigurðardóttir, for her patience and encouragement. Also to my two girls, Freyja and Iðunn, my two biggest fans.

Finally I want to thank the staff and fellow master students at NTNU, for creating a sound work and study environment.

Summary

The discovery of new antibiotics is one of the most urgent tasks facing scientists today. The escalating rate of resistant microbes creates a constant need for new or improved drugs. With the recent progress in genome sequencing a new world is opening up as researchers can explore bacterial antibiotic biosynthesis gene clusters that are normally silent under laboratory conditions. Waking up these silent gene clusters harbours the potential of novel drug discoveries to fight microbial infections, cancer or other major life-threatening conditions. In this work, genetic potential of *Streptomyces bambergiensis*, known producer of antibacterial antibiotic moenomycin, was investigated. Whole genome sequence of *S. bambergiensis* showed that it has, in addition to moenomycin cluster, 28 secondary metabolite biosynthetic gene clusters, including a giant polyketide synthase (PKS) gene cluster composed of 30 genes and spanning over 190 kb. The product of this gene cluster is unknown and likely not produced under laboratory conditions. First, moenomycin gene cluster was inactivated in *S. bambergiensis* by deletion in order to eliminate antibacterial background activity. Next, constitutive expression of the regulatory gene, a LuxR family transcriptional regulator, from the PKS gene cluster was accomplished, leading to the production of a compound with inhibitory activity against *Bacillus subtilis*. Preliminary analysis suggested that this may be a new compound that can be further studied as a potential antibiotic.

Abbreviations

dH_2O	Distilled sterile water
MoeD	Moenomycin Deletion
WT	Wild Type
C1R4	Overexpressed regulator number 4 in cluster 1
C1R5	Overexpressed regulator number 5 in cluster 1
C28R	Overexpressed regulator in cluster 28
SM	Secondary Metabolite
BuOH	Butanol
MeOH	Methanol
NaOH	Sodium Hydroxide
Kan	Kanamycin
Nal	Nalidixic Acid
Thio	Thiostrepton
Amp	Ampicillin
Cml	Chloramphenicol
Am	Apramycin
GUS	β-Glucuronidase
PCR	Polymerase Chain Reaction
OriT	Origin or Transfer

Table of content

1 Introduction

1.1 Bacterial Secondary Metabolites

The metabolites not directly essential for the development, growth and reproduction of an organism are categorized as secondary metabolites (SMs). These SMs have a wide variety of both chemical structures and biological activities. (Vining, 1992)

Producers of SMs can be found among bacteria, fungi, plants and animals of both terrestrial and marine origin. There are many medical applications of SMs from micro organisms, including but not limited to antimicrobial, antiviral, tumor suppressing and immunosuppresant drugs. (Vaishnav & Demain, 2010)

It is suggested that fitness and propagation of an organism in a particular environment is linked to its production of SMs. Most SMs are defense mechanisms of the organism, although SMs also have roles in mediation of the growth, reproduction and differentiation as signaling molecules. (Vaishnav & Demain, 2010) The environment in which the organism evolves influenses the chemical structure and bioactivity of the SM, as survivability is highly increased in those organisms that evolve powerful defense mechanisms. (Vining, 1992) Many plants and bacteria form a symbiotic relationship, where the bacteria produces SMs which protect the plant host and the bacterial symbiont then gains nutrients from the host. (Challis & Hopwood, 2003)

Whatever the exact roles of SMs are in nature, the main attraction of SMs for science and industry are their biological activities. One of the most important SMs for humans are the antibiotics.

Alexander Fleming is a name most often associated with the advent of the antibiotic era. In 1928, while investigating properties of staphylococci, he serendipitously discovered that some of his cultures had become contaminated by a fungus and that colonies surrounding the fungus, *Penicillium rubens*, had been eliminated, while those further away seemed normal. This led to the discovery of penicillin and mass production of this antibiotic was established in 1945. (Fleming, 1929) The finding of the SM produced by this fungus became an inspiration for many novel antibiotic discoveries.

The 1940s and early 1950s have been named the “golden era” of antibiotics as most groups of important antibacterial and antifungal antibiotics were discovered back then. These include tetracyclines, cephalosporins, aminoglycosides and macrolides. Between the 1970s and 1990s, there was a noticeable decline in antibiotic discovery. Although novel compounds were still

being discovered, they were mostly analogues of existing compounds or too toxic to be used as human medicines. (Bérdy, 2005)

Still, the number of microbial SMs discovered since the 1940s has increased almost exponentially. By the end of 2002, the number of known bioactive SMs from microbes was over 22,000 (Figure 1).

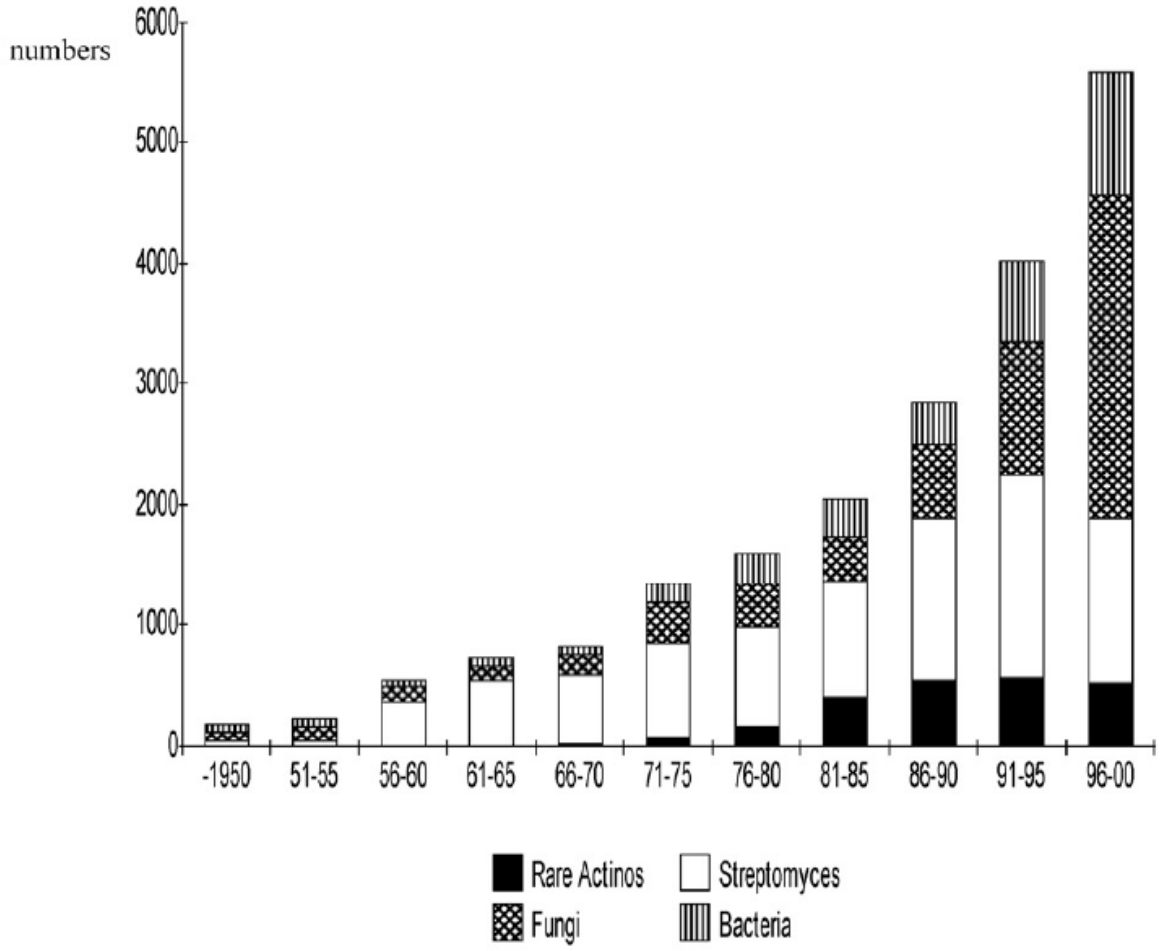


Figure 1: This figure shows the chronological distribution of discovered antibiotics from 1950 - 2000. (Bérdy, 2005)

However, the quantitative increase in antibiotic discovery did not mean qualitative results in terms of drugs on the market, as the antibiotic drug approvals from microorganisms has been on a rapid decline, especially in the last decades. Irresponsible and incorrect use of antibiotics has caused many pathogens to evolve resistance to almost all existing drugs. (Zotchev S. B., Genomics-Based Insights into the Evolution of Secondary Metabolite Biosynthesis in Actinomycete Bacteria, 2014) From 1983 to 1987, 16 new antibiotics were approved by the FDA, compared to only 2 being approved between 2008 and 2011. (Hughes, 2011) This is

largely due to frequent re-discovery of existing compounds and high cost of screening, characterization and development. (Zotchev, Sekurova, & Katz, 2012) This can be seen in the following graph (Figure 2) from the IDSA. (Infectious Diseases Society of America, 2013)

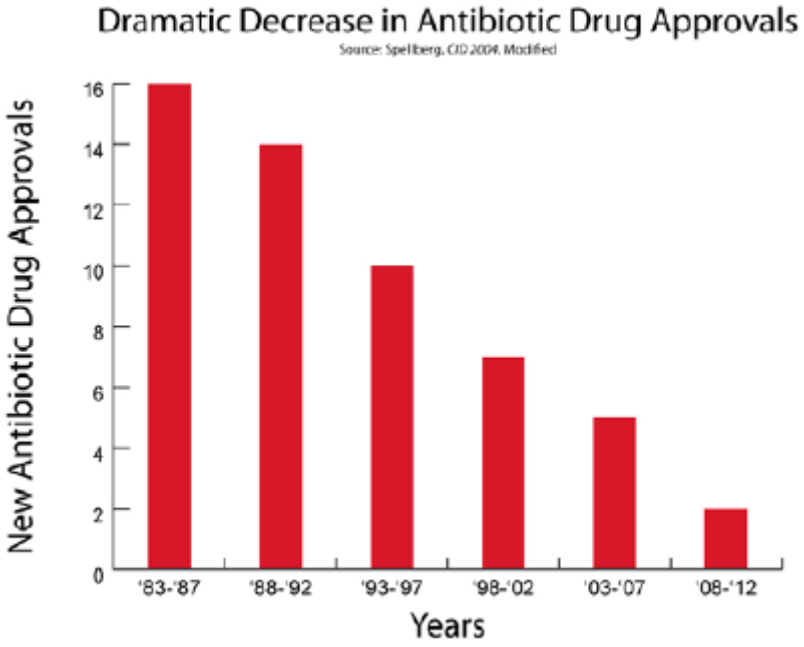


Figure 2: Chart showing the dramatic decrease in antibiotic drug approvals by the FDA.

Now in the genomic era, when sequencing of whole genomes is possible, a world of opportunities opens up as a multitude of uncharacterized SM biosynthetic pathways are revealed. It has been found that in microbial genomes that encode for SM biosynthetic pathways, the genes are very often clustered together in groups. These clusters usually also contain regulators and transport systems. This allows software, such as, antiSMASH to easily identify novel biosynthetic pathways. (Zotchev, Sekurova, & Katz, 2012) This has the potential to revolutionize the discovery of novel bioactive substances in microorganisms.

1.2 OSMAC (One Strain-Many Compounds)

Natural products are an important source of biologically active compounds and they continue to be of high interest to pharmaceutical companies. (Bertrand, et al., 2014) The search for novel natural products is one of the central subjects of industrial and academic research. OSMAC is an approach to focusing on the potential of each microorganism and elucidate the chemical diversity of natural products. (Bode, Bethe, Höfs, & Zeeck, 2002) Using the OSMAC approach, manipulation of the desired biosynthesis can be performed at different steps. The following figure (Figure 3) shows the possible ways to influence the biosynthesis of secondary metabolites.

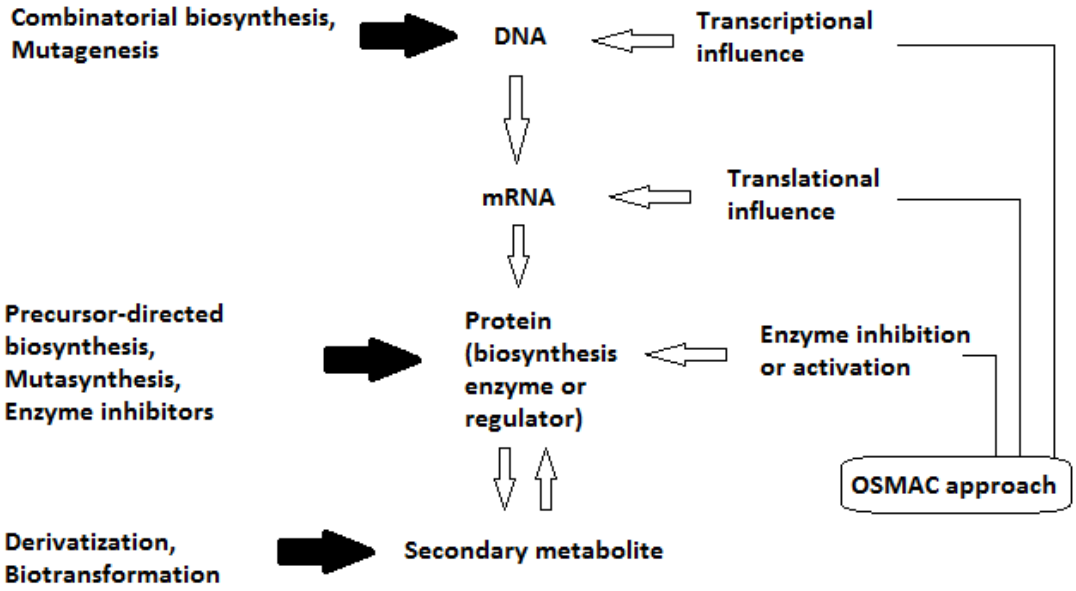


Figure 3: Possible ways to influence the biosynthesis of secondary metabolites. Using the OSMAC approach, manipulation of the biosynthetic pathway can occur at the transcriptional level, the translational level or the metabolite level. (Bode, Bethe, Höfs, & Zeeck, 2002) The main focus in this project is on the manipulation at the transcriptional level.

Manipulation of the biosynthetic pathway, using the OSMAC approach, can be performed at the translational, transcriptional and the metabolite level. The pathway towards the biosynthesis of SMs follows the usual metabolic pathway, that is, mRNA is transcribed from DNA and protein is translated from mRNA. (Bode, Bethe, Höfs, & Zeeck, 2002) The main focus of this project is the manipulation on the translational level.

As can be seen in the following figure (Figure 4), each microorganism contains a multitude of different SM gene clusters.

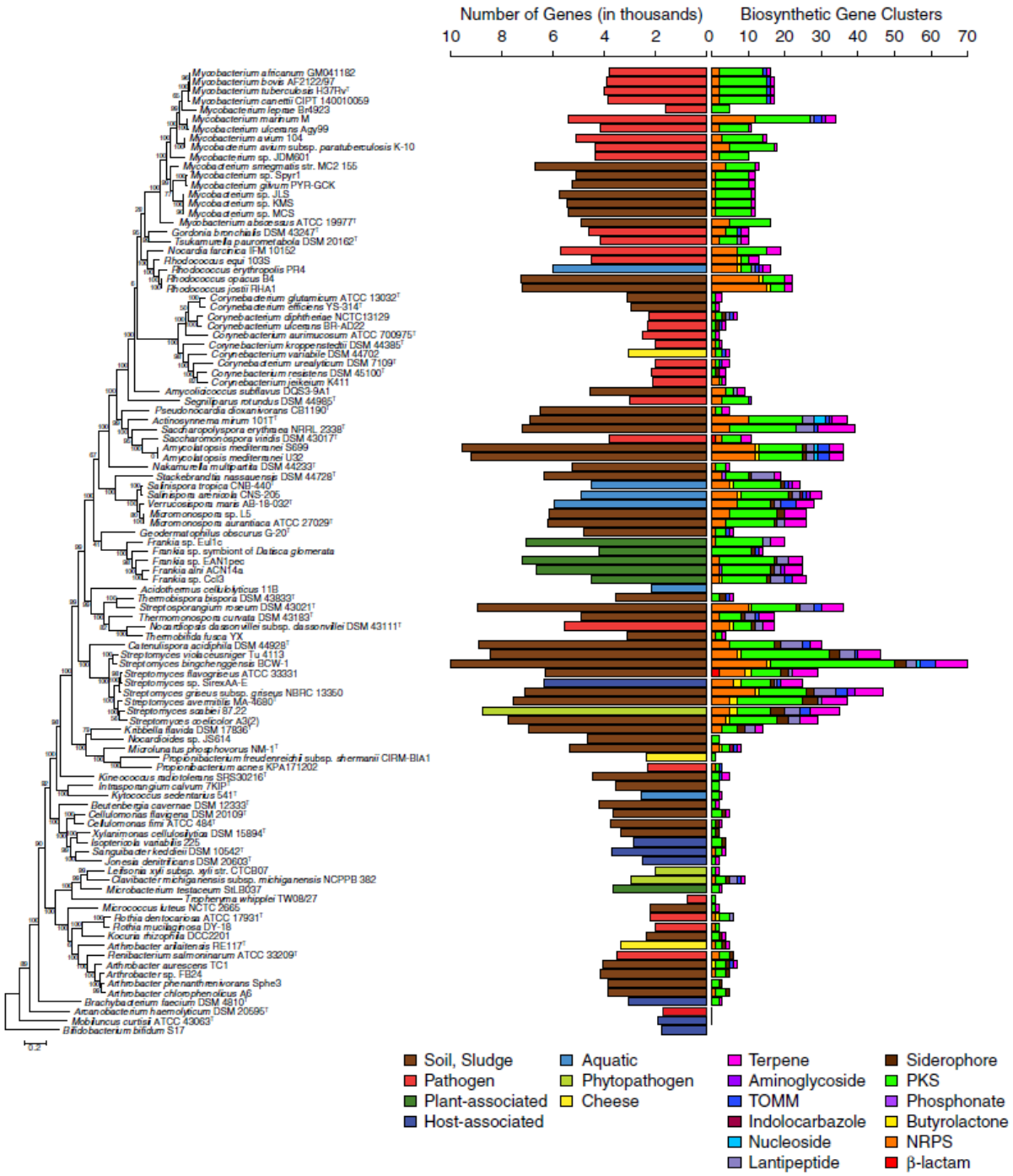


Figure 4: This phylogenetic tree shows genome size, isolation source and number of SM gene clusters of a 102 actinomycete genomes. The two bars plots are presented with species in the same order as the phylogenetic tree, representing genome size in thousands of genes on the left, coloured by habitat, and number of secondary metabolite gene clusters on the right, coloured by type of SM. (Doroghazi & Metcalf, 2013)

This shows the high potential of each microorganism to biosynthesize various natural products and illustrates the importance of the OSMAC initiative.

1.3 Actinomycetes as Producers of Secondary Metabolites

Actinomycetes are filamentous, gram-positive, aerobic, spore forming bacteria. They are known for their high G+C content, which can be over 70%. Actinomycetes derive their name from Greek, „aktis“ which means ray and „mykes“ which means fungus. This implies their resemblance to fungi as both actinomycetes and fungi form branched networks of hyphae. (Das, Lyla, & Khan, 2008) Actinomycetes are mostly terrestrial soil inhabitants, but are also found in various aquatic ecosystems.

Actinomycetes are well-known and versatile producers of bioactive SMs, including antibiotics, antitumor and immunosuppressive agents. Major actinomycete producers of SMs include species from genera *Streptomyces, Actinoplanes, Saccharopolyspora, Amycolatopsis* and *Micromonospora*. These are all producers of what are now commercially important biomolecules. (Solanki, Khanna, & Lal, 2008) The best known SMs produced by actinomycetes are arguably the antibiotics.

As can be seen in the following table (Table 1), microbes are important producers of bioactive compounds, with over 22,000 compounds discovered and of those, 17,000 have shown antibiotic activity.

Table 1: Over 22,000 bioactive compounds have been discovered from microbes and about 17,000 of those have shown antibiotic activity. (Bérdy, 2005)

Source	All known compounds	"Bioactives"	Antibiotics
Natural Products	over one million	200000 to 250000	25000 to 30000
Plant Kingdom	600000 to 700000	150000 to 200000	~25000
Microbes	over 50000	22000 to 23000	~17000
Algae, Lichens	3000 to 5000	1500 to 2000	~1000
Higher Plants	500000 to 600000	~100000	10000 to 12000
Animal Kingdom	300000 to 400000	50000 to 100000	~5000
Protozoa	several hundreds	100 to 200	~50
Invertebrates:	~100000	?	~500
Marine animals	20000 to 25000	7000 to 8000	3000 to 4000
Insects/worms/*etc.*	8000 to 10000	800 to 1000	150 to 200
Vertebrates (mammals, fishes, amphibians, *etc.*)	200000 to 250000	50000 to 70000	~1000

The following table (Table 2) shows the distribution of bioactive SM producers for the various microbial species.

Table 2: This table shows the distribution of bioactive SM producers for the various microbial species. Actinomycetes produce around 45% of the known microbial bioactive SMs and 75% of those actinomycetes were of the streptomyces species. (Bérdy, 2005)

Source	Bioactive secondary microbial metabolites				
	Antibiotics		Bioactive metabolites		Total bioactive metabolites
	Total	(with other activity)	no antibiotic activity	(antibiotics plus "other bioactives")	
Bacteria	**2900**	**(780)**	**900**	**(1680)**	**3800**
Eubacteriales	2170	(570)	580	(1150)	2750
Bacillus sp.	*795*	(*235*)	*65*	(*300*)	*860*
Pseudomonas sp.	*610*	(*185*)	*185*	(*370*)	*795*
Myxobacter	400	(130)	10	(140)	410
Cyanobacter	300	(80)	340	(420)	640
Actinomycetales	**8700**	**(2400)**	**1400**	**(3800)**	**10100**
Streptomyces sp.	6550	(1920)	1080	(3000)	7630
Rare actinos	2250	(580)	220	(800)	2470
Fungi	**4900**	**(2300)**	**3700**	**(6000)**	**8600**
Microscopic fungi	3770	(2070)	2680	(4750)	6450
Penicillium/Aspergillus	1000	(450)	950	(1400)	1950
Basidiomycetes	1050	(200)	950	(1150)	2000
Yeasts	105	(35)	35	(70)	140
Slime moulds	30	(5)	20	(25)	60
Total Microbial	**16500**	**(5500)**	**6000**	**(11500)**	**22500**
Protozoa	35	(10)	5	(45)	50

Actinomycetes are perhaps the most important producers of bioactive SMs. Around 45% of known microbial produced SMs are attributed to actinomycetes and 75% of those are of the *Streptomyces* species.

As *Streptomyces* are such active producers of SMs, they seem ideal for further studies utilizing the latest technology in molecular genetics to discover and activate cryptic biosynthetic pathways.

1.4 Secondary Metabolite Biosynthesis

Each SM produced is the result of a complex biosynthetic pathway and thus has its cost for the producing cells in terms of precursors and energy. This can be taxing for the microorganism, and suggests that production of SMs is limited to situations where a survival benefit can be achieved. Biosynthesis of SMs may be limited by the availability of primary metabolites, which most often constitute building blocks for SMs. Morover, primary and secondary metabolisms

share common cofactors (e.g.$NADPH_)$ and energy sources (e.g. ATP), making these two processes tightly interlinked (Malik, 1980)

For the synthesis of an SM, precursors from primary metabolism are utilized by the biosynthetic machinery to synthesize a molecular scaffold. This is followed by the modification of the scaffold by ”tailoring” enzymes, which add chemical groups to the scaffold, such as sugars, hydroxyl, methyl and amino groups. This results in a biologically active molecule and has the potential of self-toxicity for the organism that creates it. This is avoided by various resistance mechanisms implemented by the organism itself, and usually involves active efflux or modification of the molecule. (Zotchev S. B., Genomics-Based Insights into the Evolution of Secondary Metabolite Biosynthesis in Actinomycete Bacteria, 2014)

For the assembly of the scaffold and its modification, the organism must biosynthesize a number of enzymes in a particular order and stoichiometry. This is achieved by the clustering of the biosynthetic genes in the organism’s genome and the regulation of gene expression by pathway-specific regulatory genes. (Zotchev S. B., Antibiotics Biosynthesis, 2008)

Polyketides are a large class of secondary metabolites and includes the clinically relevant macrolide and tetracycline antibiotics. Polyketides show a remarkable functional and structural diversity. They are usually toxic for microorganisms and higher eukaryotes. (Jenke-Kodama, Sandmann, Muller, & Dittmann, 2005) Polyketides are synthesized by large enzymatic complexes called polyketide synthases (PKS). There is a large variety of different types of PKS enzymes with varying mechanistic complexities. The pathway to polyketide assembly is similar for most of them. PKSs are generally divided into three groups, the Type I, Type II and Type III. These are then further subdivided. (Zakeri & Lu, 2012)

In the following figure (Figure 5), an example of type I PKS assembly of the antibiotic, erythromycin is shown.

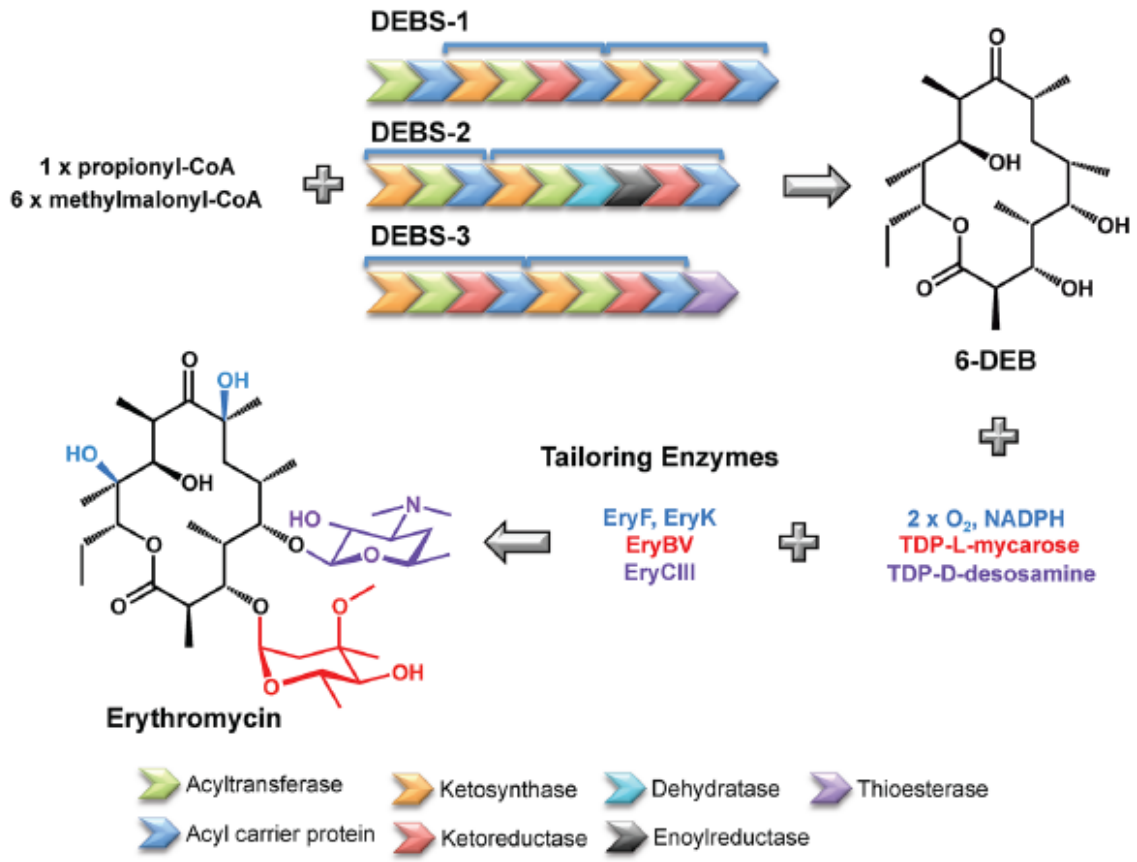


Figure 5: This figure shows Type I PKS assembly of the erythromycin molecule. Three mega enzyme complexes (DEB-1, DEB-2, and DEB-3) produce erythromycin through interaction with acyl-CoA building blocks. The mega enzyme complexes constitute 7 individual domains (blue brackets). The scaffold molecule 6-DEB is assembled. Tailoring enzymes incorporate hydroxyl and sugar moieties to produce erythromycin. (Zakeri & Lu, 2012)

Genome sequencing has revealed that each actinomycete bacterium has on average 20-30 gene clusters dedicated to the production of SMs and most of these are silent, that is, they are not expressed under laboratory conditions. (Doroghazi & Metcalf, 2013) The biosynthesis of an SM typically involves 10-30 genes, implying that this is a complicated process.

Biosynthetic clusters in *Streptomyces* generally contain pathway-specific regulatory genes that control the onset of biosynthesis. These regulators also often control genes involved in antibiotic export and self-resistance. Onset of antibiotic production in Streptomyces starts in the transition phase, which is a temporary flattening of the growth curve before exponential phase continues. Expression of regulatory genes starts in the late lag/early exponential phase. (McArthur & Bibb, 2007) This means that the expression of regulatory genes begins earlier than the onset of the antibiotic production itself. Constitutive expression of regulatory genes may lead to premature antibiotic biosynthesis and overexpression of regulatory genes can result in marked increases in the level of antibiotic production. (Bibb & Hesketh, 2009)

In *Streptomyces,* two types of regulatory mechanisms have been investigated. The first one is the pathway-specific or pyramidal cascade of regulation. The best known of the pyramidal systems controls the biosynthesis of streptomycin, tylosin and cephamycic/clavulanic acid. (Martin & Liras, 2010) Small signalling molecules called γ-butarylactones, produced by *Streptomyces* species, bind to specific cytoplasmic receptor proteins, many of which act as repressors. The binding of the signalling molecule inhibits the receptor protein from binding to specific DNA targets and thus induces the expression of the target genes. Each receptor is highly specific for its cognate γ-butarylactone signal molecule. (Takano, 2006) Pyramidal cascades of regulation are pathway-specific and usually trigger the formation of SARP (*Streptomyces* antibiotics regulatory proteins) pathway family of regulatory proteins, most of which act as positive regulators of antibiotic biosynthesis. (Martin & Liras, 2010)

The second regulatory mechanism is the global regulators, which are pleiotropic regulators, affecting several biosynthetic pathways. These are ususally two-component systems which respond to a variety of external stress signals. (Martin & Liras, 2010) Global regulatory genes are normally not found within SM gene clusters, but overexpression of global regulatory genes can result in the stimulation of SM production. (Aigle & Corre, 2012)

The three types of pathway-specific regulatory proteins emphasized in this project are the positive regulators of the SARP family and the LAL (large ATP-binding regulators of the LuxR family) family and the negative regulators of the TetR-like family.

SARP regulators contain a winged helix-turn-helix motif towards their N-termini to bind to the promoter region of the target gene. They recognize heptameric repeats within the promoter regions of the genes that they regulate. They have been found associated with secondary metabolic gene clusters that encode ribosomally and non-ribosomally synthezised peptides, aromatic polyketides and type I and type II polyketides. SARP family of proteins has only been found within actinomycetes, and most within streptomycetes. (Bibb M. J., 2005) Most characterized SARPs act as specific regulators at the bottom of regulatory cascades and turn on the expression of the biosynthetic genes in the given cluster. (Aigle & Corre, 2012)

LuxR family of transcriptional regulators are well know as key players in Quorum Sensing, coordinating the expression of a variety of genes, including genes encoding for virulence factors, biofilm formation and antibiotics biosynthesis. (Chen & Xie, 2011) LuxR family of regulators have a C-terminal helix-turn-helix motif. In actinomycetes, the best known members of the LuxR family are the LAL family. Unlike the SARP regulators, that have only been

identified in actinomycetes, the LAL regulators were first identified in protebacteria. (Schrijver & Mot, 1999) Several regulators of the LAL family have been identified in antibiotic and other SM gene clusters from actinomycetes. They are considered pathway-specific regulators. (Guerra, et al., 2012) Although knowledge regarding LAL family of regulators involved in biosynthesis of SMs in *Streptomyces* is limited, Huang et al reported enhanced antibiotic (rapamycin) production in *Streptomyces hygroscopius*, by over-expression of a pathway specific regulator of the LAL family. (Huang, et al., 2011) Also, Laureti et al showed that the constitutive overexpression of a regulator gene, encoding a protein that is similar to LAL family proteins, of a gene cluster in *Streptomyces ambofaciens* resulted in the production of a novel macrolide with antitumor activity. (Laureti, et al., 2011) (Zotchev, Sekurova, & Katz, 2012)

TetR-like family of repressors have a helix-turn-helix motif. TetR regulator proteins are involved in the transcriptional control of multidrug efflux pumps, pathways for the biosynthesis of antibiotics, response to osmotic stress and toxic chemicals, control of catabolic pathways, differentiation processess, and pathogenicity. Usually TetR binds to target operator and represses *tetA* transcription and is released in the presence of tetracycline, but the pathway can be more complex. (Ramos, et al., 2005) γ-butarylactone receptor proteins are TetR-like repressors. Gottelt et al showed that the deletion of ScbR, which is a γ-butarylactone receptor protein of *S. coelicolor* that binds to specific DNA sequences in front of its target genes, repressing their transcription, resulted in the expression of a silent gene cluster, which yielded a previously undetected compound with antibacterial activity. (Gottelt, Kol, Gomez-Escribano, Bibb, & Takano, 2010)

1.5 Genome Based Mining for Novel Secondary Metabolites

Now in the age of genome sequencing, one can obtain almost whole genomes of organisms in a matter of weeks. This opens up incredible oportunities, but also requires a considerable effort in data handling. (Zotchev, Sekurova, & Katz, 2012) The development of the software antiSMASH has helped in this regard. AntiSMASH is able to accurately identify gene clusters encoding SMs of nearly all known chemical classes. (Blin, et al., 2013) AntiSMASH provides links to the online software BLAST for the identification of proteins encoded by genes in a given cluster. Using this, an annotation of the gene clusters and prediction of their products can be achieved.

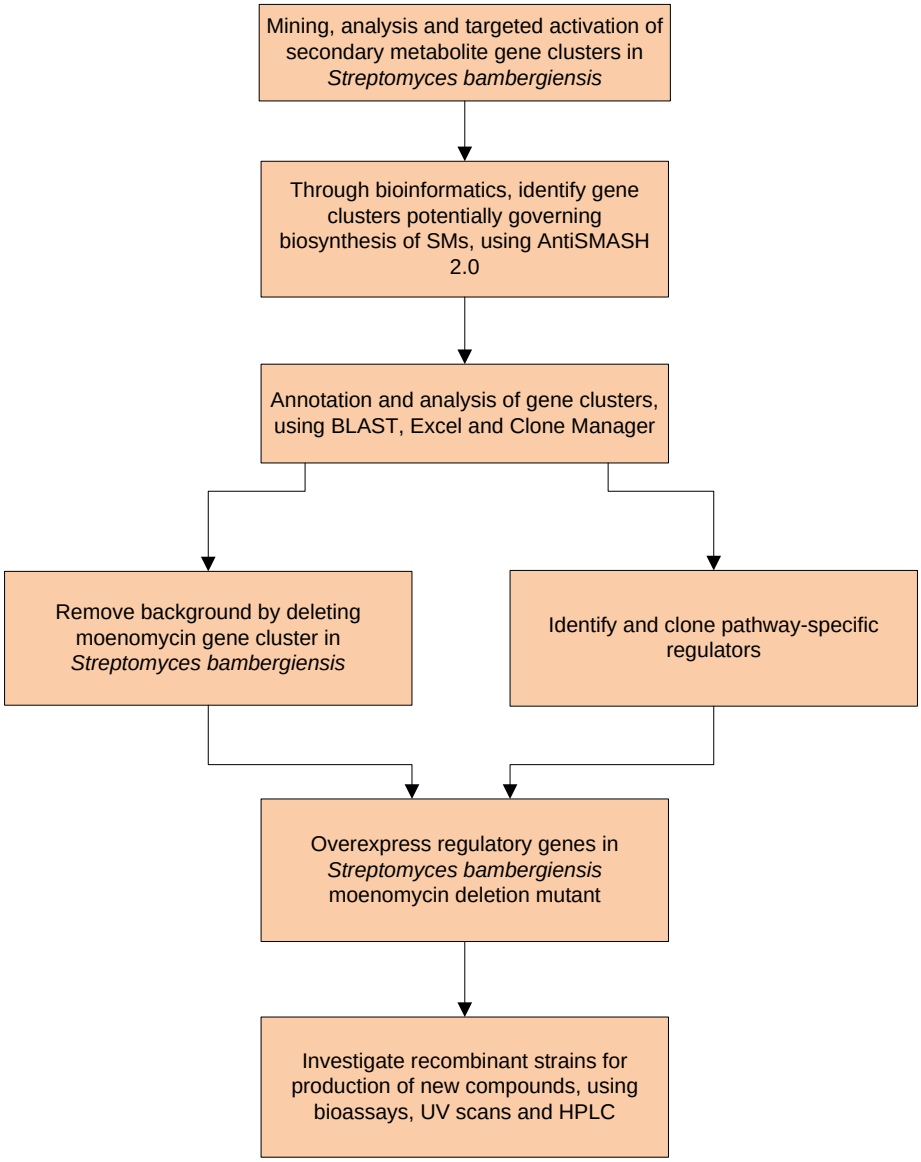
1.6 Project Background

Streptomyces are known producers of a wide variety of biologically active SMs. Recent sequencing of several *Streptomyces* genomes revealed many unknown gene clusters for biosynthesis of potentially novel biologically active SMs, which have not been detected in a conventional screening. Most likely these gene clusters are not expressed under laboratory conditions, and need to be activated to achieve production of novel compounds. However, it is not fesible to activate all gene clusters present in *Streptomyces* bacterium, which is on average between 20-30 per genome. (Doroghazi & Metcalf, 2013) A careful bioinformatic analysis of the gene clusters can be helpful in selection of the most interesting ones to be targeted for activation.

Almost all SM gene clusters contain regulatory genes. These genes control the expression of the biosynthetic genes. These regulator genes can be manipulated in order to coax the organism into producing more of the SMs. Inactivation of a repressor gene or overexpression of a positive regulator are examples of methods currently being used to manipulate organisms into increasing their production of SMs. *S. bamgergiensis* ATCC 13879 is a producer of the antibiotic moenomycin (flavomycin) complex (Zotchev, Rozenfeld, & Zhdanov, Transformation of streptomyces bambergiensis S712 protoplasts with plasmid DNA, 1988). This antibiotic is a phosphoglycolipid antibiotic inhibiting peptidoglycan formation in Gram-positive bacteria, and no other SMs have so far been reported from *S. bambergiensis*. (Ostash & Walker, 2010) Using a recently obtained complete genome sequence of *S. bambergiensis*, 29 SM biosynthesis gene clusters have been identified in this bacterium using the online software antiSMASH.

1.7 Project Goals

A flowchart of the project can be seen on the next page.



The major goal of the project was to attempt a discovery of novel SM through genome mining of *S. bambergiensis*.

The sub-goals were:

- To identify, using bioinformatics, biosynthetic gene clusters for potentially novel SMs
- To inactivate moenomycin biosynthesis in *S. bambergiensis*
- To activate two selected target gene clusters via overexpression of their pathway-specific regulators
- To test biological activity of the activated recombinant strains

2 Materials

2.1 Bacterial Strains and Vectors Used in This Work

Bacterial strain or vector	Relevant properties	Source or reference
S. bambergiensis		
S. bambergiensis Wild Type	Wild type, moenomycin producing	(Zotchev, Soldatova, Orekhov, & Schrempf, 1992)
S. bambergiensis MoeD	*S. bambergiensis* Moenomycin deletion mutant	This work
S. bambergiensis MoeD C1R4	*S. bambergiensis* Moenomycin deletion mutant with overexpressed regulator 4 in cluster 1	This work
S. bambergiensis MoeD C1R5	*S. bambergiensis* Moenomycin deletion mutant with overexpressed regulator 5 in cluster 1	This work
S. bambergiensis MoeD C28R	*S. bambergiensis* Moenomycin deletion mutant with overexpressed regulator in cluster 28	This work
E. Coli		
XL1-blue	Strain used for transformation	Stratagene
ET12567	Strain used for transformation of non-methylated DNA	(Flett, Mersinias, & Smith, 1997)
Vectors		
pSOK806	Vector used for introduction of regulatory genes	(Sekurova, et al., 2004)
pSOK808	Vector used for deletion of moenomycin gene cluster	(Sekurova, et al., 2004)
pSOK806 C1R4	Vector containing regulatory gene 4 in cluster 1	This work
pSOK806 C1R5	Vector containing regulatory gene 5 in cluster 1	This work
pSOK806 C28R	Vector containing regulatory gene in cluster 28	This work
pSOK808 MoeD	Vector contining A and B parts flanking the moenomycin gene cluster	This work

2.2 Media

2 X YT

Difco Bacto Tryptone	16 g
Difco Bacto Yeast Extract	10 g
NaCl	5 g
dH_2O	1000 ml

Autoclaved at 120°C for 20 minutes. Stored at room temperature.

LB

Difco Bacto Tryptone	10 g
Difco Bacto Yeast Extract	5 g
NaCl	5 g
dH_2O	1000 ml

Autoclaved at 120°C for 20 minutes. Stored at room temperature.

LA

Difco Bacto Tryptone	10 g
Difco Bacto Yeast Extract	5 g
NaCl	5 g
Agar	15 g
dH_2O	1000 ml

Autoclaved at 120°C for 20 minutes. Stored at room temperature.

ISP4

Difco™ ISP Medium 4	37 g
dH_2O	1000 ml

Autoclaved at 120°C for 20 minutes. Plates stored at 4 °C.

ISP4 + $MgCl_2$

Difco™ ISP Medium 4	37 g
1 M $MgCl_2$ Stock Solution	10 ml
dH_2O	1000 ml

Autoclaved at 120°C for 20 minutes. Plates stored at 4 °C.

TSS (Transformation Storage Solution)

LB Broth	42.5 ml
PEG_{6000}	5 g
1 M $MgCl_2$ Stock Solution	2.5 ml
DMSO	2.5 ml

PH adjusted to 6.5. Autoclaved at 120°C for 20 minutes, DMSO added after autoclavation.

TSB Medium

Tryptone Soya Broth	30 g
dH_2O	1000 ml

Autoclaved at 120°C for 20 minutes. Stored at room temperature.

PM4-1 Medium

Glucose	150 g
Soy Meal	15 g
Corn Steep Solids (Sigma)	5 g
$CaCO_3$	2 g
TMS 1 Stock Solution	6 ml
dH_2O	1000 ml

Glucose solution (150g/L) was prepared and sterilized. 5 g of Corn Steep Solids were dissolved in 890 mL of distilled water. Corn Steep Solids were poured into 3 flasks (297 mL in each). While stirring, 5 g of Soy Meal and 0.7 g of $CaCO_3$ were added to each flask. The flasks were then stirred for 5 minutes and autoclaved at 120°C for 20 minutes. After sterilization, under a sterile hood, 100 ml of sterile 150 g/L glucose solution and 2 mL of TMS1 stock solution was added to each flask.

Lysis Buffer

Tris HCL - pH 8.0	1 ml
EDTA - pH 8.0	1 ml
Triton Stock Solution	6.25 ml
dH_2O	41.75 ml

5x Isothermal Mix (Gibson Mix)

5x Isothermal Buffer	160 µl
$10u^{-1}$ µl T5 Exonuclease	0.32 µl

$2u^{-1}$ µl T5 Phusion DNA Polymerase	10 µl
$40u^{-1}$ µl Taq DNA Ligase	80 µl
dH_2O	up to 600 µl

Stored in 15 µl aliquots at -20°C.

5010 Medium

Sucrose	30 g
$NaNO_3$	2 g
KH_2PO_4	1 g
$MgSO_4$ x 7 H_2O	0.5 g
KCl	0.5 g
$FeSO_4$ x 7 H_2O	0.01 g
dH_2O	1000 ml

Set to pH 7.8 then autoclave for 20 minutes at 120°C.

5288 Medium

Glycerol	15 g
Soy Meal	10 g
NaCl	5 g
$CaCO_3$	1 g
$CoCl_2$ x 6 H_2O	0.001 g
dH_2O	1000 ml

Set to pH 7.8 then autoclave for 20 minutes at 120°C.

5321 Medium

Peptone	10 g
Glucose	20 g
$CaCO_3$	2 g
$CoCl_2$ x 6 H_2O	0.001 g
dH_2O	1000 ml

Set to pH 7.8 then autoclave for 20 minutes at 120°C.

5333 Medium

Yeast Extract	4 g
Soluble Starch	15 g
KH_2PO_4	1 g
$MgSO_4$ x 7 H_2O	0.5 g
dH_2O	1000 ml

Set to pH 7.8 then autoclave for 20 minutes at 120°C.

2.3 Stock Solutions

PCR Mixes

PCR mixes N7 and N9 were used in this project. These are mixes from Master Amp™, High Fidelity, Extra-Long PCR Kit, from Epicentre Biotechnologies.

TMS 1 (Filter sterilized)	mg/l
$FeSO_4*7H_2O$	5000
$CuSO_4*5H_2O$	390
$ZnSO_4*7H_2O$	440
$MnSO_4*H_2O$	150
$Na_2MoO_4*2H_2O$	10
$CoCl_2*6H_2O$	20
HCl	50 ml/l

Glycerol 20%

Glycerol	200 ml
dH_2O	800 ml

$MgCl_2$ - 1M

$MgCl_2$	203.3 g
dH_2O	1 l

Tris HCl 1 M - pH 8.0

Tris	121.14 g
dH_2O	1 l

TfBI Solution (500 ml)

Potassium Acetate	1.47 g
$MnCl_2$	4.95 g
RbCl	6.05 g
$CaCl_2$	0.74 g
Glycerol	75 ml

pH 5.8 with 0.2 M Acetic Acid
Filter sterilize and store at 4°C

TfBII Solution (100 ml)

100 mM MOPS pH 7.0	10 ml
$CaCl_2$	1.10 g
RbCl	0.12 g
Glycerol	15 ml

Autoclave and store at 4°C in the dark

2.4 Primers

2.4.1 pMoeD Construct

PCR for pSOK808 part

Forward primer:

5‘ – GCGCATCGTATGCGCTCTTCGGTGAACTTCGCACTTCG – 3‘

Reverse primer:

5‘ – CCTCTAGAGTCCGTCTGACGCGTGTGGGCCTTCCGCGAGAAGTTACG – 3‘

PCR for A part

Forward primer:

5‘ – TCTGACGCGTGTGGGCCTTCCGCGAGAAGTTACGG – 3‘

Reverse primer:

5‘ – GTCGACGTATCCGGCCATCAGGTTCCTTCGCCGAAT – 3‘

PCR for B part

Forward primer:

5‘ – CCGGCCATCAGGTTCCTTCGCCGAATCCGTACTGG – 3‘

Reverse primer:

5‘ – GCCCAGAAGGAGCGCATCGTATGCGCTCTTCGGTGAA – 3‘

2.4.2 C1R4 Regulator Construct

PCR for pSOK806 part

Forward primer:

5‘ – GGATTTCCTGACCACCTGCAGGTCGACTCTAGAGGATCCG – 3‘

Reverse primer:

5‘ – GGACAATCGTGCCGGTTGGTAGGATCCATGAGTATTCGTCACCG – 3‘

PCR for C1R4 regulator part

Forward primer:

5‘ – GCCGGTTGGTAGGATCCATGAGTATTCGTCACCGGTGAGCCC – 3‘

Reverse primer:

5‘ – CCGTCGGAGTGGATTTCCTGACCACCTGCAGGTCGACTCTAGAGGA – 3‘

2.4.3 C1R5 Regulator Construct

PCR for pSOK806 part

Forward primer:

5‘ – GAGACCGTTCTCTCTGCAGGTCGACTCTAGAGGATCCG – 3‘

Reverse primer:

5‘ – GGACAATCGTGCCGGTTGGTAGGATCGATCAGCAGCCGATC – 3‘

PCR for Regulator part

Forward primer:

5‘ – GGTTGGTAGGATCGATCAGCAGCCGATCAGTACGGTTGG – 3‘

Reverse primer:

5‘ – CCGGCGTGATCAGAGACCGTTGTCTCTGCAGGTCGACTC – 3‘

2.4.4 C28R Regulator Construct

PCR for pSOK806 part

Forward primer:

5‘ – GGTTGATGAGTTCGTCGTCGTCCTGCAGGTCGACTCTAGAGGATCCG – 3‘

Reverse primer:

5‘ – GGACAATCGTGCCGGTTGGTAGGATCCACCACAACTGCTGG – 3‘

PCR for Regulator part

Forward primer:

5‘ – CGGTTGGTAGGATCCACCACAACTGCTGGTCCGGC – 3‘

Reverse primer:

5‘ – GGTTGATGAGTTCGTCGTCGTCCTGCAGGTCGACTCTAGAGG – 3‘

3 Methods

3.1 In Silico Mining

The online resource tool AntiSMASH 2.0 (http://www.secondarymetabolites.org/) was used for the identification of secondary metabolite gene clusters of *S. bambergiensis.*

Uploading the whole genome sequence of *S. bambergiensis*, showed that the bacteria contains 29 gene clusters.

Using a combination of antiSMASH, Excel and the online database the National Centre for Biotechnology Information (NCBI), (http://www.ncbi.nlm.nih.gov/), an annotation of all the clusters was made. Each putative gene was blasted in the NCBI databank and the gene with the highest homology was inserted into the Excel sheet (See Annex 1).

3.2 Storage of Bacterial Strains

All bacterial strains were stored at -80°C in 20% glycerol stocks.

Spore suspensions from plates with freshly sporulated *S. bambergiensis* were prepared as follows:

1. 5 ml of 20% glycerol was poured on the plate

2. Using a sterile pipette tip, the surface of the plate was gently scratched so as to get the spores suspended

3. Suspension was then drained of the plate and filtered through sterile cotton

4. Spore suspensions were stored at -80° C

Thawing of *Streptomyces* strains was always done on ice as strains will soon lose viability if germination starts. (Kieser T. , Bibb, Chater, & Hopwood, 2000)

3.3 Preparation of Antibiotic Stock Solutions

Antibiotics stock solutions were prepared as follows:

Apramycin or Am_{100} was prepared by mixing 100 mg of Apramycin with 1 ml of dH_2O.

Ampicillin or Amp_{100} was prepared by mixing 100 mg of Ampicillin with 1 ml of dH_2O.

Thiostrepton or $Thio_{30}$ was prepared by mixing 30 mg of Thiostrepton with 1 ml of DMSO.

Nalidixic Acid or Nal_{30} was prepared by mixing 30 mg of Nalidixic Acid with 1 ml of 0.1 M NaOH.

Chloramphenicol or Cml_{25} was prepared by mixing 25 mg of Chloramphenicol with 1 ml of Ethanol absolute.

Kanamycin or Kan_{25} was prepared by mixing 25 mg of Kanamycin with 1 ml of dH_2O.

Stock solutions were stored in aliquots at -20°C.

3.4 Preparation of Competent Cells

In nature, competent bacteria are able to take up exogenous DNA which possibly leads to genetic transformation. Artificial competence can be achieved by altering the permeability of the cell membrane. (Chen & Dubnau, 2004)

20 ml of LB was inoculated with a single colony of XL1-Blue or ET12567, and incubated at 37°C on a shaker overnight.

This culture was transferred to 300 ml pre-warmed LB media and grown for another two hours in shaking incubator at 37°C until the OD at 600 nm reached 0.3 – 0.4.

Culture was kept on ice from this point on.

Cells were chilled for 5 minutes on ice and then transferred into six chilled 50 ml tubes. The tubes were centrifuged for 10 minutes at 4000 rpm at 4° in a table top centrifuge. Supernatant was discarded, and each pellet resuspended in 15 ml of chilled TfBI media.

Tubes were incubated on ice for 30 minutes, then centrifuged for 10 minutes at 4000 rpm at 4°C and supernatant discarded. Each pellet was resuspended in 1 ml of chilled TfBII media.

Aliquots of 100 µl were snap frozen using dry ice and stored at -80°C.

3.5 Extraction of Plasmid DNA from Bacteria

Extraction of plasmid DNA from bacteria was done using the kit Wizard Plus SV Minipreps DNA Purification System.

Bacteria with plasmid of interest would be incubated in LB liquid medium overnight at 37°C on a shaker. The next day the overnight cultures would be centrifuged for 5 minutes at 10,000 rpm. The supernatant was discarded and the pellet was resuspended in 250 µl of cell Resuspension solution. 250 µl of cell Lysis solution was added to the sample and the tube was inverted 4 times to mix. 10 µl of Alkaline protease solution was added and the tube was inverted 4 times to mix. The tube was incubated for 5 minutes at room temperature. 350 µl of Neutralization solution was added and the tube was inverted 4 times to mix. The tube was

centrifuged for 10 minutes at 10,000 rpm. A spin column was placed in a collection tube and the cleared lysate was decanted into the spin column. The spin column was centrifuged for 1 minute at 10,000 rpm at room temperature. Flow through was discarded and column was reinserted into collection tube. 750 µl of Washing solution was added onto column. Column was centrifuged for 1 minute at 10,000 rpm. Flow through was discarded and column reinserted into collection tube. This step was repeated with 250 µl Washing solution. Column was centrifuged for 2 minutes at 10,000 rpm. Spin column was transferred to a sterile 1.5 ml micro centrifuge tube. 100 µl of sterile, nuclease-free water was added to the spin column. Column was centrifuged for 1 minute at 10,000 rpm at room temperature. Column was discarded and sample was stored at -20°C.

3.6 Isolation of Total (Genomic) DNA

Isolation of total DNA of *S. bambergiensis* wild type and *S. bambergiensis* moenomycin deletion mutant was done using the Qiagen Dneasy Kit.

Spores were inoculated in 2 ml TSB medium and incubated at 30°C on shaker overnight. Next day, 1 ml of the culture was transferred into an eppendorf tube, mixed with 0.5 ml sterile water, and spun down at 13,000 rpm for 3 minutes. The supernatant was discarded, and the pellet was resuspended in 180 µl lysis buffer (See Materials chapter). Lysozyme was added just before the experiment. The mixture was incubated at 37°C for 15 minutes and the suspension mixed gently by tipping the end of the tube every 5 minutes. 25 µl of Proteinase K was added and mixed by pipetting, then 200 µl of Lysis buffer (AL) was added and mixed by pipetting. The mixture was incubated at 55° for 30 minutes. 200 µl of 96% ethanol were added, mixed well by pipetting, and then applied on a column sitting in a 2 ml collection tube. The column was centrifuged for 2 minutes at 13,000 rpm, and transferred to a new 2 ml tube. 500 µl of Washing buffer (AW1) was applied on the column and spun for 1 minute at 13,000 rpm. The column was transferred to a new 2 ml tube and 500 µl of Washing buffer (AW2) was applied on the column and spun for 3 minutes at 13,000 rpm. The column was then transferred into a new eppendorf tube and 125 µl of Elution buffer (AE), preheated to 50°C, was applied. This tube was incubated at room temperature for 15 minutes, and centrifuged for 1 minute at 13,000 rpm.

The DNA was then checked using gel electrophoresis.

3.7 Gel Electrophoresis

Gel electrophoresis was used in this project to check the size of DNA fragments and also to confirm the presence of a product. 0.8% agarose with Gel Green stain was used to prepare the gels.

3.8 Excision and Purification of DNA from Agarose Gels

PCR products were excised from agarose gels in order to purify the DNA product, so it could be used in the Gibson assembly. QIAquick Gel extraction kit was used.

PCR products were applied on agarose gels and run on 110V for 60-80 minutes. The gel was placed inside the Gel Doc imager and photographed. Ultra violet light was used to see the PCR products on the gel. The correct DNA products were excised from the gel using a clean, sharp scalpel. The slice was placed in an eppendorf tube and weighed. 3 volumes of the buffer QG were added for every 1 volume of gel. For example if the slice was 150 mg, then 450 µl of QG buffer was added. The eppendorf containing the slice and buffer were incubated at 50°C for 10 minutes, vortexed every 3 minutes during incubation. When the slice was completely dissolved, 1 gel volume of isopropanol was added and mixed. QIAquick spin columns were placed in 2 ml collection tubes. The sample from the eppendorf was applied on the column and it was centrifuged for 1 minute on 13,000 rpm. Flow through was discarded and QIAquick column was placed back in the same collection tube. 750 µl of PE buffer was added onto the column and it was centrifuged for 1 minute on 13,000 rpm. Flow through was discarded and the column was centrifuged for an additional 1 minute on 13,000 rpm. The column was placed in a sterile eppendorf tube and 50 µl of EB buffer was applied to the centre of the column membrane and then it was centrifuged for 1 minute on 13,000 rpm. This sample was then checked on a gel to confirm correct size. The product was stored at -20°C.

3.9 Gibson Assembly

All assemblies of DNA fragments were carried out using the Gibson assembly method.

The Gibson assembly is an isothermal, single reaction method (Figure 6). It is used to assemble DNA fragments with overlapping regions. This method uses 5´ exonucleases, Phusion DNA polymerase and Taq DNA ligase. Having two adjacent DNA fragments that share terminal sequence overlaps, they can be joined into a covalently sealed molecule in a one-step isothermal reaction. T5 exonuclease removes nucleotides from the 5’ ends of double stranded DNA molecules. The 3’ complementary single-stranded DNA overhangs then anneal and the Phusion

DNA polymerase fills the gaps and Taq DNA ligase seals the nicks. T5 exonuclease is heat-labile and is inactivated during the 50°C incubation. (Gibson, et al., 2009)

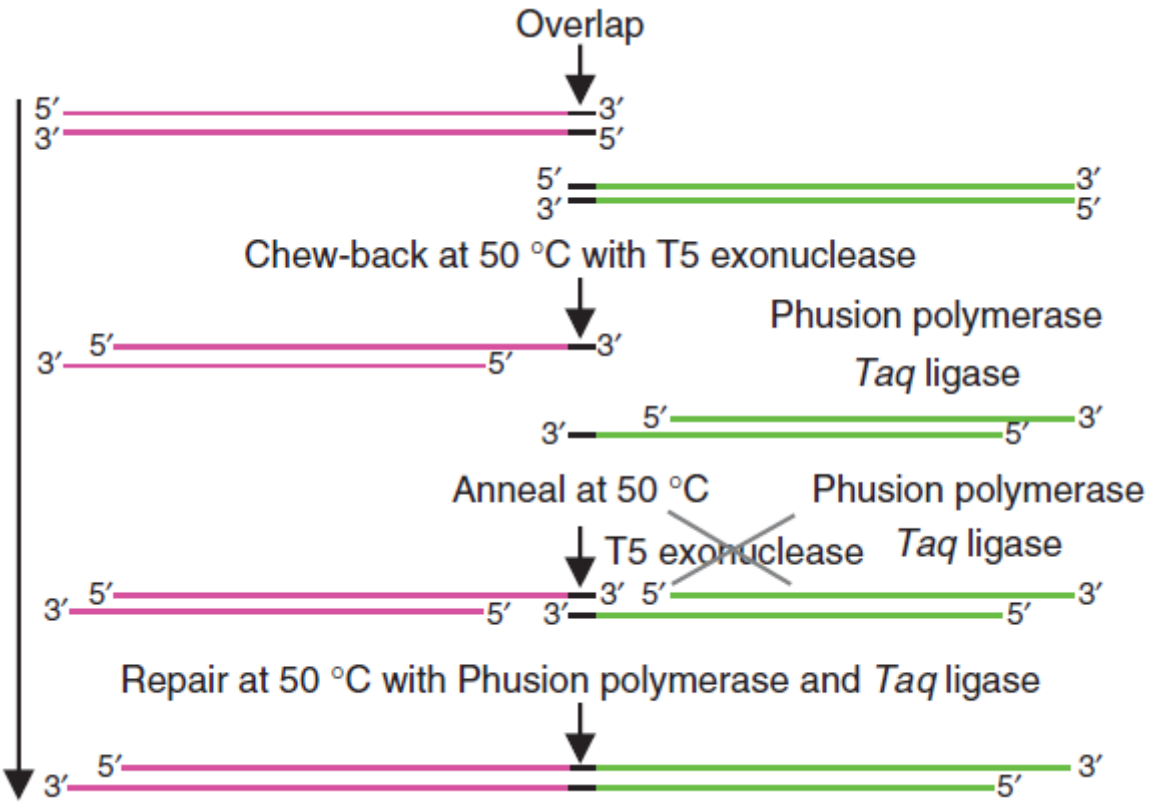


Figure 6: The Gibson assembly is an isothermal reaction. Two adjacent fragments (magenta and green) share terminal sequence overlaps (black). These fragments are joined into a covalently sealed molecule. T5 exonuclease removes nucleotides from the 5' ends of the double-stranded DNA molecules. Complementary single-stranded DNA overhangs anneal and Phusion DNA polymerase fills the gaps and Taq DNA ligase seals the nicks. T5 exonuclease is heat-labile and is inactivated during the 50°C incubation. (Gibson, et al., 2009)

All vectors were first treated with a DpnI restriction enzyme. DpnI cuts methylated DNA. The pSOK806 and pSOK808 templates used in the PCR contain methylated DNA. DpnI treatment chews up the template plasmid but not the PCR product. (Zhng, Baumann, & Reymond, 2004)

For the DpnI treatment of the vector parts, 14 µl of purified PCR vector product, 2 µl Cutsmart 10X Buffer, 1 µl DpnI enzyme and dH_2O up to 20 µl, were mixed in an eppendorf tube and placed in a PCR machine. The program for the DpnI treatment is shown in the following table (Table 3).

Table 3: Program for DpnI treatment of vector part.

Step	Temperature (° C)	Duration (Minutes)
1.	37	60
2.	80	20
3.	4	Forever

The Gibson reaction for the moenomycin deletion vector consisted of 15 µl of Gibson mix, 0.5 µl of DpnI treated vector part, 2.5 µl of PCR amplified part A and 2.0 µl of PCR amplified part B.

The Gibson reaction for the regulator parts consisted of 15 µl of Gibson mix, 1 µl of DpnI treated vector, and 4 µl PCR amplified DNA for the regulatory gene.

The program used in the PCR machine for the Gibson assembly method is shown in the following table (Table 4).

Table 4: Program for Gibson assembly.

Step	Temperature (° C)	Duration (Minutes)
1.	50	60
3.	4	Forever

3.10 Transformation

Transformation is the incorporation of foreign DNA into bacterial cells. In this work plasmid vectors were introduced into competent XL1-Blue *E. coli* cells. XL1-Blue gives high yields of transformants but as many *Streptomyces* strains possess a restriction system that cuts methylated DNA, a totally non-methylating *E. coli* host is needed as a DNA donor for conjugation. (Kieser T. , Bibb, Buttner, Chater, & Hopwood, 2000)

DNA constructs made in this work, by Gibson assembly, were transformed into XL1-Blue cells and the plasmids were isolated, and analysed by restriction analysis. Confirmed by Gel Electrophoresis, constructs were transformed into the non-methylating ET12567 *E. coli* cells.

Competent cells were taken from -80°C storage and placed on ice for 5 minutes. 1 µl of relevant plasmid DNA was added, mixed by tapping the tube with a finger and placed back to ice for 15 minutes. The tube was then transferred to a thermoblock at 42°C for 1 minute for heat shock and set back to ice for 2 – 10 minutes. 800 µl of LB media was added, and the tube placed into

shaking incubator at 37°C for 40 minutes to 2 hours. 100 µl of the suspension was plated on LA plates with appropriate antibiotics for selection.

Plates were placed in 37°C for overnight incubation.

3.11 Conjugation

Bacterial conjugation involves the direct transfer of DNA from a donor cell to a recipient cell through a specialized intercellular channel, which forms between them. (Snustad & Simmons, 2012)

In this work, the *E. coli* strain, ET12567 with helper plasmid pUZ8002 was used for conjugation. This is a non-methylating strain. A totally non-methylating *E. coli* cell is needed as a DNA donor for conjugation as many *Streptomyces* contain a methyl specific restriction system. (Kieser T. , Bibb, Chater, & Hopwood, 2000) After successful transformation of plasmid into *E.* coli strain ET12567, conjugation into *Streptomyces* could proceed. The conjugation process transferred the plasmid into the *Streptomyces* species. Transferred plasmid integrates into host genome by way of homologous recombination.

Transfer of DNA from donor cell to recipient cell is initiated at the oriT site (Origin of Transfer) on the plasmid. During conjugation, one strand of the plasmid is cut at the oriT by an enzyme and one end is transferred into the recipient cell through a channel that forms between the two conjugating cells. (Snustad & Simmons, 2012)

Conjugations in this work involved the plasmids pSOK808 for deletion of moenomycin and pSOK806 for overexpression of regulators. (Sekurova, et al., 2004)

Plasmid pSOK808 contains an RP4 oriT site for initiation of transfer, Am^r to confer resistance to apramycin and a *gusA* site for gusA assays (See Figure 7).

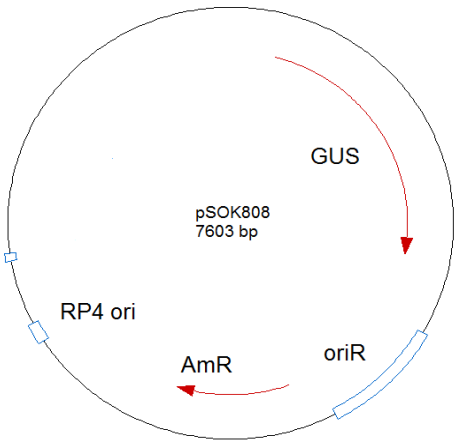


Figure 7: Plasmid pSOK808 was used for deletion of moenomycin gene cluster. The part of the plasmid used in this project contains an RP4 oriT site, AmR site and a gusA site.

Plasmid pSOK806 contains an RP4 oriT site for initiation of transfer, Am^r to confer resistance to apramycin, attP site and integrase gene for site specific integration and an *ermE*p* up strong constitutive promoter region for overexpression of regulators (Figure 8).

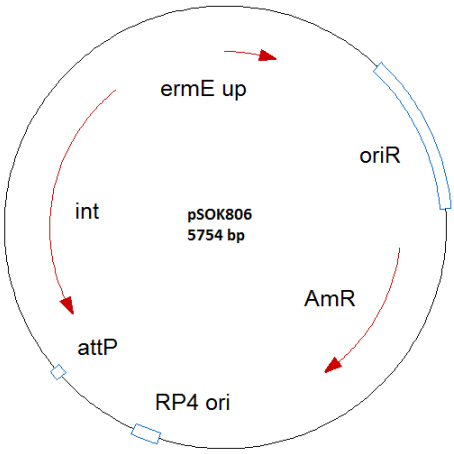


*Figure 8: Plasmid pSOK806 was used for overexpression of regulator parts. The part of the plasmid used in this project contains an RP4 oriT site, AmR, attP site and integrase gene and an ermE*p up strong constitutive promoter region.*

The day before conjugations, overnight cultures of ET12567 cells were prepared on appropriate selection plates.

The next day spore suspensions of *Streptomyces* strains were taken from -80°C storage and thawed out on ice. 50 µl of this spore suspension was added into 350 µl of 2 X YT media. This suspension was put for 5 – 10 minutes at 50°C for heat shock, and then allowed to cool down at room temperature for 10 – 15 minutes.

An area (ca. 1.5x1.5 cm) of the ET12567 cells from fresh selective plate was swiped and suspended in 500 µl of 2 X YT media.

400 µl of the heat shocked *Streptomyces* and 100 µl of ET12567 cell suspension were mixed carefully by pipetting, and then spun down for 1 min at 5000 rpm. 350 µl of supernatant was removed and the remaining suspensions plated on dry ISP4 + $MgCl_2$ plates. Then stored at 30° for 18 – 20 hours.

18 – 20 hours later selection for recombinant *Streptomyces* was done by adding on each plate 1 ml dH_2O + 30 µl of Nal_{30} and 15 µl of Am_{100}, gently rubbed into media and let dry out for about 20 minutes. Incubation continued for 2 – 3 days at 30°C.

Transconjugants were picked with sterile toothpicks and plated on new ISP4 with Nalidixic acid and another appropriate antibiotic for selection of *Streptomyces*. This was further incubated for about 3 – 4 days or until good growth was observed. The most visibly viable colonies were transferred to new ISP4 with appropriate antibiotic and incubated until sporulation.

Spore suspensions of recombinant *Streptomyces* strains were made and stored at -80°C for further analysis.

3.12 Construction of *S. bambergiensis* Moenomycin Deletion Mutant

In order to get a clear background for the detection of novel secondary metabolites, a *S. bambergiensis* moenomycin deletion mutant was needed, as *S. bambergiensis* is a known producer of the antibiotic moenomycin.

The moenomycin gene cluster was identified using the online resource tool, AntiSMASH 2.0 (See Method 3.1). (Blin, et al., 2013)

A strategy was devised for the deletion of the scaffold genes for moenomycin production, where primers were designed for two regions, A and B, which flank the moenomycin gene cluster. These regions were then PCR amplified and cloned into a vector, pSOK808, using the Gibson assembly method. This vector was then introduced into the *E. coli* cells by transformation and these cells were then in turn conjugated to the wild type *S. bambergiensis.*

Following a double crossover event, the moenomycin deletion mutant was made (Figure 9).

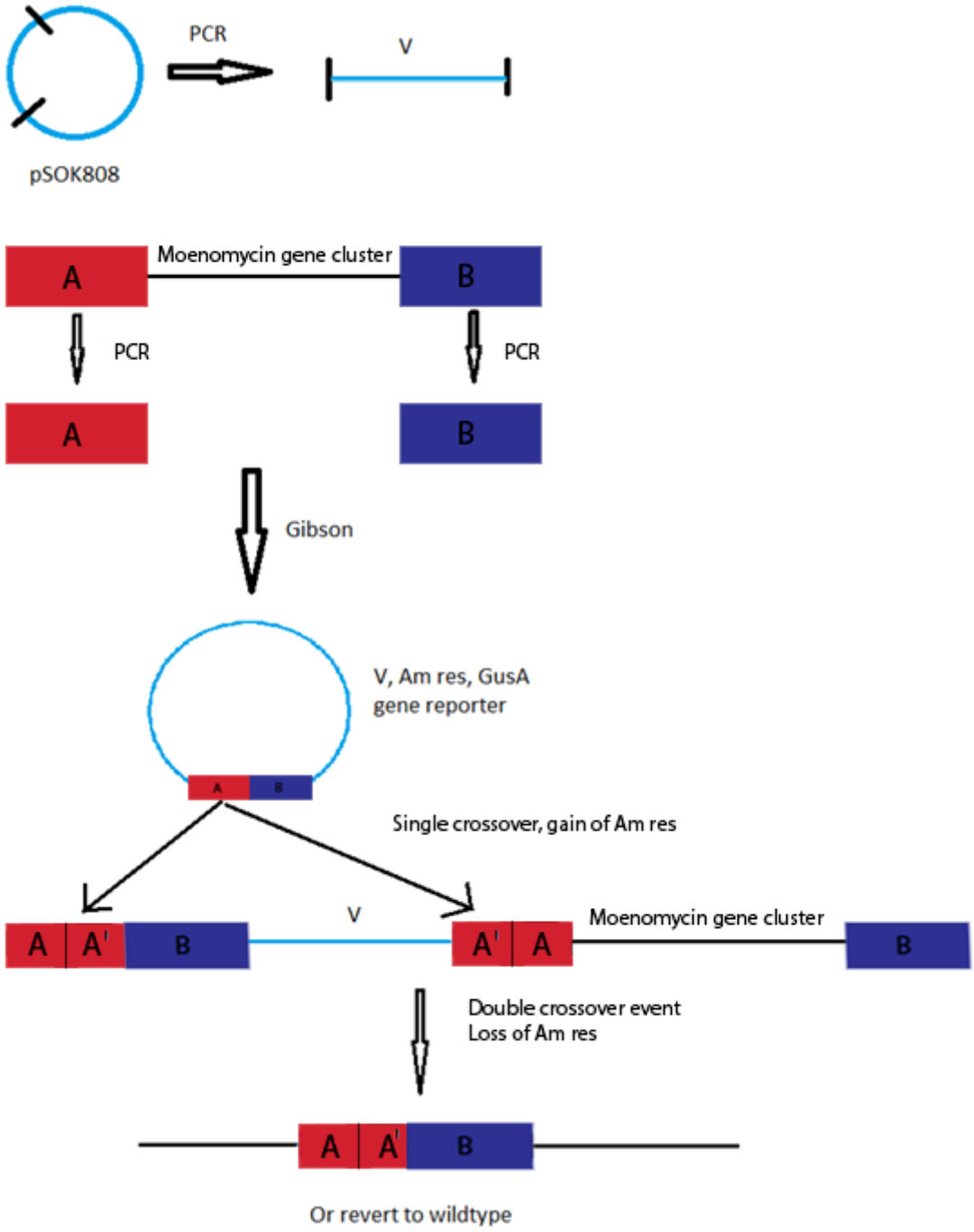


Figure 9: Schematic of moenomycin deletion strategy. Using designed primers, part of the plasmid pSOK808 was amplified using PCR and also A and B, flanking the moenomycin gene cluster in S. bambergiensis. Using the Gibson assembly method, these 3 parts were assembled and introduced into E. coli cells via transformation and then into S. bambergiensis via conjugation. Single and double crossover events would then later occur as explained in detail in the text. For the S. bambergiensis moenomycin deletion mutant, a scar of A + B parts would be observed.

A more detailed explanation of the method follows.

Using the j5 program (https://j5.jbei.org/) and Clone Manager, primers were designed for amplifying the flanking regions, A and B, of the moenomycin gene cluster. Primers were also designed for the vector pSOK808. The part of the pSOK808 vector contained the Am^r region and the *gusA* reporter gene. The following program was used for the PCR amplification (Table 5).

Table 5: PCR program used to amplify regions A and B of the moenomycin gene cluster and also for the pSOK808 vector part.

Step	Temperature (° C)	Duration	Number of cycles
1.	95	3 minutes	
2.	94	45 seconds	x25
3.	56	45 seconds	
4.	68	10 minutes	
5.	68	10 minutes	
6.	4	Forever	

The PCR consisted of 1 µl DNA, 0.5 µl of each primer, 20 µl master mix (N7 for A and B parts, N9 for pSOK808), (See Materials, Stock Solutions), 0.8 µl polymerase Q5 and dH_2O up to 40 µl.

The PCR product of vector part was treated with DpnI before assembly. For the DpnI treatment, 14 µl of purified PCR vector product, 2 µl Cutsmart 10X Buffer, 1 µl DpnI enzyme and dH_2O up to 20 µl, were mixed in an eppendorf tube and placed in a PCR machine. The program for the DpnI treatment PCR is shown in the following table (Table 6).

Table 6: PCR program for DpnI treatment of pSOK808 vector part.

Step	Temperature (° C)	Duration (Minutes)
1.	37	60
2.	80	20
3.	4	Forever

These three parts were then assembled using the Gibson assembly method (Gibson, et al., 2009). The volume of the Gibson reaction was 20 µl and consisted of 15 µl Gibson Mix, 0.5 µl of

vector (pSOK808) DNA, 2.5 µl of A part DNA and 2.5 µl of B part DNA. This was gently mixed and a reaction was performed using the following program (Table 7).

Table 7: Program for the Gibson assembly reaction of pMoeD.

Step	Temperature (° C)	Duration
1.	50	60 minutes
2.	4	Forever

A figure of the J5 software generated pMoeD molecule can be seen in the following figure (Figure 10).

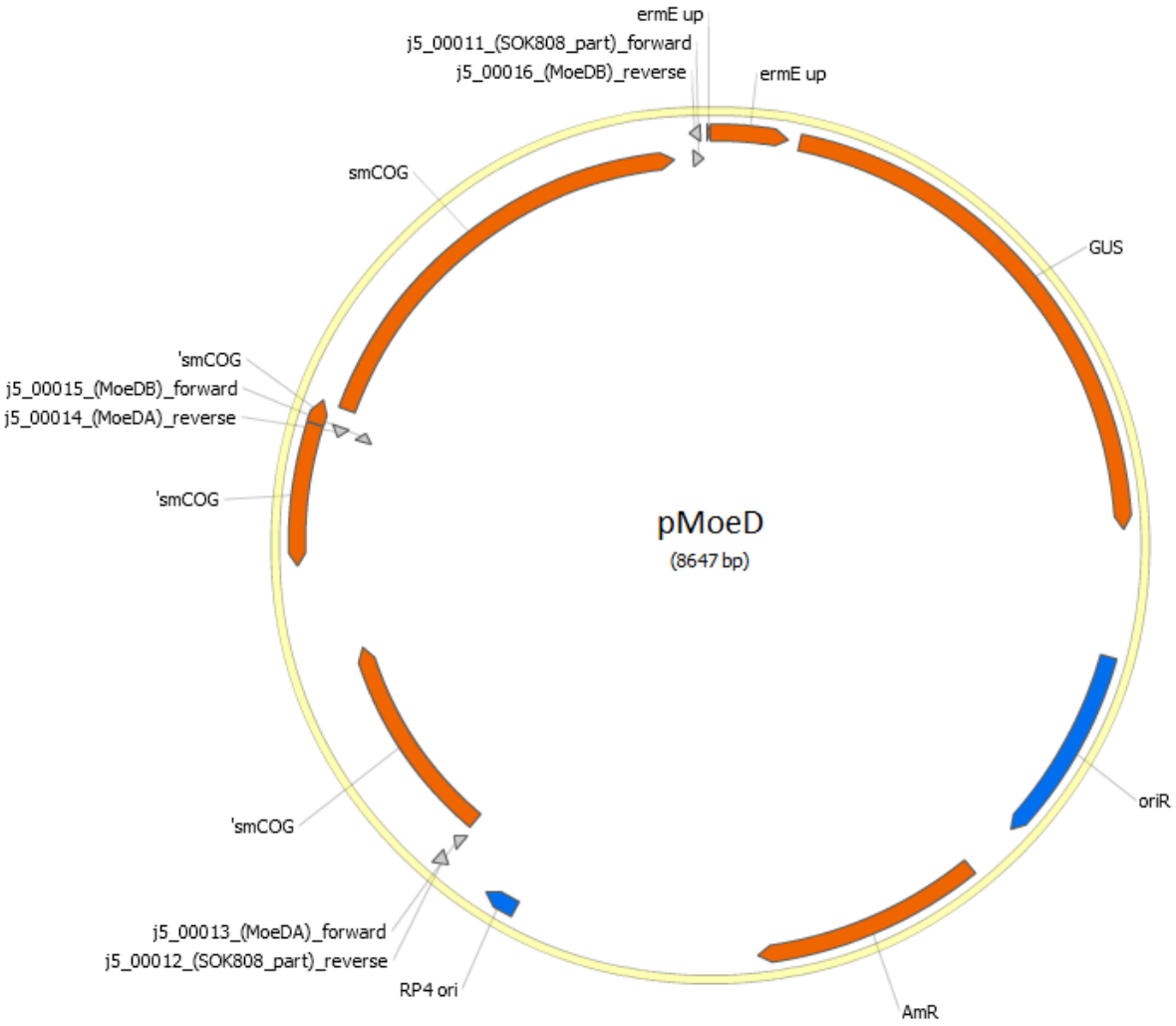


Figure 10: J5 software representation of the moenomycin deletion construct pMoeD.

This construct was transformed into XL1-BLue competent cells, using the transformation procedure as explained in the methods (See Method 3.10).

Next step was the plasmid isolation from the transformed XLBlue1 cells (See Method 3.5). Restriction analysis was done to confirm the correct clone using BamHI (See Method 3.14).

Restriction assay consisted of 2 µl of DNA, 2 µl of NEB buffer 3.1, 0.8 µl of BamHI and dH_2O up to 20 µl.

After confirming the correct construct, the plasmid, designated as pMoeD, was then transformed into competent ET12567 cells using the transformation method (See Method 3.10).

PMoeD was then conjugated into the wild type *S. bambergiensis*, as described in the methods (See Method 3.11).

Once the single crossover had occurred, the *S. bambergiensis* with pMoeD integrated into its chromosome, had attained an Am^r, and can be selected on ISP4 plates with Am. Crossover mutants were selected to obtain double crossover mutants. For this purpose 3 rounds of sporulation on the ISP4 plates without selection with following gusA assay were performed (See Method 3.13).

After 3 full rounds of sporulation, spore suspension was plated on ISP4 with appropriate antibiotic as explained in methods. Those colonies that hadn´t gained a blue colour were checked for sensitivity for Am, as the double crossover *S. bambergiensis* would have lost their Am^r.

The genomic DNA from the colourless, Am sensitive colonies was then isolated using the DNA isolation method (See Method 3.6)

To confirm the 2nd crossover candidates, a PCR was performed using forward primer of the A part and reverse primer of the B part. The plasmid pMoeD was used as a template for positive control in the reactions.

For this PCR, 1 µl of genomic/plasmid DNA was used, 0.5 µl of each primer, 20 µl of N7 PCR mix, 0.8 µl of Q5 Polymerase and dH_2O up to 40 µl. The program used was as follows (Table 8).

Table 8: PCR program used for confirmation of moenomycin deletion mutant.

Step	Temperature (° C)	Duration	Number of cycles
1.	95	60 minutes	
2.	94	45 seconds	x25
3.	56	45 seconds	
4.	68	10 minutes	
5.	68	10 minutes	
6.	4	Forever	

3.13 GusA Assay for Selection of Second Crossover Mutant

GusA, encoding enzyme β-glucuronidase (GusA) is a reporter gene which is a part of the pSOK808 vector used in this research. GusA converts the substrate, 5-bromo-4-chloro-3-indolyl β-D-glucuronide cyclohexylammonium salt, and cells appear blue on media plates (Figure 11). (Siegl & Luzhetskyy, 2012)

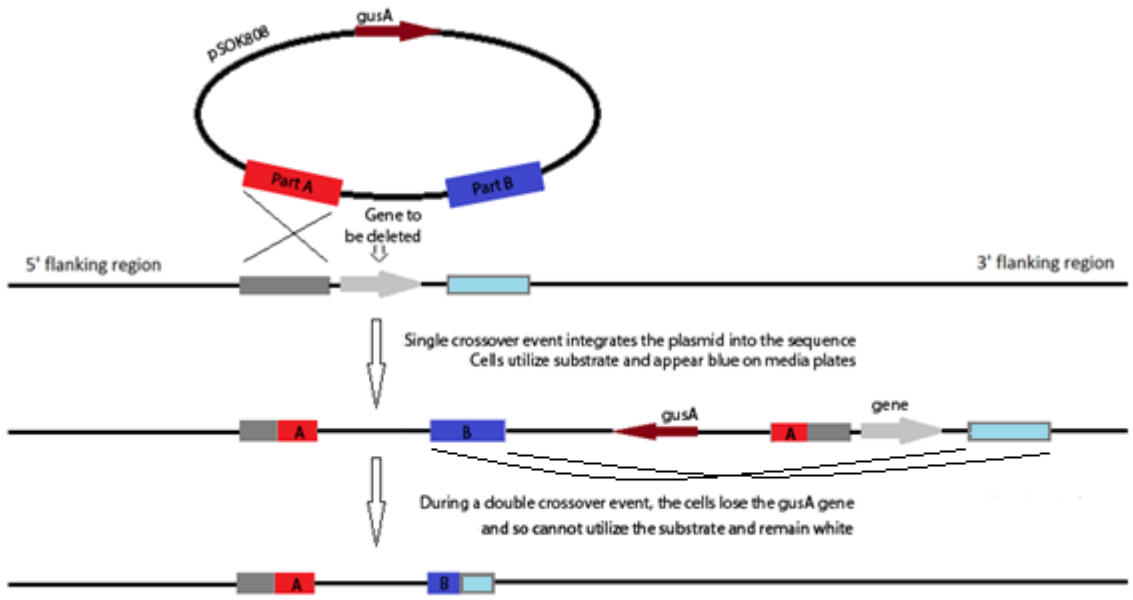


Figure 11: Using Gibson assembly, genes A and B, flanking the moenomycin gene cluster, were cloned into the pSOK808 vector. pSOK808 contains the gusA reporter gene. When this plasmid is inserted into the moenomycin gene cluster via single crossover, the cell gains the ability to utilize the added substrate and turn blue. During a double crossover event, the cells lose this ability and remain white. In this way the cells that have the desired deletion can easily be selected. (Siegl & Luzhetskyy, 2012)

Several dilutions of spore suspensions of candidate for second crossover were made in order to obtain single colonies on a plate for GusA assay. 1 ml of sterile dH_2O was added to 7 eppendorf vials. 100 µl of *S. bambergiensis* MoeD second crossover mutants were added to the first one and mixed by pipetting. 100 µl from the first eppendorf was added to the second and mixed and so on, until a dilution series up to -7 was reached. Dilutions -5, -6 and -7 were plated and dilution -5 was selected as the most appropriate as a colony count of around 200 colonies was observed.

3 tablets of 5-bromo-4-chloro-3-indolyl β-D-glucuronide cyclohexylammonium salt (GusA substrate) were added to 300 µl of dH_2O. This was vortexed until completely dissolved. This was added to 300 ml of autoclaved ISP4 medium, mixed and poured onto plates.

150 µl of -5 dilution of *S. bambergiensis* candidate for 2nd crossover (moenomycin deletion mutant), was spread on plates w. added GusA substrate, and incubated at 30°C. Daily observations ensued. After 3 days of incubation, 8 non-blue colonies were selected and checked for Am sensitivity.

3.14 Construction of Vector for Overexpression of Regulatory Gene from *S. bambergiensis*

In order to overexpress a regulatory gene from *S. bambergiensis*, it was cloned into the integrative vector under the control of strong constitutive promoter *ermE*p*, and introduced into moenomycin deletion mutant (Figure 12). (Sezonov, Duchene, Friedman, Guerineau, & Pernodet, 1998)

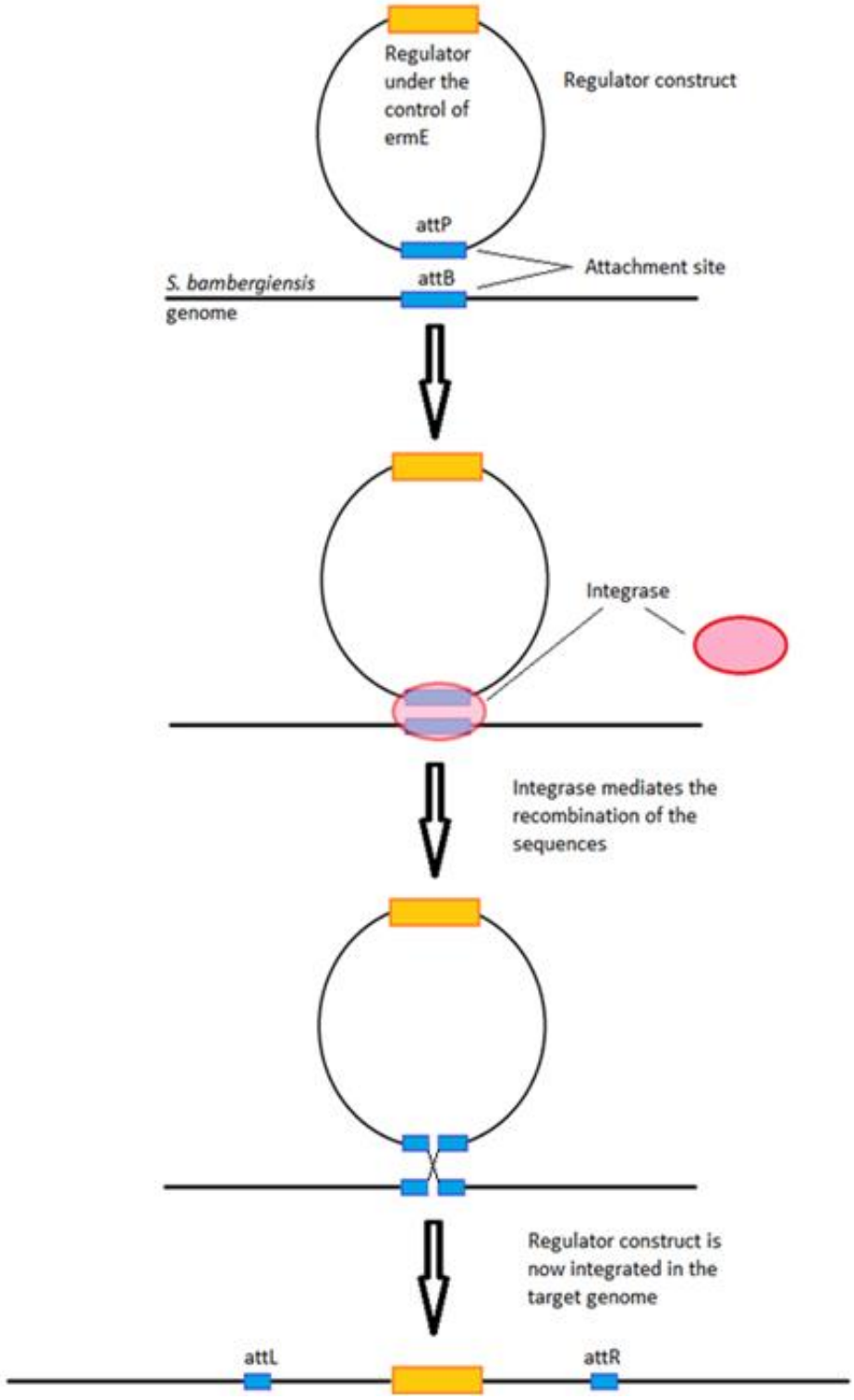


*Figure 12: Once regulator part is integrated with the pSOK806 vector part, the vector for overexpression of regulatory gene is complete. The vector for overexpression of regulatory gene contains an integrase gene and an attP site, along with the strong constitutive promoter ermE*p, regulatory gene, AmR (apramycin resistance) and the oriR (origin of replication). With the aid of an integrase enzyme, the attP site of the construct recombines with the attB site on the target genome.*

Using AntiSMASH, regulators were selected for over expression (See Method 3.1). Regulatory genes 4 and 5 from cluster 1 and the single regulator from cluster 28, designated as C1R4, C1R5 and C28R, respectively, were chosen for this experiment.

Primers for amplifying the regulators and the pSOK806 vector, were designed using the online program j5 (https://j5.jbei.org/) and the Clone Manager software.

PCR amplification for the pSOK806 vector was performed using 0.5 µl pSOK806 DNA, 20 µl N9 enzyme mix (see appendix), 0.8 µl of polymerase Q5, 0.5 µl forward primer, 0.5 µl reverse primer and dH_2O up to 20 µl

PCR amplification for the regulators was performed using 1 µl of *S. bambergiensis* total DNA, 20 µl N7 PCR mix, 0.8 µl of polymerase Q5, 0.5 µl forward primer, 0.5 µl reverse primer and dH_2O up to 20 µl.

The PCR program used was as follows (Table 9).

Table 9: PCR program used for amplification of regulatory genes and pSOK806 parts.

Step	Temperature (° C)	Duration	Number of cycles
1.	95	60 minutes	
2.	94	45 seconds	x25
3.	56	45 seconds	
4.	68	10 minutes	
5.	68	10 minutes	
6.	4	Forever	

Gel electrophoresis was performed using 0.8% agarose containing Gel Green. DNA stain. The gel was run on 100V for 1 hour.

Bands were cut out of the gel and purification of DNA was performed as explained in the methods (See Method 3.8).

The purified DNA fragments were then assembled via Gibson assembly method (See Method 3.9)

The vector parts were all treated with DpnI before assembly. For the DpnI treatment of the vector parts, 14 µl of purified PCR vector product, 2 µl Cutsmart 10X Buffer, 1 µl DpnI enzyme and dH_2O up to 20 µl, were mixed in an eppendorf tube and placed in a PCR machine. The reaction was performed as it is shown in the following table (Table 10).

Table 10: Program for DpnI treatment of pSOK806 vector part.

Step	Temperature (° C)	Duration (Minutes)
1.	37	60
2.	80	20
3.	4	Forever

The two parts were then assembled using the Gibson assembly (Gibson, et al., 2009). The volume of the Gibson reaction was 20 µl and consisted of 15 µl Gibson Mix, 1 µl of the vector (pSOK806) DNA, 4 µl of the regulatory gene DNA. This was gently mixed and an isothermal Gibson reaction was performed as in the following table (Table 11).

Table 11: PCR program for the Gibson assembly reaction of regulators

Step	Temperature (° C)	Duration (Minutes)
1.	50	60
2.	4	Forever

These constructs were transformed into XL1BLue competent cells, using the transformation procedure as explained in the methods (See Method 3.10)

Next step was the plasmid isolation from the transformed XL1Blue cells (See Method 3.5). Restriction analysis was done to confirm the correct clone.

Restriction assay was done with various restriction enzymes and it consisted of 2 µl of DNA, 2 µl of NEB buffer 3.1, 0.8 µl of each restriction enzyme and dH_2O up to 20 µl.

The restriction enzyme used for C1R4 construct was PvuII. For C1R5 construct, EcoRI and NhEI were used. For C28R construct, EcoRI was used.

After confirming the correct constructs (See Figure 13), they were transformed into competent ET12567 cells as described in the methods (See Method 3.10).

Conjugations with the ET12567 cells w. regulator constructs and the *S. bambergiensis* MoeD followed. The conjugation was carried out as described in the methods (See Method 3.11).

Mutants containing integrated vectors for overexpression of regulatory genes were designated *S. bambergiensis* MoeD C1R4, *S. bambergiensis* MoeD C1R5 and *S. bambergiensis* MoeD C28R.

Spore suspensions were made of all mutants and stored at -80°C.

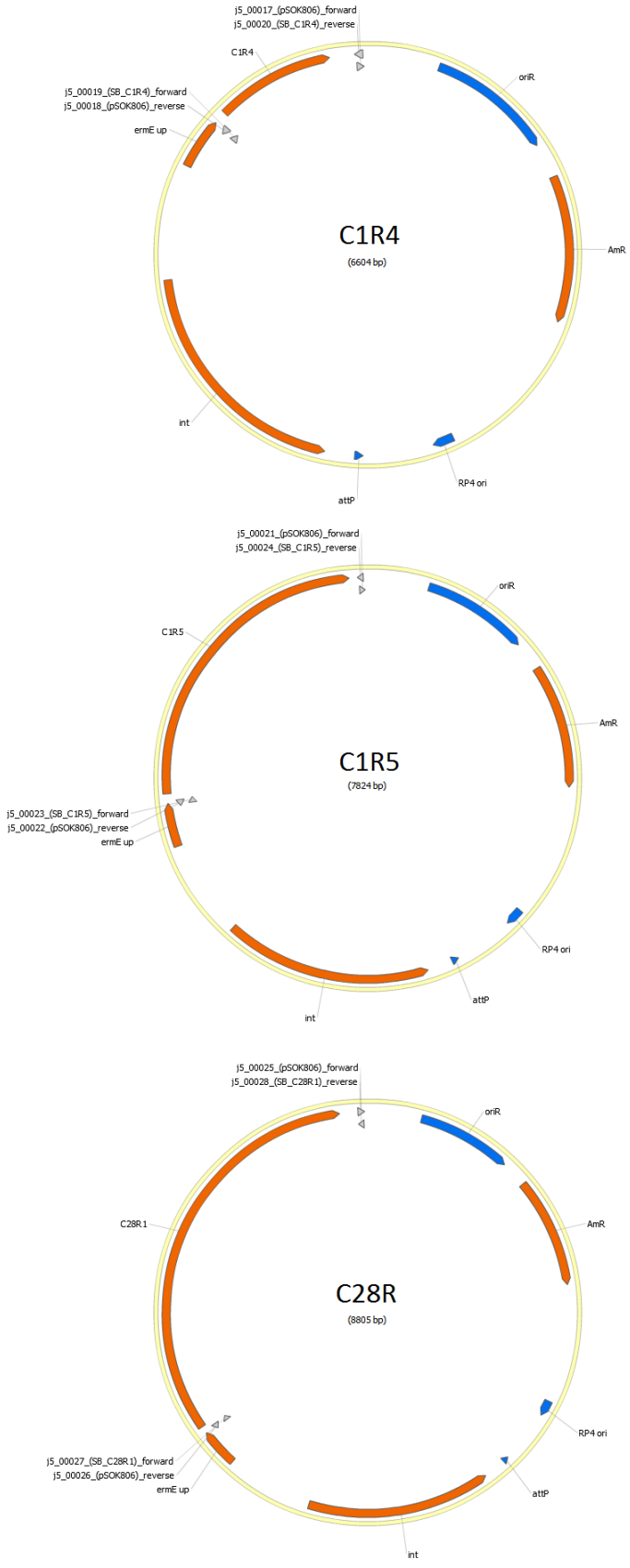


*Figure 13: Assembled regulator constructs containing the attachment site, ermE*p strong constitutive promoter, the regulatory gene, origin of replication, AmR and the integrase gene.*

3.15 Specific Digestion of DNA by Use of Restriction Endonucleases

In order to confirm that the plasmid constructs made in this project were the correct ones, restriction endonucleases were used to cut the constructs. Using Clone Manager, appropriate restriction enzymes were chosen, based on the plasmids recognition sites, to cut the molecules into 2 – 3 fragments. The fragments were then applied to a gel and separated.

Restriction enzymes are naturally produced by bacteria in order to defend against infection by viruses or other mobile genetic elements (MGEs). They do this by degrading the DNA of the MGEs. The restriction enzyme binds to DNA at a specific sequence of nucleotides or what is called the recognition sequence. The enzyme cuts both strands of the DNA within the sequence producing restriction fragments (Figure 14). (Klug, Cummings, Spencer, & Palladino, 2009)

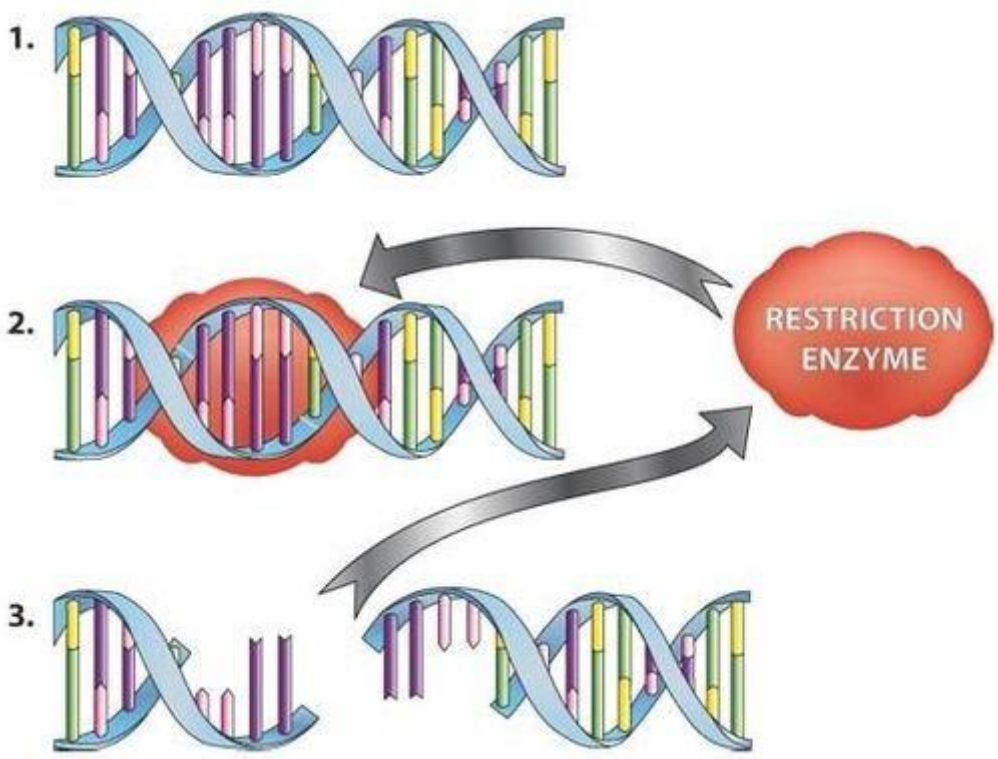


Figure 14: This figure shows how a restriction enzyme cleaves a DNA molecule. In this case leaving sticky ends. (Image taken from http://biotechlearn.org.nz/themes/dna_lab/images/restriction_enzyme)

Most restriction sequences are palindromes. Which means that the nucleotide sequence reads the same on both strands of the DNA molecule when read in the 3' to 5' direction. (Pingoud & Jeltsch, 2001)

2.0 ml of CutSmart Buffer, 0.8 ml of restriction enzyme, 3 ml of DNA and dH_2O up to 20 ml was pipetted to eppendorf tubes. These tubes were placed in 37° for 2 hours. Gel electrophoresis

was performed using 0.8% agarose containing Gel Green. DNA stain. The gel was run on 100V for 1 hour. 20 µl of sample, 3.5 µl of dye and 3 µl of ladder were used.

3.16 Primer Design

Primer design was done using the program Clone Manager version 6.0 and then the online resource j5.jbei.org.

3.17 Growth Inhibition Bio Assay

In order to get a first indication of any antimicrobial activity of the *Streptomyces* MoeD mutants, a growth inhibition bio assay was performed.

3.17.1 Bioassay with Prokaryotic and Eukaryotic Strains

E. coli (Gram - prokaryote), *Bacillus subtilis* (Gram + prokaryote) and *Saccharomyces cerevisiae* (eukaryote), were chosen as test organisms for bioassay.

E. coli and *B. subtilis* cells were incubated overnight in 2 ml of LB media at 37°C and *S. cerevisiae* was incubated overnight in 2 ml of YPD liquid media at 30°C.

3.17.2 Fermentations

100 µl of spore suspension from *S. bambergiensis* wild type, *S. bambergiensis* MoeD, *S. bambergiensis* MoeD C1R4, *S. bambergiensis* MoeD C1R5 and *S. bambergiensis* MoeD C28R were inoculated in 10 ml of medium 2XYT and incubated for 48 hours in shaking incubator at 30°C, 250 rpm.

After 48 hours, 2 ml of culture were inoculated into 50 ml of fermentation medium in 250 ml baffled flasks. These flasks were placed in a shaking incubator at 30°C, 250 rpm and incubated for 6 days.

3.17.3 Methanol Extractions

For methanol extractions, *S. bambergiensis* MoeD and *S. bambergiensis* MoeD C28R were fermented. For this purpose, 1 ml of each of the 6 day fermentations in different fermentation media (S800, 5010, 5288, 5321 and 5333) was placed into 13 ml tubes. These were centrifuged for 7 minutes at 8000 rpm. The cell pellet was resuspended in 3 ml of ethanol and glass beads were added. The tubes were placed into a shaking incubator at 16°C, 900 rpm for 60 minutes. The tubes were then centrifuged at 8000 rpm for 7 minutes. Aproximately 3 ml of supernatant was placed into eppendorf tubes.

100 µl of overnight culture of *S. cerevisiae* was spread onto a YPD plate. 100 µl of overnight cultures of *E. coli* and *B. subtilis*, were plated onto LA plates.

40 µl of methanol extracts were placed on sterile paper disks, allowed to dry out for 15 minutes in a sterile hood, and then the disks were set on the plates with test organism. Plates were then incubated overnight at 30°C for *S. cerevisiae* and 37°C for *E. coli* and *B. subtilis.*

3.17.4 Butanol Extraction

1 ml of each of the 6 day fermentation (PM4 medium) was placed into 13 ml tubes. 0.5 µl of n-butanol was added and the tubes were vortexed for 1 minute. The mixes were transferred into eppendorf tubes and spun for 3 minutes at 13,000 RPM. About 0.4 ml was taken from the upper phase (extracts) placed in new eppendorf tubes.

3.18 UV Scans

To get a first indication of novel compound production in recombinant strains, Ultraviolet (UV) scans were used.

Using the 6 day fermentations, 4 ml were extracted and placed in 10 ml tubes. 2 ml of n-butanol was added to each tube and they were spun down for 3 minutes at 10,000 RPM. About 1.5 ml of the upper phase was moved to eppendorfs.

S. bambergiensis MoeD was used as reference or control sample.

1 ml of the upper phase liquid was used for the scan. The scan range was 220 – 650 nm, with a step increment of 5 nm.

S. bambergiensis MoeD C1R4, *S. bambergiensis* MoeD C1R5 and *S. bambergiensis* MoeD C28R were sampled.

3.19 HPLC

Two strains of *S. bambergiensis* (WT and MoeD) were sent to the University of Napoli for High Performance Liquid Chromatography or HPLC.

HPLC stands for High-Performance Liquid Chromatography. HPLC is a technique used to separate, identify and quantify components in a mixture. Sample is run through a column filled with solid adsorbent material. Depending on the interaction of the component to the adsorbent material in the column, different flow rates for different compounds result and this leads to the separation of the components as they flow through the column. (Gerber, et al., 2004)

At Napoli, the 2 strains were cultivated using PM4 medium (inoculum in TSB + 10% Sucrose). An extraction was performed of the culture medium after sonication with BuOH. The organic layers were dried, resuspended in MeOH at 10 mg/ml and subjected to LC-MS (Liquid

Chromatography-Mass Spectrometry). A 5 μm Kinetex column (100x2.1 mm), maintained at room temperature, was used. Injection: 5 μl, Gradient program at 200 μl min^{-1}: 10% MeOH 3 min, 10-95% MeOH over 25 min, 95% MeOH 15 min. Negative ion mode, m/z 150-2000.

The BuOH extracts were dried and resuspended in MeOH at 1 mg/ml and subjected to LC-MS in the same conditions as above.

HPLC for the methanol extractions was carried out at the NTNU facilities. Methanol extractions from *S. bambergiensis* MoeD and *S. bambergiensis* MoeD C28R were analysed.

4 Results

4.1 Genome Analysis for SM Gene Clusters

Using antiSMASH, an annotation sheet of all the gene clusters was made.

BLASTing genes in gene cluster 2 showed homology to moenomycin associated genes in the well explored moenomycin biosynthetic pathway of *S. ghanaensis.* (Ostash & Walker, 2010)

AntiSMASH analysis of *S. bambergiensis* showed that gene cluster 28 was likely to be a gene cluster for biosynthesis of a putative 66-membered macrolide. Cluster 28 is a type I PKS, contains 30 genes and spans over 190 kb. This would make it one of the largest polyketide biosynthetic gene cluster identified to date. (Laureti, et al., 2011) Gene cluster 28 also contains 157 enzymatic domains organized into 23 modules. This cluster was specifically selected for targeted activation. An antiSMASH representation of the polyketide synthase is seen in the following figure (Figure 15).

Figure 15: AntiSMASH representation of the polyketide synthase involved in biosynthesis of putative macrolide in gene cluster 28. It contains 157 domains contained in 23 modules. These all have specific roles in assembling the secondary metabolite scaffold.

Cluster 1, a non-ribosomal peptide synthetase (NRPS), was also selected for targeted activation. Non-ribosomal peptides are SM and often antibiotics. (Finking & Marahiel, 2004) BLASTing genes of gene cluster 1 showed low homology to genes in the NCBI databank.

4.2 Construction of *S. bambergiensis* Moenomycin Deletion Mutant

The construction of the *S. bambergiensis* moenomycin deletion mutant can be split into 6 main steps:

- PCR amplification of pSOK808 vector part and the A and B parts, flanking the moenomycin gene cluster;
- Gibson assembly of the three parts, transformation of Gibson mix into competent *E. coli* cells and selection of clones containing assembled vector;
- Transformation of moenomycin deletion vector into non-methylating *E. coli* strain ET 12567 , followed by conjugation into *S. bambergiensis* wild type, and selection of a first crossover mutants
- Propagation of the first crossover mutants without selective pressure for several generations
- GusA assay for selection of second crossover mutant(s)
- PCR to confirm that the moenomycin gene cluster was deleted.

4.2.1 PCR of Vector Part and the A and B parts

Using the j5 (Hillson, 2011) and Clone manager tools, PCR primers were designed.

These primers were used in the PCR of pSOK808 vector part and the A and B fragments flanking the moenomycin gene cluster. Gel electrophoresis was run to check the PCR products. The sizes should be 1703 bp for the A part, 1602 bp for the B part and 5341 bp for the vector part (Figure 16).

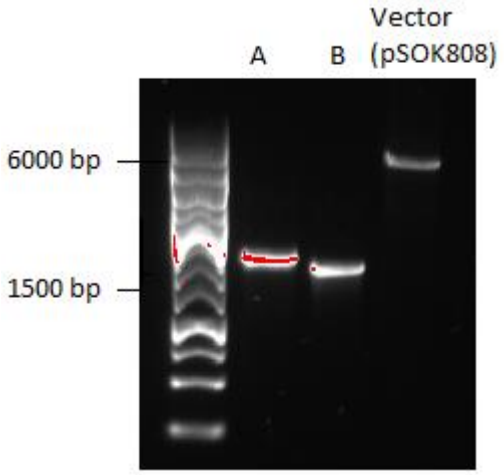


Figure 16: Gel electrophoresis of the A, B and vector parts for the moenomycin deletion vector. Part A was 1703 bp, part B 1602 and the vector part (pSOK808) 5341 bp in sizes.

4.2.2 Gibson Assembly

Using the Gibson assembly method (See Method 3.9), the three parts were combined and transformed into competent *E. coli* DH5α cells. Confirmation of the correct clone was done by isolating the plasmid DNA and running a restriction analysis using *Bam*HI, which cuts this construct at 3 sites and gives three bands of sizes 1022 bp, 1944 bp and 5681 bp on a gel.

Correct constructs are seen in samples 2, 4 and 7 in the following figure (Figure 17).

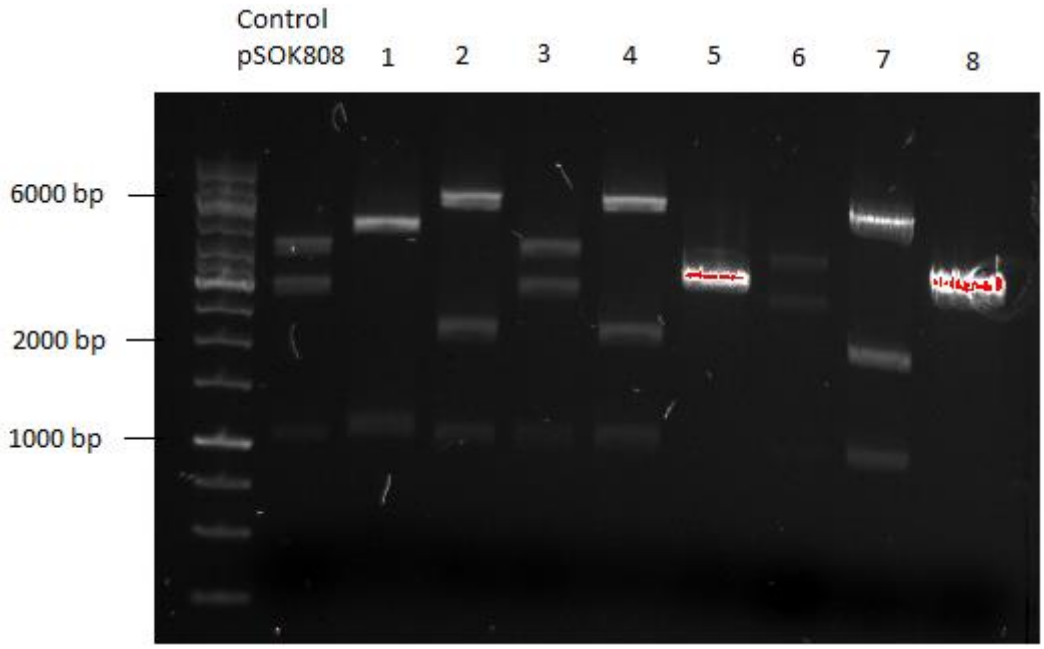


Figure 17: Gel picture of restriction assay of moenomycin deletion construct. The restriction enzyme used was BamHI, which cuts the correct construct at three sites. Wells are counted after the control. In wells 2, 4 and 7 the correct clones were observed.

Correct plasmid (pMoeD) was transformed into competent *E. coli* ET12567 cells, where from it was conjugated into *S. bambergiensis* WT.

4.2.3 Transformation and Conjugation

Moenomycin deletion vector was transformed into non-methylating *E. coli* strain, ET12567. This was followed by conjugation into *S. bambergiensis* wild type, and selection of a first crossover mutants.

4.2.4 Propagation of First Crossover Mutants

Propagation of the first crossover mutants was done without selective pressure for several generations as explained in the methods (See Method 3.12).

4.2.5 GusA Assays

The insertion of the pMoeD into the chromosome of *S. bambergiensis* is done via a single crossover, where the plasmid is integrated into the DNA of the *S. bambergiensis* by homologous recombination. For the double crossover event, the *S. bambergiensis* was incubated for three rounds of sporulation, as explained in the methods (See Method 3.12). After three rounds, the strains were checked for moenomycin deletion by using the GusA assay, as explained in the methods (See Method 3.13).

Colonies that had undergone the double crossover event would show no colour in the assay, while the ones that remained unchanged and kept the plasmid integrated into the chromosome would retain the ability to utilize the GusA substrate, thus turning blue (Figure 18).

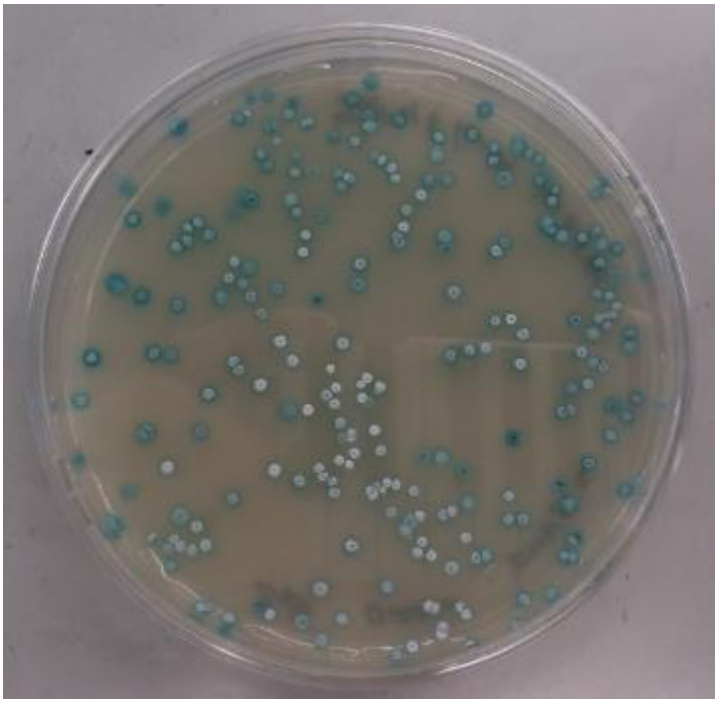


Figure 18: GusA assay. The colonies that did not show a blue colour were the ones that had putatively undergone the 2nd crossover. They were incubated on new ISP4 plates.

4.2.6 PCR of "Scar" Region of *S. bambergiensis* MoeD Double Crossover Mutants

Using the forward primer for the A part and the reverse primer for the B part, a PCR was done using the genomic DNA isolated from cultures representing the white colonies.

The expected "scar" was 3305 bp in size and the correct clones were confirmed using gel electrophoresis of PCR products. The following gel figure (Figure 19) shows that in wells 1, 3 and 4, the correct PCR product is observed.

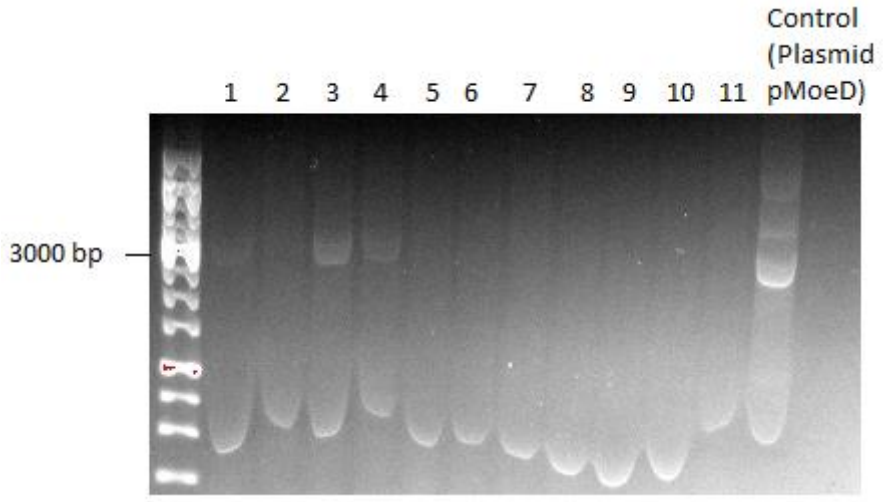


Figure 19: Gel showing the scar of the moenomycin cluster deletion in wells 1, 3 and 4. This scar is about 3300 bp in size and consists of the A + B parts of the original moenomycin gene cluster. Plasmid pMoeD was used as a template for positive control PCR.

4.3 LC-MS of Butanol Extracts

The *S. bambergiensis* moenomycin deletion mutant (MoeD) along with the wild type *S. bambergiensis* (WT) and moenomycin standard (flavomycin) were sent to the University of Napoli for liquid chromatography mass spectrometry (LC-MS). This was done to confirm the deletion of moenomycin gene cluster in *S. bambergiensis.*

Moenomycin has an m/z of 789.8233. Both the flavomycin standard and the *S. bambergiensis* WT showed such peaks. No moenomycin analogue was detected in the MoeD strain as can be seen in the following figure (Figure 20).

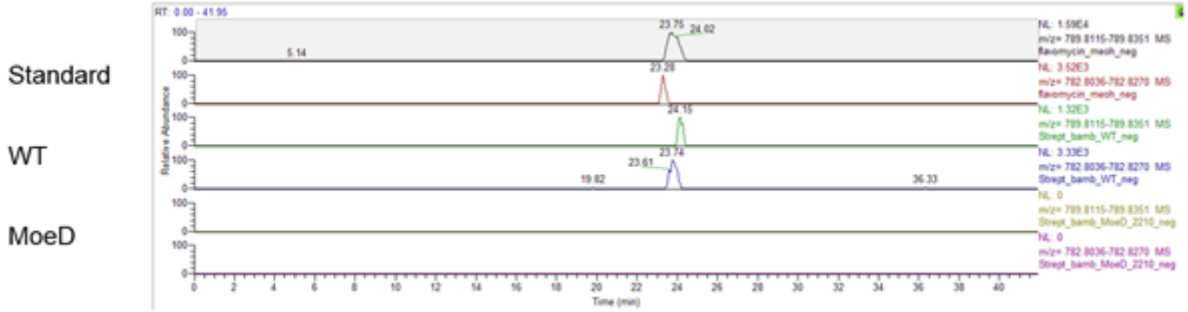


Figure 20: Moenomycin has an m/z 789.8233. Here a peak for this m/z can be seen at 24 minutes for the WT and the flavomycin standard. The MoeD mutant strain shows no such peak.

4.4 Construction of *S. bambergiensis* Regulator Mutants

Three *S. bambergiensis* MoeD mutants, MoeD C1R4, MoeD C1R5 and MoeD C28R, have been obtained by introduction of vectors, containing regulatory genes under control of strong

constitutive promotor *ermE*p* into *S. bambergiensis* MoeD mutant strain, non-producer of moenomycin.

Gene cluster 1 contains 5 regulatory genes (See Annex 2). Regulators 4 and 5 were chosen for overexpression. Regulator 4 is a LuxR family regulator, a regulatory gene known to be involved in the biosynthesis of antibiotics. (Chen & Xie, 2011) Regulator 5 is of the SARP family of regulators. SARP family of regulators are known activators of biosynthetic genes in gene clusters. (Aigle & Corre, 2012)

Gene cluster 28 contains 1 regulatory gene (See Annex 2), a LuxR family transcriptional regulator. As LuxR family regulators are involved in regulation of many SMs, this was a good candidate for overexpression. (Chen & Xie, 2011)

The construction of the *S. bambergiensis* regulatory mutants can be split into 3 main categories:

- PCRs of vector part and regulatory genes
- Gibson assembly of vector part and regulatory genes, and transformations of assembly mixes into *E.coli* strains
- Conjugation of confirmed overexpression vectors into *S.bambergiensis* MoeD strain

4.4.1 PCR of Vector Part and Regulatory Genes

Using the j5 programme and Clone manager tools, primers for Gibson assembly were designed.

These primers were used for the PCR of pSOK806 vector and the regulatory genes.

Sizes for the respective parts can be seen in the following table (Table 12).

Table 12: Table showing the respective sizes of each part of the regulator constructs.

Part	Full Size (bp)	Regulatory Gene Size(bp)	pSOK806 Part Size (bp)
C1R4 Construct	6604	850	5754
C1R5 Construct	7824	2070	5754
C28R Construct	8805	3051	5754

Gel electrophoresis was run to see if the PCR products were of correct sizes (Figures 21 and 22).

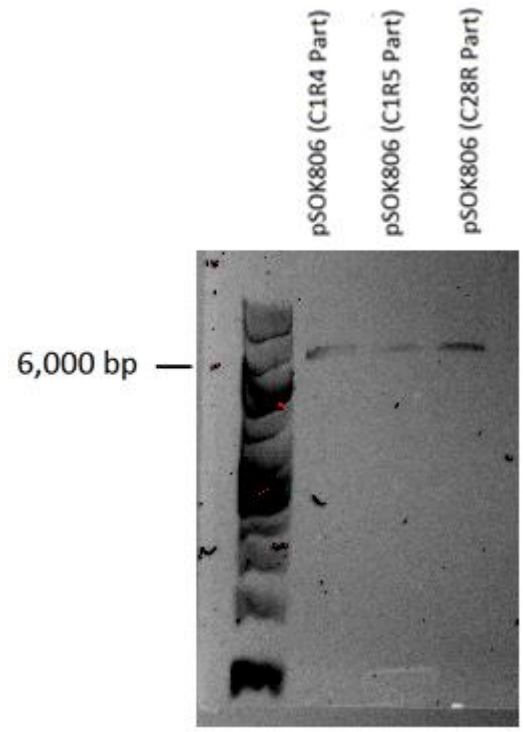


Figure 21: Gel electrophoresis of pSOK806 parts. This figure confirms the sizes of PCR products (5754 bp).

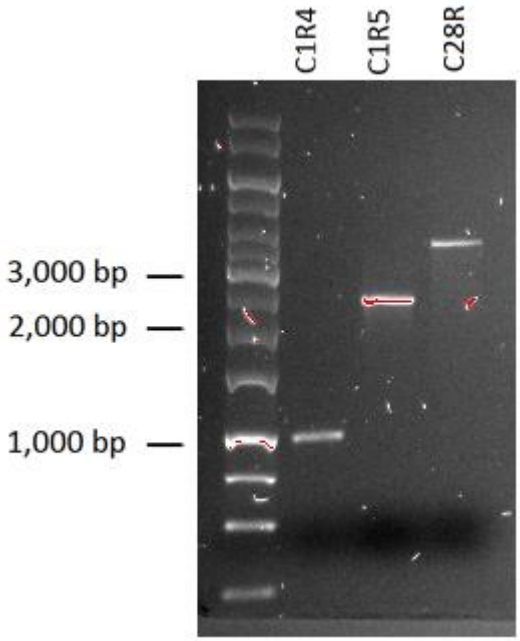


Figure 22: Gel electrophoresis of regulatory gene parts. This figure shows the PCR products of correct sizes, 850 bp for C1R4, 2070 bp for C1R5 and 3051 bp for C28R.

4.4.2 Gibson Assembly of Vector Part and Regulatory Gene

Using the Gibson assembly method, the respective two parts were combined and transformed into competent *E. coli* DH5α cells. Confirming the correct clone was done by isolating the plasmid DNA and running a restriction analysis using appropriate restriction enzymes.

Construct pC1R4 was cut with PvuII and this would cut the DNA molecule into two pieces, yielding 4578 bp and 2029 bp fragments (Figure 23).

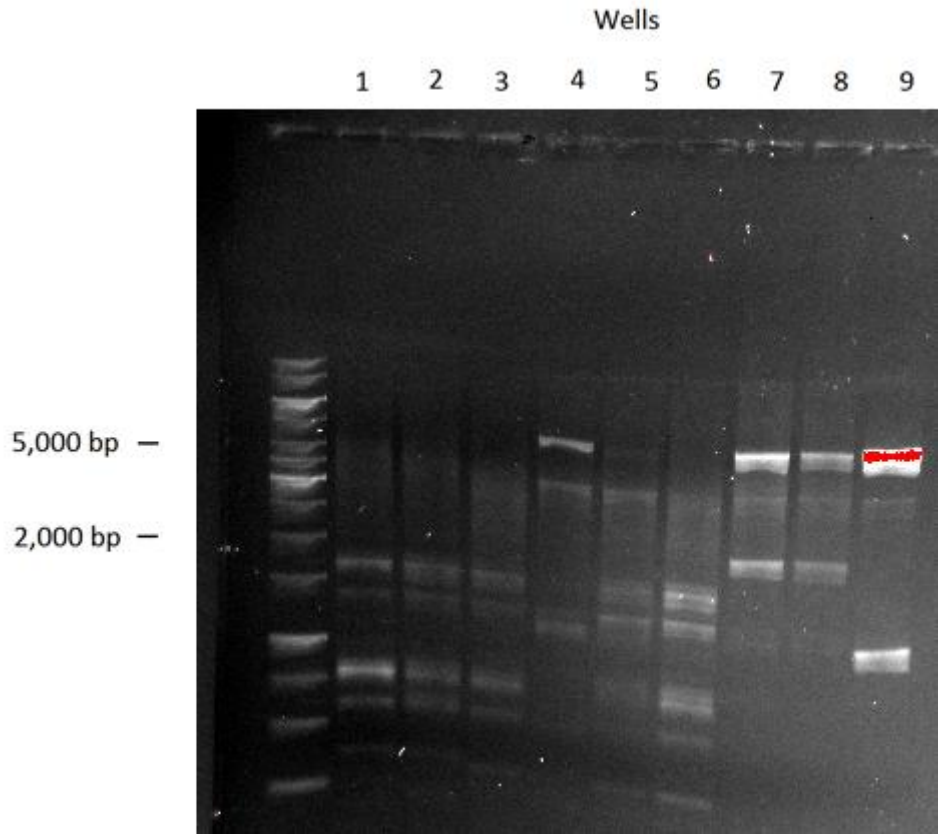


Figure 23: Restriction assay for regulator C1R4. Wells 7 and 8 contain the correct clones as they are of appropriate sizes (4575 and 2029 respectively).

Construct pC1R5 was cut with EcoRI and NhEI and this would cut the DNA molecule into two pieces, yielding 5541 bp and 2283 bp fragments (Figure 24).

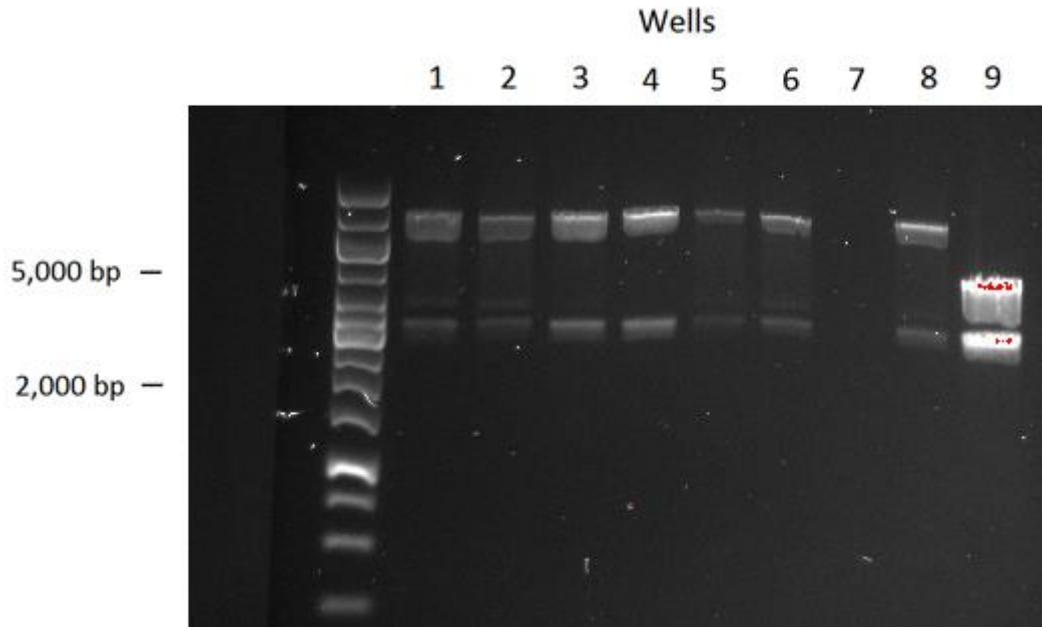


Figure 24: Restriction assay for regulator C1R5. Wells 1-6 show correct clones. For this research, clones # 3 and 4 were used. These clones were of appropriate sizes (2283 and 5541 respectively).

Construct pC28R was cut with EcoRI. After restriction, fragments of 5781 bp and 3024 bp should be observed on a gel (Figure 25).

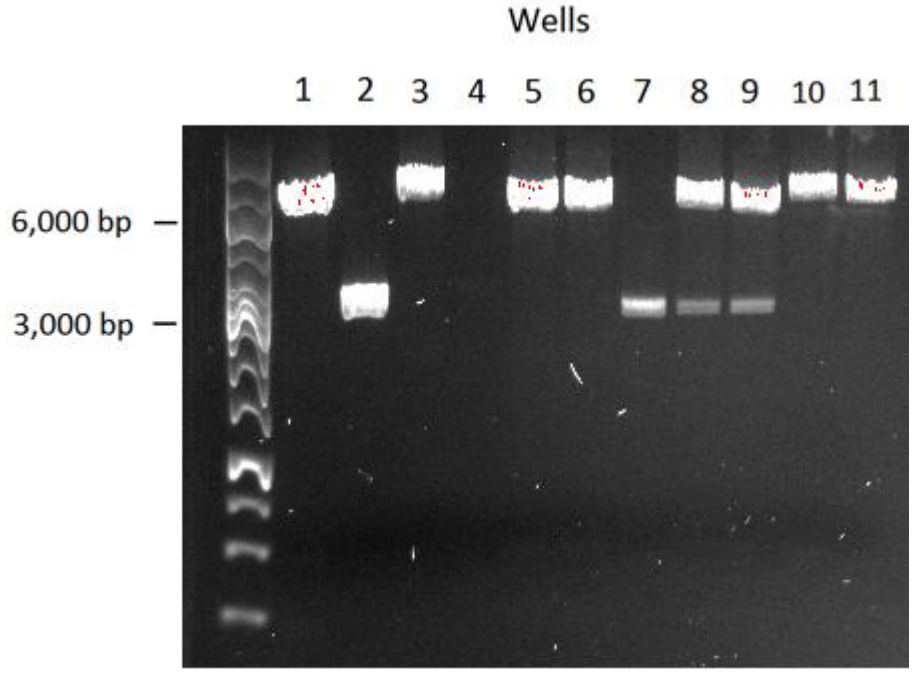


Figure 25: Restriction assay for pC1R4. Wells 8 and 9 contain the DNA fragments of correct sizes (5781 and 3024 respectively).

4.4.3 Conjugation of Constructs for Expression of Regulatory Genes into *S. bambergiensis*

As a totally non-methylating *E. coli* host is needed as DNA donor for conjugation into *Streptomyces*, constructs containing regulatory genes were transformed into competent *E. coli* ET12567 cells. Cells were incubated on LA plates with Am_{100}, as the vector has an Am – resistance marker, and then on plates with Am, Cml and Kan to support both vector with regulatory gene and helper plasmid pUZ8002. (Kieser T. , Bibb, Buttner, Chater, & Hopwood, 2000)

The vectors were then introduced into *S. bambergiensis* MoeD strain. *Streptomyces* clones, which received plasmid during conjugation, were selected by Am_{50}, and Nal_{30} was used to get rid of *E. coli* cells.

4.5 UV Scans

In order to get first indications of novel products, UV scans of butanol extractions of *S. bambergiensis* MoeD C1R4, *S. bambergiensis* MoeD C1R5, *S. bambergiensis* MoeD C28R were done.

Strains were fermented for 6 days as described in the methods and UV scans were made using butanol extractions from those fermentations.

For UV scans, the moenomycin deletion mutant, *S.* bambergiensis MoeD, was used as control. No peaks were seen from the butanol extractions of *S. bambergiensis* MoeD C1R4 and *S. bambergiensis* MoeD C1R5, but a peak was obsereved from the butanol extractions of *S. bambergiensis* MoeD C28R as can be seen in the following figure (Figure 26).

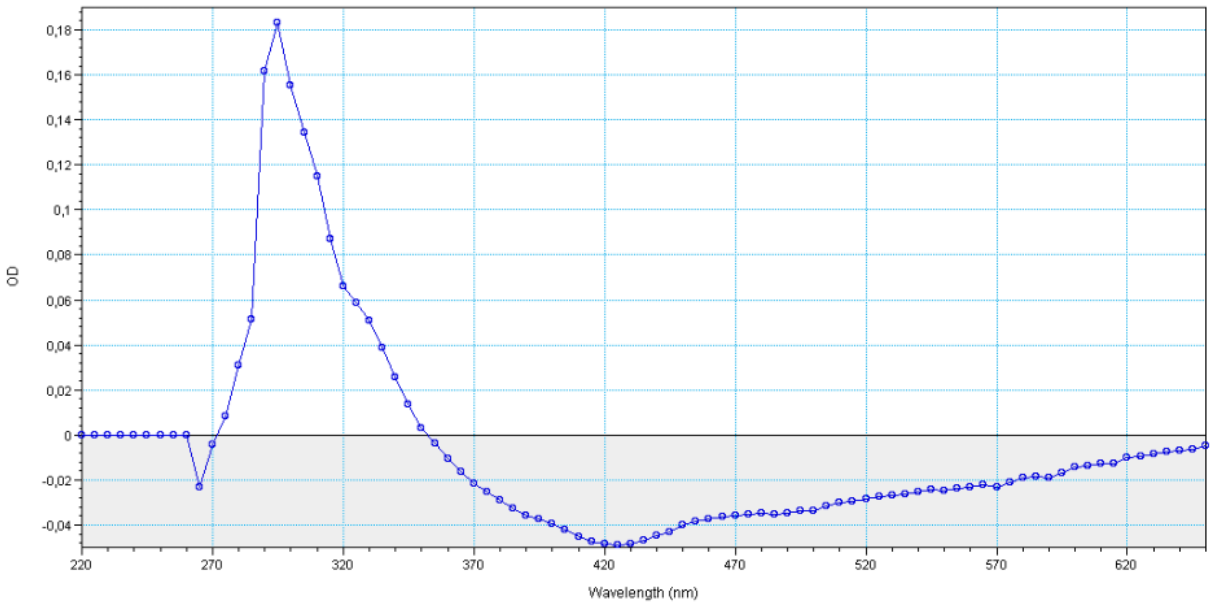


Figure 26: An absorbance peak of about 0.19 was observed at about 290 nm.

4.6 Growth Inhibition Bio Assay

For growth inhibition bio assay, methanol extracts from the cells of recombinant *Streptomyces* strains were used.

Growth inhibition can be seen by the formation of clear zones around putative bio active substance placed on a lawn of an indicator strain. Methanol extracts were placed on paper disks, which were dried and then transferred onto plates containing *E. coli* (Gram negative), *B. subtilis* (Gram positive) and *S. cerevisiae* (Eukaryote), respectively. Zone of inhibition was observed in LA plates with *B. subtilis* and paper disks with methanol extraction of *S. bambergiensis* MoeD C28R, fermented in media S800 and 5288. Small activity can be seen on the LA plates with *B. subtilis* and paper disks with methanol extraction of *S. bambergiensis* MoeD, fermented in media 5288 suggesting that the target gene cluster may be weakly expressed in this medium (Figure 27).

No inhibition zones were observed on *S. cerevisiae* and *E. coli* plates.

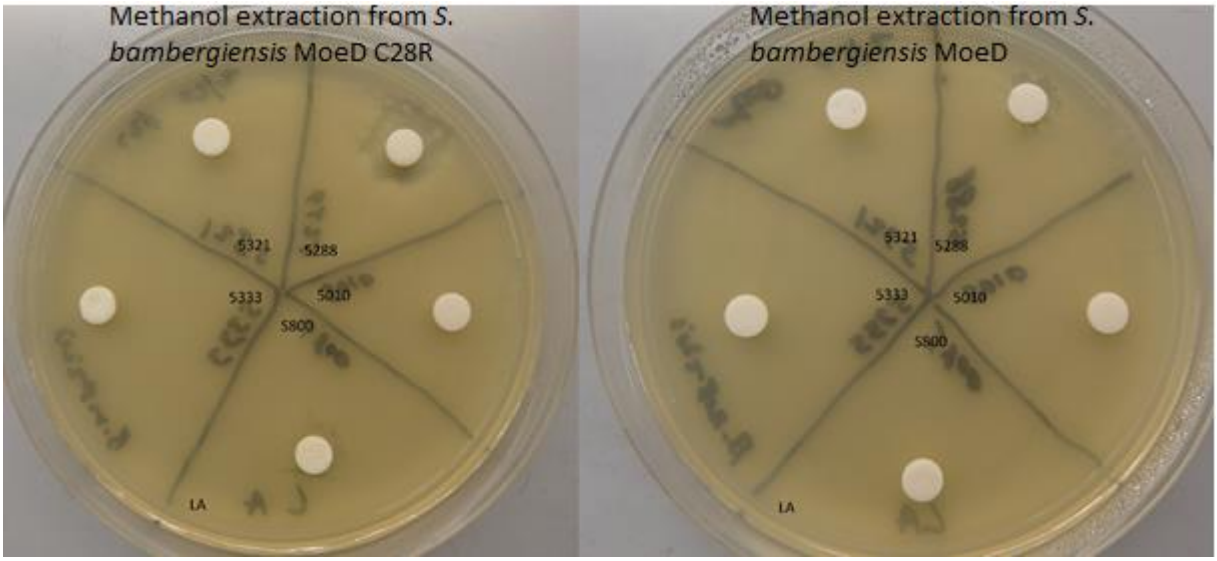


Figure 27: Figure showing the LA plates with B. subtilis. Paper disks were inoculated with methanol extractions from S. bambergiensis MoeD and S. bambergiensis MoeD C28. Inhibition zones formed in methanol extraction of S. bambergiensis MoeD C28 from S800 and 5288 fermentation media, respectively, indicating antibiotic activity. Some activity is seen in media 5288 on the plate with the S. bambergiensis MoeD extractions.

4.7 LC-MS of the Methanol Extracts

LC-MS was done with the methanol extractions of *S. bambergiensis* MoeD and *S. bambergiensis* MoeD C28R, from the 5288 fermentation media. *S. bambergiensis* MoeD was used as control.

As can be seen in the following figure (Figure 28), a clear peak was observed on LC at 3.739 minutes in the *S. bambergiensis* MoeD C28R extract, which was absent in the control strain.

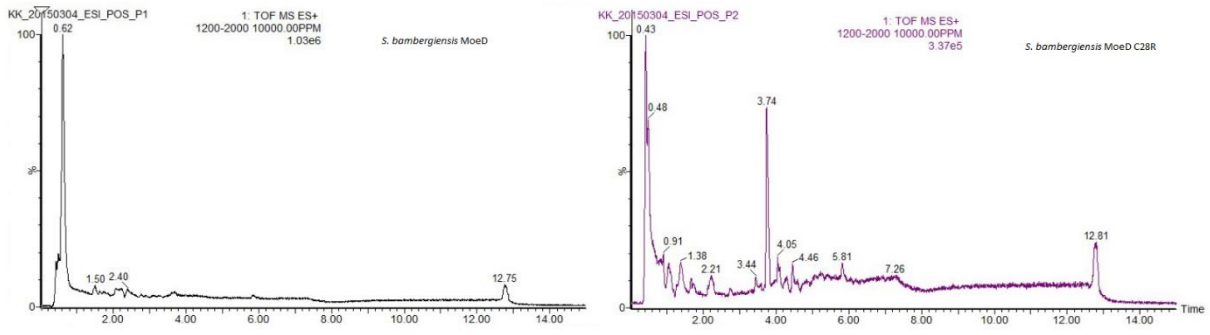


Figure 28: This figure shows the chromatogram comparison between the control, S. bambergiensis MoeD, extraction (left) and S. bambergiensis MoeD C28R extraction (right). A clear peak can be seen at the 3.739 minute mark in the C28R extraction. Other smaller peaks.

Analysing the peak at the 3.739 minute mark gives a m/z ratio of 1563 for the molecule. This can be seen in the following figure (Figure 29).

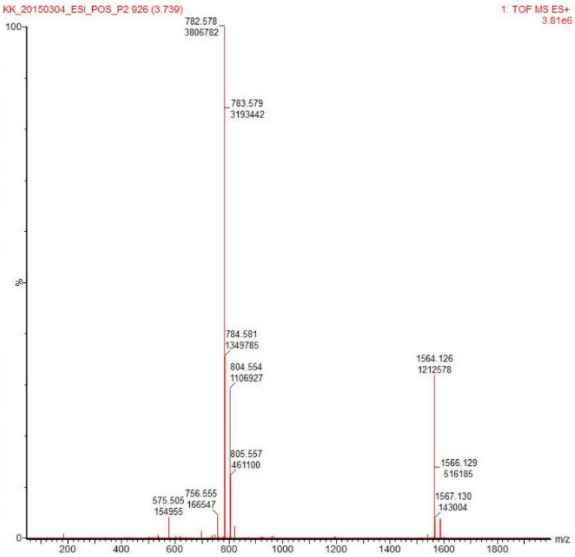


Figure 29: Analysing the spectra of the peak at the 3.739 mark, shows that the molecule has an m/z ratio of 1563.

5 Discussion

5.1 Genome Mining for Interesting Gene Clusters

The potential of each microorganism can be vast as it becomes apparent that each strain possesses a multitude of biosynthetic gene clusters (Figure 4). (Doroghazi & Metcalf, 2013) Many of these gene clusters are not expressed under laboratory conditions. Activating these cryptic pathways for the biosynthesis of novel SMs is an innovative way to discover novel bioactive compounds. There are a number of methods currently employed by scientists in order to activate silent gene clusters. Ribosome engineering is one method, where introduction of a mutation in the RNA polymerase has led to activation of biosynthetic pathways. (Ochi & Hosaka, 2013) Co-cultivation is another exciting method used for inducing production of silent metabolites. Cultivating two bacterial strains together mimics in a way conditions encountered in nature. This has been shown to be an effective method to discover novel bioactive compounds. (Ochi & Hosaka, 2013) In this project, the method of constitutive overexpression of putative pathway-specific regulators was chosen as the method for waking up silent gene clusters. The incentive for choosing this method was that the whole genome sequence of *S. bambergiensis* was recently obtained and that this bacteria produces only one known SM, the antibiotic moenomycin. (Ostash & Walker, 2010) *Streptomyces* bacteria are historically proliferate producers of biologically active SMs and so *S. bambergiensis* has a potential to biosynthesize a number of different SMs. (Bérdy, 2005)

Genome mining is a term applied to the use of bioinformatics, molecular genetics and natural product analytical chemistry in order to access the metabolic product of a gene cluster found in the genome of an organism. (Gomez-Escribano & Bibb, 2014) Now, in the age of genomics, thousands of gene clusters are becoming available. With the overflow of information, there is a need to filter out the gene clusters that are interesting as possible natural products. Using software and databanks such as AntiSMASH and NCBI, the task of filtering the interesting gene clusters beomes much less laborious, yet even with the aid of such bioinformatic tools, there still are many factors to consider when deciding which gene clusters to research further.

Currently two methods are used for genome mining in microorganisms, heterologous expression and homologous expression, which is the method employed in this project.

Heterologous expression is the cloning of a fragment of the producing organisms genome, containing the gene cluster, and expressing it in a suitable host. (Gomez-Escribano & Bibb, 2014) This can enable characterisation of biosynthetic pathways in a genetically amenable host.

(Ongley, Bian, Neilan, & Muller, 2013) There are generally many different rationales for choosing to heterologously express certain genes, such as slow growth rate and difficulty in laboratory cultivation of original strain, host strain being well characterised and proving to be a superior producer to original strain, for instance (Jones, et al., 2012) There are also a number of difficulties associated with heterologous expression. The compound produced might be too toxic for new host, some additional cofactors and interactions might be missing. Improper post-translational folding of proteins in heterologous host can be problematic also and as many gene clusters are very large, the transfer of genetic material to another host can be very difficult. (Ongley, Bian, Neilan, & Muller, 2013)

Obviously many of these challenges are not an issue during homologous expression. Other problems arise instead, such as poorly characterized regulator pathways and regulators may not be controlling expression of all necessary genes in the cluster. (Bibb M. J., 2005)

Homologously overexpressing genes can be done in one of two ways, by over expressing cluster specific activators or by inactivating repressors. (Zotchev, Sekurova, & Katz, 2012) In this project, the goal was to identify biosynthetic gene clusters of *S. bambergiensis* for potentially novel SMs and activate these gene clusters via overexpression of their pathway-specific regulators. After careful in silico mining using the bioinformatics tool AntiSMASH and the BLAST feature of the NCBI website, the gene clusters chosen for targeted activation were clusters 1 and 28. In both these clusters the regulators in question were activators.

According to in silico mining using AntiSMASH, gene cluster 1 of *S. bambergiensis* is a non-ribosomal peptide synthetase. The fact that many known antibiotics are produced by non-ribosomal synthetases (Finking & Marahiel, 2004) and low homology to genes in the NCBI databank, made this cluster an interesting choice for targeted activation.

AntiSMASH analysis of gene cluster 28 showed that this cluster was a type I polyketide synthase gene cluster. Polyketides are a large class of diverse compounds. Many important antibiotics belong to the polyketide class, including macrolides. (Alekhova & Novozhilova, 2001) The putative polyketide produced by the polyketide synthase in this cluster would be a 66-membered macrolide, the largest identified to date. The cluster itself spans over 190 kb and contains 30 putative biosynthetic genes.

5.2 Removing Background for the Discovery of Novel Compounds

For the discovery of novel bioactive compounds, it is necessary to eliminate as much background interference as possible. The only known bioactive compound produced in the case

of *S. bambergiensis* is moenomycin. (Ostash & Walker, 2010) Moenomycin is an inhibitor of peptidoglycan glycosyltransferases and so it inhibits cell wall biosynthesis. The only explored phosphoglycolipid biosynthetic pathway is that of *Streptomyces ghanaensis*. (Ostash & Walker, 2010) Genes with high homology to genes identified in *S. ghanaensis* relating to biosynthesis of moenomycin, were found in *S. bambergiensis* as a result of in silico mining in AntiSMASH (See Annex 1). These genes were found in gene cluster 2. Targeted gene inactivation with the help of *gusA* reporter gene is a well established method. (Siegl & Luzhetskyy, 2012) This method was used to succesfully delete gene cluster 2 from the *S. bambergiensis*, resulting in the recombinant strain, *S. bambergiensis* MoeD or moenomycin deletion mutant of *S. bambergiensis*, incapable of moenomycin production.

5.3 Selecting Regulatory Genes for Overexpression

In silico mining of *S. bambergiensis* showed that cluster 28 contains one regulatory gene. AntiSMASH analysis showed this to be a LuxR family transcriptional regulator (See Appendix 2). LuxR regulatory genes are known to be involved in the biosynthesis of antibiotics as activators of gene expression, so this was a good candidate for overexpression. (Chen & Xie, 2011) AntiSMASH analysis of cluster 1 showed that this cluster contains 5 regulatory genes (See Appendix 2). Regulatory gene 4 is a putative LuxR family transcriptional regulator and so using the same rationale as for gene cluster 28, this was a good candidate for overexpression. Regulatory gene 5 is of the SARP family of regulators. SARP family of regulators are known activators of biosynthetic genes in gene clusters. (Aigle & Corre, 2012) This made regulatory gene 5 in cluster 1 also a good candidate for overexpression.

5.4 Discovery of a Novel Bioactive Compound

Once the regulatory constructs were successfully integrated into the recombinant strain, *S. bambergiensis* MoeD, a bioassay was done to check for novel bioactive compounds. Following 6 days of fermentation, an extraction was done as explained in the methods. These extraction were used for the bioassays. The extraction of the 6 day fermentation of *S. bambergiensis* MoeD C28R showed inhibition activity against the Gram + bacteria, *B. subtilis*. No inhibition was shown against the Gram – *E. coli* or the eukaryotic *S. cerevisiae*. LC-MS of the extract showed a putative molecule with an m/z ratio of 1563.

5.5 Significance

Laureti et al discovered a 51-membered macrolide (stambomycin) by waking up a gene cluster in *S. ambofaciens*. This molecule was elucidated and found to have an m/z ratio of 1375. This macrolide was shown to have inhibitory activity against Gram + bacteria but no activity against

Gram – bacteria. Interestingly this macrolide inhibits proliferation of human adenocarcinoma cell lines. (Laureti, et al., 2011) The m/z ratio of the putative compound from this project correlates with the predicted 66-membered polyketide and has predicted inhibitory activity of a macrolide, that is, shows activity against Gram + bacteria and no activity against Gram – bacteria.

5.6 Way Forward

The compound discovered in this project has a high commercial potential. The way forward for this work would be to purify the compound and elucidate the structure of the molecule. It would be very interesting to see if this compound has antiproliferative capabilities against human cancer cell lines, as was evident with the macrolide discovered by Laureti et al.

This project focused on two gene clusters in *S. bambergiensis,* there are still numerous gene clusters unexplored and so this bacteria holds even more potential. In the course of this work two additional mutants were successfully integrated with overexpression vectors. Unfortunately these were created successfully towards the end of the project so due to time restriction, these could not be investigated thoroughly.

6 Conclusion

By waking up a silent gene cluster with the overexpression of regulators within the gene cluster, this project led to the successful production of a putative biologically active product.

The discovery of a 51-membered macrolide by Laureti et al in 2011 led to the identification of stambomycins A-D. These have shown moderate activity against Gram positive bacteria, but more importantly, they inhibited proliferation of human adenocarcinoma (HT29) cell lines. Laureti et al also observed significant antiproliferative activities against the human breast (MCF7), lung (H460) and prostate (PC3) cancer cell lines. This would indicate that stambomycins might represent promising new leads for anticancer drug discovery. (Laureti, et al., 2011)

There is promising evidence that the *S. bambergiensis* MoeD C28R strain is producing the antiSMASH predicted 66-membered macrolide. Further identification of this molecule and tests against human cell lines might prove very interesting.

7 References

Aigle, B., & Corre, C. (2012). Waking up Strepomyces Secondary Metabolism by Constitutive Expression of Activators or Genetic Disruption of Repressors. In D. A. Hopwood, *Methods in ENZYMOLOGY vol 517* (pp. 344-366). Elsevier Inc.

Alekhova, T. A., & Novozhilova, T. Y. (2001). Biosynthesis of Polyketide Anibiotics by Various Streptomyces Species that Produce Actinomycins. *Applied Biochemistry and Microbiology*, 267-273.

Bentley, S. D., Chater, K. F., Cerdeno-Tarraga, A. M., Challis, G. L., Thomson, N. R., James, K. D., . . . Hopwood, D. A. (2002). Complete genome sequence of the model actinomycete Streptomyces coelicolor A3 (2). *Nature*, 141-147.

Bertrand, S., Bohni, N., Schnee, S., Schumpp, O., Gindro, K., & Wolfender, J. L. (2014). Metabolite induction via microorganism co-culture: A potential way to enhance chemical diversity for drug discovery. *Biotechnology Advances*, 1180-1204.

Bérdy, J. (2005). Bioactive Microbial Metabolites. *Journal of Antibiotics*, 1-26.

Bibb, M. J. (2005). Regulation of secondary metabolism in streptomycetes. *Current Opinion in Microbiology*, 208-215.

Bibb, M., & Hesketh, A. (2009). Analyzing the Regulation of Antibiotic Production in Streptomyces. In H. D. (editor), *Methods in Enzymology Volume 458* (pp. 93-116). Burlington: Elsevier Inc.

Blin, K., Medema, M. H., Kazempour, D., Fischbach, M. A., Breitling, R., Takano, E., & Weber, T. (2013). antiSMASH 2.0 - a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Research*, 1-9.

Bode, H. B., Bethe, B., Höfs, R., & Zeeck, A. (2002). Big Effects from Small Changes: Possible Ways to Explore Nature´s Chemical Diversity. *ChemBioChem*, 619-627.

Challis, G. L., & Hopwood, D. A. (2003). Synergy and contingency as driving forces for the evolution of multiple secondary metabolite production by Streptomyces species. *Proc Natl Acad Sci*, 14555-61.

Chen, I., & Dubnau, D. (2004). DNA uptake during bacterial transformation. *Nature Reviews*, 241-249.

Chen, J., & Xie, J. (2011). Role and Regulation of Bacterial LuxR-Like Regulators. *Journal of Cellular Biochemistry*, 2694-2702.

Das, S., Lyla, P. S., & Khan, S. A. (2008). Distribution and generic composition of culturable marine actinomycetes from the sediments of Indian continental slope of Bay of Bengal. *Chinese Journal of Oceanology and Limnology*, 166-177.

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

Finking, R., & Marahiel, M. A. (2004). Biosynthesis of Nonribosomal Peptides. *Annual Review Microbiology*, 453-488.

Fleming, A. (1929). On the Anitibacterial Action of Cultures of a Penicillium with Special Reference to their use in the isolation of B. Influenze. *British Journal of Experimental Pathology*, 226 - 236.

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

Gaynor, M., & Mankin, A. S. (2003). Macrolide Antibiotics: Binding Site, Mechanism of Action, Resistance. *Current Topics in Medicinal Chemistry*, 949-961.

Gerber, F., Krummen, M., Potgeter, H., Roth, A., Siffrin, C., & Spoendlin, C. (2004). Practical aspects of fast reversed-phase high-performance liquid chromatography using 3 μm particle packed columns and monolithic columns in pharmaceutical development and production working under current good manufacturing practice. *Journal of Chromatography*, 127-133.

Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison III, C. A., & Smith, H. O. (2009). Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases. *Nature*, 343 - 347.

Gomez-Escribano, J. P., & Bibb, M. J. (2014). Heterologous expression of natural product biosynthetic gene clusters in Streptomyces coelicolor: From genome mning to manipulation of biosynthetic pathways. *Journal of Indus trial Microbiology and Biotechnology*, 425-431.

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*, 2343-53.

Guerra, S. M., Rodriquez-Garcia, A., Santos-Aberturas, J., Vicente, C. M., Payero, T. D., Martin, J. F., & Aparicio, J. F. (2012). LAL Regulators SC00877 and SC07173 as Pleiotropic Modulators of Phosphate Starvation Response and Actinorhodin Biosynthesis in Streptomyces coelicolor. *PLOS one*, 1-11.

Hillson, N. J. (2011). *https://j5.jbei.org/index.php/Main_Page*. Retrieved from https://j5.jbei.org/index.php/Main_Page: https://j5.jbei.org/index.php/Main_Page

Huang, M., Li, M., Feng, Z., Liu, Y., Chu, Y., & Tian, Y. (2011). Enhanced rapamycin production in Streptomyces hygroscopicus by integrative expression of aveR, a LAL family transcriptional regulator. *World J Microbiol Biotechol*, 2103-2109.

Hughes, J. M. (2011). Preserving the Lifesaving Power of Antimicrobial Agents. *Journal of the American Medical Association*, 1027-1028.

Infectious Diseases Society of America. (2013). *Facts on Antibiotic Resistance.* IDSA.

Jenke-Kodama, H., Sandmann, A., Muller, R., & Dittmann, E. (2005). Evolutionary Implications of Bacterial Polyketide Synthases. *Molecular Biological Evolution*, 2027-2039.

Jones, A. C., Ottilie, S., Eustaquio, A. S., Edwards, D. J., Gerwick, L., Moore, B. S., & Gerwick, W. H. (2012). Evaluation of Streptomyces coelicolor A3(2) as a heterologous expression host for the cyanobacterial protein kinase C activator lyngbyatoxin A. *FEBS Journal*, 1243-1251.

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

Kieser, T., Bibb, M. J., Chater, K. F., & Hopwood, D. A. (2000). *Practical Streptomyces Genetics.* Norwich: John Innes Centre.

Klug, W. S., Cummings, M. R., Spencer, C. A., & Palladino, M. A. (2009). *Concepts of Genetics.* San Francisco: Pearson Education Inc.

Laureti, L., Song, L., Huang, S., Corre, C., Leblond, P., Challis, G. L., & Aigle, B. (2011). Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in Streptomyces ambofaciens. *PNAS*, 6258-6263.

Malik, V. S. (1980). Microbial secondary metabolism. *TIBS*, 68-72.

Martin, J. F., & Liras, P. (2010). Engineering of regulatory cascades and networks controlling antibiotics biosynthesis in Streptomyces. *Current Opinion in Microbiology*, 263-273.

McArthur, M., & Bibb, M. J. (2007). Manipulating and understanding antibiotic production in Streptomyces coelicolor A3 (2) with decoy oligonucleotides. *PNAS*, 1020-1025.

Ochi, K., & Hosaka, T. (2013). New strategies for drug discovery: activation of silent or weakly expressed microbial gene clusters. *Appl Microbiol Biotechnol*, 87-98.

Ongley, S. E., Bian, X., Neilan, B. A., & Muller, R. (2013). Recent advances in the heterologous expression of microbial natural product biosynthetic pathways. *Nat. Prod. Rep*, 1121-1138.

Ostash, B., & Walker, S. (2010). Moenomycin family antibiotics: chemical synthesis, biosynthesis, biological activity. *Nat Prod Rep*, 1-42.

Pingoud, A., & Jeltsch, A. (2001). Structure and function of type II restriction endonucleases. *Nucleic Acids Research*, 3705-3727.

Ramos, J. L., Martinez-Bueno, M., Molina-Henares, A. J., Terán, W., Watanabe, K., Zhang, X., . . . Tobes, R. (2005). The TetR Family of Transcriptional Regulators. *Microbiology and Molecular Biology Reviews*, 326-356.

Schrijver, A. D., & Mot, R. D. (1999). A subfamily of MalT-related ATP-dependent regulators in the LuxR family. *Microbiology*, 1287.

Sekurova, O. N., Brautaset, T., Sletta, H., Borgos, S. E., Jakobsen, Ø. M., Ellingsen, T. E., . . . Zotchev, S. B. (2004). In Vivo Analysis of the Regulatory Genes in the Nystatin Biosynthetic Gene Cluster of Streptomyces noursei ATCC 11455 Reveals Their Different Control Over Antibiotic Biosynthesis. *Journal of Bacteriology*, 1345-1354.

Sezonov, G., Duchene, A.-M., Friedman, A., Guerineau, M., & Pernodet, J.-L. (1998). Replicase, Excisionase, and Integrase Genes of the Streptomycs Element pSAM2

Constitute an Operon Positively Regulated by the pra Gene. *Journal of Bacteriology*, 3056-3061.

Siegl, T., & Luzhetskyy, A. (2012). Actinomycetes genome engineering approaches. *Antonie van Leeuwenhoek*, 503-516.

Snustad, D. P., & Simmons, M. J. (2012). *Genetics.* Minnesota: John Wiley and Sons, Inc.

Solanki, R., Khanna, M., & Lal, R. (2008). Bioactive compounds from marine actinomycetes. *Indian Journal of Microbiology*, 410-431.

Takano, E. (2006). y-Butyrolactone: Streptomyces signalling molecules regulating antibiotic production and differentiation. *Current Opinion in Microbiology*, 287-294.

Vaishnav, P., & Demain, A. L. (2010). Unexpected applications of secondary metabolites. *Biotechnology Advances*, 223-229.

Vining, L. C. (1992). Secondary metabolism, inventive evolution and biochemical diversity - a review. *Gene*, 135-140.

Zakeri, B., & Lu, T. K. (2012). Synthetic Biology of Antimicrobial Discovery. *ACS Synthetic Biology*, 358-372.

Zhng, L., Baumann, U., & Reymond, J. L. (2004). An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Research*, 1-5.

Zotchev, S. B. (2008). Antibiotics Biosynthesis. *Wiley Encyclopedia of Chemical Biology*, 1-9.

Zotchev, S. B. (2014). Genomics-Based Insights into the Evolution of Secondary Metabolite Biosynthesis in Actinomycete Bacteria. In P. Pontarotti (ed.), *Evolutionary Biology: Genome Evolution, Speciation* (pp. 35-45). Springer International Publishing Switzerland.

Zotchev, S. B., Rozenfeld, S. M., & Zhdanov, V. G. (1988). Transformation of streptomyces bambergiensis S712 protoplasts with plasmid DNA. *Anitbiotiki i Khimioterapiya*, 200-203.

Zotchev, S. B., Sekurova, O. N., & Katz, L. (2012). Genome-based bioprospecting of microbes for new therapeutics. *Current Opinion in Biotechnology*, 941-947.

Zotchev, S. B., Soldatova, L. I., Orekhov, A. V., & Schrempf, H. (1992). Characterization of a linear extrachromosomal DNA element (pBL1) isolated after interspecific mating between Streptomyces bambergiensis and S. lividans. *Res. Microbiol.*, 839-845.

8 Appendices

8.1 Appendix 1

Annotations of gene clusters using antiSMASH, Excel and BLAST. Function is based on highest hit in BLAST.

Cluser Nr	Designation	Location	Type	Function	Of Composition
1	ctg1 orf00367	163248 - 164186	Nrps	alpha/beta hydrolase	
1	ctg1 orf00380	169005 - 169994	Nrps	taurine catabolism dioxygenase	
1	ctg1 rof00382	169991 - 177766	Nrps	glutamate racemase	A - PCP - C - A -PCP - C - A - PCP
1	ctg1 orf00385	177763 - 185877	Nrps	Non-ribosomal peptide synthetase	C - A - PCP - C - A - PCP - TE - TE
1	ctg1 orf00388	185874 - 198302	Nrps	non-ribosomal peptide synthetase	C - A - PCP - C - A - PCP - C - A - PCP - C - A - PCP
1	ctg1 orf00390	198299 - 211780	Nrps	non-ribosomal peptide synthetase	C - A - PCP - C - A - PCP - C - A - PCP - C - A - PCP
1	ctg1 orf00416	228146 - 229024	Nrps	alpha/beta hydrolase	
2	ctg1 orf00517	269783 - 270664	Nucleoside-phosphoglycolipid	NADPH:quinone reductase	
2	ctg1 orf00526	273243 - 274943	Nucleoside-phosphoglycolipid	AMP-dependent synthetase	
2	ctg1 orf00527	275090 - 276286	Nucleoside-phosphoglycolipid	5-aminolevulinate synthase	
2	ctg1 allorf003068	280371 - 281389	Nucleoside-phosphoglycolipid	MoeE5	
2	ctg1 orf00533	281396 - 283345	Nucleoside-phosphoglycolipid	asparagine synthase	
2	ctg1 orf00534	283383 - 284561	Nucleoside-phosphoglycolipid	Moenomycin biosynthesis protein	
2	ctg1 orf00536	284558 - 286099	Nucleoside-phosphoglycolipid	asparagine synthetase	
2	ctg1 orf00538	287822 - 288964	Nucleoside-phosphoglycolipid	hypothetical protein [Streptomyces viridosporus]	
2	ctg1 orf00541	291441 - 292283	Nucleoside-phosphoglycolipid	Moenomycin biosynthesis protein	
2	ctg1 orf00547	294717 - 295835	Nucleoside-phosphoglycolipid	predicted protein [Streptomyces]	
2	ctg1 orf00550	296702 - 298312	Nucleoside-phosphoglycolipid	hypothetical protein [Streptomyces viridosporus]	
2	ctg1 orf00552	298385 - 299503	Nucleoside-phosphoglycolipid	polysaccharide biosynthesis protein CapD	
2	ctg1 orf00554	299913 - 300743	Nucleoside-phosphoglycolipid	polysaccharide biosynthesis protein	
3	ctg1 orf00707	365624 - 366538	Other	F420-dependent oxidoreductase	
3	ctg1 orf00738	381168 - 383909	Other	peptide synthetase	
3	ctg1 orf00739	383914 - 387069	Other	peptide synthetase	
3	ctg1 orf00741	387106 - 388113	Other	SyrP	
3	ctg1 orf00744	388110 - 389180	Other	ornithine cyclodeaminase	
3	ctg1 orf00745	389222 - 390274	Other	cysteine synthase	
3	ctg orf00775	406809 - 407921	Other	carboxylate-amine ligase	
4	ctg1 orf00925	489762 - 492032	Nrps-t1pks	cytochrome P450	
4	ctg1 orf00931	494318 - 495586	Nrps-t1pks	D-amino acid dehydrogenase	
4	ctg1 orf00936	498152 - 501562	Nrps-t1pks	pyruvate carboxylase	
4	ctg1 orf00940	503031 - 504272	Nrps-t1pks	cytochrome P450	
4	ctg1 orf00942	504537 - 505754	Nrps-t1pks	ferredoxin reductase	
4	ctg1 orf00944	506206 - 507684	Nrps-t1pks	glutamate--cysteine ligase	
4	ctg1 orf00952	510500 - 523249	Nrps-t1pks	peptide synthetase	CAL - PCP - KS - AT - KR - PCP - C - A - PCP - C
4	ctg1 orf00965	528476 - 529861	Nrps-t1pks	cytochrome P450	
4	ctg1 orf00968	530074 - 530916	Nrps-t1pks	non-heme chloroperoxidase	
4	ctg1 orf00978	534252 - 535853	Nrps-t1pks	glycerol-3-phosphate dehydrogenase	
5	ctg1 orf01354	750607 - 751275	Lantipeptide	cytochrome P450	
5	ctg1 orf01355	751288 - 752511	Lantipeptide	cytochrome P450	
5	ctg1 orf01364	754904 - 755032	Lantipeptide	hypothetical protein	
5	ctg1 orf01366	755123 - 755269	Lantipeptide	hypothetical protein	
5	ctg1 orf01370	755352 - 758450	Lantipeptide	lantibiotic dehydratase	
5	ctg1 orf1373	758437 - 759696	Lantipeptide	Lanthionine biosynthesis cyclase LanC	
5	ctg1 orf1386	768424 - 769293	Lantipeptide	NADPH:quinone reductase	
6	ctg1 orf01557	865697 - 867739	Nrps	para-aminobenzoate synthase	
6	ctg1 orf01561	869054 - 869686	Nrps	FAD dependent oxidoreductase	
6	ctg1 orf01564	870590 - 871111	Nrps	O-methyltransferase	
6	ctg1 orf01570	872517 - 873476	Nrps	O-methyltransferase	
6	ctg1 orf01575	875833 - 876021	Nrps	protein mbtH	
6	ctg1 orf01577	876127 - 876924	Nrps	histidinol-phosphate aminotransferase	
6	ctg1 orf01581	877333 - 882531	Nrps	amino acid adenylation domain protein	
6	ctg1 orf01584	888853 - 899850	Nrps	amino acid adenylation protein	
6	ctg1 orf01585	901304 - 902374	Nrps	methyltransferase MppJ	
6	ctg1 orf01586	907692 - 909671	Nrps	gamma-glutamyltransferase	
6	ctg1 orf01593	912526 - 913038	Nrps	methionyl-tRNA formyltransferase	

Cluser Nr	Designation	Location	Type	Function	Of Composition
7	ctg1 orf01662	951822 - 953117	Type 2 pks	sugar ABC transporter substrate-binding protein	
7	ctg1 orf01674	960141 - 961793	Type 2 pks	monooxygenase	
7	ctg1 orf01683	963734 - 965002	Type 2 pks	beta-ACP synthase	
7	ctg1 orf01684	964999 - 966264	Type 2 pks	polyketide beta-ketoacyl synthase	
7	ctg1 orf01685	966292 - 966552	Type 2 pks	Curamycin polyketide synthase acyl carrier protein	
7	ctg1 orf01686	966555 - 967037	Type 2 pks	polyketide cyclase	
7	ctg1 orf01688	967066 - 967404	Type 2 pks	polyketide synthase	
7	ctg1 orf01690	967520 - 968548	Type 2 pks	O-methyltransferase	
8	ctg1 orf03066	1816730 - 1817581	Ectoine	ABC transporter substrate-binding protein	
8	ctg1 orf03067	1817805 - 1818908	Ectoine	aspartate aminotransferase	
8	ctg1 orf03069	1819564 - 1820070	Ectoine	2,4-diaminobutyric acid acetyltransferase	
8	ctg1 orf03070	1820161 - 1821432	Ectoine	diadenosine tetraphosphatase	
8	ctg1 orf03071	1821593 - 1821991	Ectoine	ectoine synthase	
8	ctg1 orf03075	1823044 - 1824141	Ectoine	class V aminotransferase	
8	ctg1 orf03080	1825030 - 1825536	Ectoine	acetyltransferase	
8	ctg1 orf03083	1826331 - 1827860	Ectoine	aldehyde dehydrogenase	
9	ctg1 orf04003	2394841 - 2396511	Type 1 pks	aldehyde dehydrogenase	
9	ctg1 orf04004	2396566 - 2397675	Type 1 pks	isobutylamine N-hydroxylase	
9	ctg1 orf04006	2397705 - 2398991	Type 1 pks	ornithine aminotransferase	
9	ctg1 orf04013	2400327 - 2401361	Type 1 pks	VlmB	
9	ctg1 orf04015	2401358 - 2402383	Type 1 pks	3-oxoacyl-ACP synthase	
9	ctg1 orf04017	2402489 - 2402719	Type 1 pks	phosphopantetheine-binding protein	
9	ctg1 orf04018	2402716 - 2403939	Type 1 pks	beta-ketoacyl synthase	
9	ctg1 orf04023	2406732 - 2407508	Type 1 pks	thioesterase	
9	ctg1 orf04024	2407505 - 2408500	Type 1 pks	F420-dependent oxidoreductase	
9	ctg1 orf04026	2408658 - 2409065	Type 1 pks	aromatic compound degradation protein PaaI	
9	ctg1 orf04030	2410199 - 2414713	Type 1 pks	beta-ketoacyl synthase	
9	ctg1 orf04032	2414746 - 2415105	Type 1 pks	phosphopantetheine-binding protein	
9	ctg1 orf04031	2415220 - 2415777	Type 1 pks	Putative decarboxylase	
9	ctg1 orf04041	2418935 - 2419489	Type 1 pks	NADPH-flavin oxidoreductase	
9	ctg1 orf04043	2420128 - 2421789	Type 1 pks	long-chain-fatty-acid--CoA ligase	
9	ctg1 orf04057	2428455 - 2429399	Type 1 pks	ACP S-malonyltransferase	
9	ctg1 orf04058	2429422 - 2430423	Type 1 pks	3-oxoacyl-ACP synthase	
9	ctg1 orf04060	2430531 - 2430779	Type 1 pks	acyl carrier protein	
9	ctg1 orf04061	2430806 - 2432185	Type 1 pks	3-oxoacyl-ACP synthase	
10	ctg1 orf04641	2794742 - 2795638	Melanin	alpha/beta hydrolase	
10	ctg1 orf04650	2798379 - 2798795	Melanin	tyrosinase	
10	ctg1 orf04656	2800941 - 2802284	Melanin	glycosyl transferase family 1	
10	ctg1 orf04657	2802281 - 2803723	Melanin	transferase	
11	ctg1 orf04787	2889789 - 2891234	Siderophore	pyridoxal-dependent decarboxylase	
11	ctg1 orf04789	2891215 - 2892510	Siderophore	monooxygenase	
11	ctg1 orf04791	2892507 - 2893085	Siderophore	acetyltransferase	
11	ctg1 orf04792	2893082 - 2894890	Siderophore	siderophore biosynthetic enzyme	
11	ctg1 orf04798	2896467 - 2898284	Siderophore	glutamine amidotransferase	
12	ctg1 orf08303	5054531 - 5054917	Nrps-type 1 pks	methyltransferase	
12	ctg1 orf08310	5059545 - 5060741	Nrps-type 1 pks	cytochrome P450	
12	ctg1 orf08311	5060886 - 5061941	Nrps-type 1 pks	alcohol dehydrogenase	-ER -
12	ctg1 orf08313	5061938 - 5063815	Nrps-type 1 pks	phytoene dehydrogenase	
12	ctg1 orf08314	5063863 - 5065479	Nrps-type 1 pks	FAD-dependent oxidoreductase	
12	ctg1 orf08317	5065712 - 5075383	Nrps-type 1 pks	polyketide synthase	KS - ST - DH - KR - ACP - C - A - PCP - TE
12	ctg1 orf08335	5083942 - 5084868	Nrps-type 1 pks	alpha/beta hydrolase	
12	ctg1 orf08336	5084904 - 5085899	Nrps-type 1 pks	beta-lactamase	
12	ctg1 orf08339	5088081 - 5089661	Nrps-type 1 pks	serine protease	
13	ctg1 orf08606	5250459 - 5251073	Terpene	peptide deformylase	
13	ctg1 orf08608	5251555 - 5252478	Terpene	multidrug MFS transporter	
13	ctg1 orf08611	5252475 - 5253845	Terpene	cytochrome P450	

Cluser Nr	Designation	Location	Type	Function	Of Composition
14	ctg1 orf09695	5961970 - 5963226	Type 1 pks	carboxylate--amine ligase	
14	ctg1 orf09705	5968369 - 5669901	Type 1 pks	FAD dependent oxidoreductase	
14	ctg1 orf09711	5970176 - 5971309	Type 1 pks	aldehyde dehydrogenase	- ER -
14	ctg1 orf09712	5971758 - 5973284	Type 1 pks	glycerol kinase	
14	ctg1 orf09724	5979103 - 5980344	Type 1 pks	FAD-binding monooxygenase	
14	ctg1 orf09731	5982007 - 5991681	Type 1 pks	Polyketide synthase	- A - PCP - KS - ACP - KS - AT -
14	ctg1 orf09734	5991678 - 5998508	Type 1 pks	Polyketide synthase	- KR - ACP - KS - ACP - KS
14	ctg1 orf09735	5998573 - 6001080	Type 1 pks	polyketide synthase	- KR - ACP -
14	ctg1 orf09738	6001085 - 6004810	Type 1 pks	Polyketide synthase	- KS - ACP -
14	ctg1 orf09740	6005044 - 6005898	Type 1 pks	Polyketide synthase, partial	- KS
14	ctg1 orf09741	6005932 - 6007446	Type 1 pks	Polyketide synthase	- AT -
14	ctg1 orf09744	6008557 - 6012564	Type 1 pks	aminotransferase	- KR - ACP - AMINOTRAN_1_2
15	ctg1 orf10137	6244548 - 6246011	Bacteriocin	flavoprotein oxidoreductase	
15	ctg1 orf10140	6247267 - 6248703	Bacteriocin	protoporphyrinogen oxidase	
15	ctg1 orf10144	6249759 - 6251102	Bacteriocin	hypothetical protein	
15	ctg1 orf10149	6253257 - 6254054	Bacteriocin	Methyltransferase	
16	ctg1 orf10202	6281695 - 6283857	Terpene	Geosmin synthase	
16	ctg1 orf10206	6285512 - 6287257	Terpene	malto-oligosyltrehalose trehalohydrolase	
16	ctg1 orf10215	6290155 - 6292599	Terpene	alpha-amylase	
16	ctg1 orf10217	6292691 -6294838	Terpene	glycogen debranching protein	
17	ctg1 orf10386	6397598 - 6398806	Siderophore	aminotransferase AlaT	
17	ctg1 orf10394	6401876 - 6403462	Siderophore	iron transporter	
17	ctg1 orf10395	6403500 - 6405086	Siderophore	iron transporter	
17	ctg1 orf10400	6407845 - 6408708	Siderophore	2-hydroxy-6-ketonona-2,4-dienedoic acid hydrolase	
17	ctg1 orf10402	6408785 - 6409700	Siderophore	alpha/beta hydrolase fold	
17	ctg1 orf10403	6409720 - 6410886	Siderophore	acyl-CoA dehydrogenase	
18	ctg1 orf10419	6419432 - 6420271	Terpene	alpha/beta hydrolase	
18	ctg1 orf10421	6420300 - 6421472	Terpene	acyl-CoA dehydrogenase	
18	ctg1 orf10434	6426858 - 6428081	Terpene	lycopene cyclase	
18	ctg1 orf10436	6428078 - 6428804	Terpene	ubiquinone biosynthesis methyltransferase	
18	ctg1 orf10437	6428806 - 6430353	Terpene	isorenieratene synthase	
18	ctg1 orf10440	6431623 - 6432681	Terpene	phytoene synthase	
18	ctg1 orf10442	6432668 - 6434269	Terpene	phytoene dehydrogenase	
18	ctg1 orf10443	6434266 - 6435387	Terpene	geranylgeranyl pyrophosphate synthase	
18	ctg1 orf10444	6435634 - 6436560	Terpene	short-chain dehydrogenase	
18	ctg1 orf10450	6437056 - 6438309	Terpene	cytochrome P450	
18	ctg1 orf10451	6438579 - 6439904	Terpene	Monooxygenase, FAD-binding	
18	ctg1 orf10457	6440831 - 6441511	Terpene	O-methyltransferase	
18	ctg1 orf10459	6441544 - 6442548	Terpene	alcohol dehydrogenase	
19	ctg1 orf10443	6434266 - 6435387	Type 3 pks	geranylgeranyl pyrophosphate synthase	
19	ctg1 orf10444	6435634 - 6436560	Type 3 pks	short-chain dehydrogenase	
19	ctg1 orf10450	6437056 - 6438309	Type 3 pks	cytochrome P450	
19	ctg1 orf10451	6438579 - 6439904	Type 3 pks	putative FAD-dependent monooxygenase	
19	ctg1 orf10457	6440831 - 6441511	Type 3 pks	O-methyltransferase	
19	ctg1 orf10459	6441544 - 6442548	Type 3 pks	alcohol dehydrogenase	
19	ctg1 orf10460	6443427 - 6444644	Type 3 pks	cytochrome P450	
19	ctg1 orf10464	6447380 - 6448444	Type 3 pks	methyltransferase	
19	ctg1 orf10465	6448503 - 6449573	Type 3 pks	methyltransferase	
19	ctg1 orf10468	6449863 - 6450837	Type 3 pks	putative adenosine kinase	
19	ctg1 orf10474	6452499 - 6453920	Type 3 pks	FAD-dependent oxygenase	
19	ctg1 orf10477	6455038 - 6456120	Type 3 pks	Type III polyketide synthase RppA	
19	ctg1 orf10478	6456117 - 6457115	Type 3 pks	O-methyltransferase 2	
19	ctg1 orf10486	6460871 - 6462061	Type 3 pks	acyl-CoA dehydrogenase type 2 domain-containing protein	
19	ctg1 orf10485	6462058 - 6462651	Type 3 pks	flavin reductase domain-containing protein	
19	ctg1 orf10492	6464688 - 6466241	Type 3 pks	aldehyde dehydrogenase	
19	ctg1 orf10508	6473505 - 6744185	Type 3 pks	4-diphosphocytidyl-2C-methyl-D-erythritol synthase	

Cluser Nr	Designation	Location	Type	Function	Of Composition
20	ctg1 orf10545	6495409 - 6497040	Nrps-type 1 pks	glycosyltransferase	
20	ctg1 orf10549	6498613 - 6499614	Nrps-type 1 pks	NAD-dependent dehydratase	- NAD -
20	ctg1 orf10551	6499737 - 6500450	Nrps-type 1 pks	mannose-1-phosphate guanyltransferase	
20	ctg1 orf10553	6500457 - 6501485	Nrps-type 1 pks	NDP-hexose 4-ketoreductase	
20	ctg1 orf10562	6505289 - 6501485	Nrps-type 1 pks	beta-lactamase domain-containing protein	
20	ctg1 orf10565	6507760 - 6508746	Nrps-type 1 pks	aldo/keto reductase	
20	ctg1 orf10569	6508757 - 6509572	Nrps-type 1 pks	thioesterase	- TE -
20	ctg1 orf10570	6509569 - 6511092	Nrps-type 1 pks	cytochrome P450	
20	ctg1 orf10573	6512827 - 6514386	Nrps-type 1 pks	condensation domain-containing protein	- C -PCP
20	ctg1 orf10575	6514383 - 6517850	Nrps-type 1 pks	long-chain-fatty-acid--CoA ligase	- PCP - C - A - PCP
20	ctg1 orf10578	6517850 - 6519025	Nrps-type 1 pks	cytochrome P450	
20	ctg1 orf10581	6519027 - 6522812	Nrps-type 1 pks	Beta-ketoacyl synthase	- KS - AT - PCP - TE
20	ctg1 orf10583	6524194 - 6525477	Nrps-type 1 pks	cytochrome P450	
20	ctg1 orf10585	6525479 - 6526216	Nrps-type 1 pks	thioesterase	- TE -
20	ctg1 orf10586	6526207 - 6527031	Nrps-type 1 pks	short-chain dehydrogenase/reductase SDR	- KR -
20	ctg1 orf10588	6527056 - 6529767	Nrps-type 1 pks	long-chain-fatty-acid--CoA ligase	- A - PCP -
20	ctg1 orf10593	6533303 - 6534502	Nrps-type 1 pks	3-oxoadipate enol-lactonase	
20	ctg1 orf10603	6538242 - 6540068	Nrps-type 1 pks	gamma-glutamyltransferase	
20	ctg1 orf10606	6543107 - 6543874	Nrps-type 1 pks	methionine aminopeptidase	
20	ctg1 orf10611	6546008 - 6546850	Nrps-type 1 pks	apolipoprotein acyltransferase	
20	ctg1 orf10613	6546847 - 6548133	Nrps-type 1 pks	aminotransferase class III	- Aminotran_3 -
21	ctg1 orf11165	6914237 - 6914989	Terpene	nucleotide sugar-1-phosphate transferase	
21	ctg1 orf11167	9614977 - 6916038	Terpene	dehydrogenase	
21	ctg1 orf11176	6919843 - 6920796	Terpene	squalene synthase	
21	ctg1 orf11178	6920793 - 6921743	Terpene	phytoene synthase	
21	ctg1 orf11182	6923421 - 6924485	Terpene	dimethylallyltranstransferase	
21	ctg1 orf11185	6924653 - 6926701	Terpene	squalene-hopene cyclase	
21	ctg1 orf11190	6928375 - 6929790	Terpene	aminotransferase	
22	ctg1 orf11368	7037750 - 7038787	Nrps	epoxide hydrolase	
22	ctg1 orf11382	7051032 - 7051784	Nrps	oleoyl-(acyl-carrier-protein) hydrolase	- TE -
22	ctg1 orf11385	7052769 - 7053893	Nrps	cytochrome P450	
22	ctg1 orf11388	7059788 - 7055337	Nrps	alpha/beta hydrolase	
22	ctg1 orf11395	7055378 - 7074142	Nrps	NRPS protein	- A - PCP - C - A - PCP - E - C - A - PCP - E - C - A - PCP - E - C - A - PCP -
22	ctg1 orf11401	7074129 - 7094216	Nrps	NRPS protein	- C - A - PCP - E - C - A - PCP - C - A - PCP - E - C - A - PCP - C - A - PCP - E -
22	ctg1 orf11404	7094626 - 7095138	Nrps	acetyltransferase	
22	ctg1 orf11415	7101647 - 7102453	Nrps	acetyltransferase	
22	ctg1 orf11428	7109829 - 7111115	Nrps	sugar-binding protein	
23	ctg1 orf11836	7370600 - 7371850	Terpene	cytochrome P450	
23	ctg1 orf11840	7371944 - 7373065	Terpene	terpene cyclase	
24	ctg1 orf12218	7609286 - 7610233	Bacteriocin - t1pks	theronine dehydrogenase-like Zn-dependent dehydrogenase	- CAL - ACP - KS - AT - DH - KR - ACP - TE
24	ctg1 orf12231	7614516 - 7622954	Bacteriocin - t1pks	polyketide synthase type I	
24	ctg1 orf12247	7630796 - 7632157	Bacteriocin - t1pks	FAD dependent oxidoreductase	
25	ctg1 orf12273	7644361 - 7645452	Nrps	cystathione synthase	
25	ctg1 orf12275	7645585 - 7646652	Nrps	ornithine cyclodeaminase	
25	ctg1 orf12277	7646661 - 7646903	Nrps	MbtH-like protein	
25	ctg1 orf12278	7646945 - 7649398	Nrps	M28 family peptidase	
25	ctg1 orf12283	7651337 - 7656598	Nrps	peptide synthetase	- CAL - C - A - PCP
25	ctg1 orf12285	7656595 - 7657398	Nrps	Thioesterase	- TE -
25	ctg1 orf12290	7658201 - 7667869	Nrps	grsb-like gramicidin s synthetase 2	- C - A - PCP - C - A - PCP - C - A - PCP
25	ctg1 orf12293	7667869 - 7671126	Nrps	non-ribosomal peptide synthase	- C - A - PCP -
25	ctg1 orf12295	7671291 - 7675094	Nrps	hypothetical protein	-C -
25	ctg1 orf12297	7676275 - 7680090	Nrps	amino acid adenylation domain-containing protein	- C - A - PCP - TE -
25	ctg1 orf12304	7682022 - 7683932	Nrps	glycosyl transferase	
25	ctg1 orf12312	7687535 - 7688524	Nrps	aldo/keto reductase	
25	ctg1 orf12333	7699977 - 7701533	Nrps	monooxygenase	

Cluser Nr	Designation	Location	Type	Function	Of Composition
26	ctg1 orf12346	7707360 - 7708352	Butyrolactone	gamma-butyrolactone biosynthesis protein SrrX	
26	ctg1 orf12348	7708927 - 7709784	Butyrolactone	methylmalonyl-CoA carboxyltransferase	
26	ctg1 orf12356	7712204 - 7715473	Butyrolactone	non-ribosomal peptide synthetase	
27	ctg1 orf12394	7741635 - 7743968	Nrps-type 1 pks	beta-ketoacyl synthase	- KR - ACP
27	ctg1 orf12396	7746331 - 7748838	Nrps-type 1 pks	polyketide synthase	- KR - ACP
27	ctg1 orf12397	7752175 - 7752858	Nrps-type 1 pks	hypothetical protein	- KS -
27	ctg1 orf12398	7756323 - 7758704	Nrps-type 1 pks	putative polyketide synthase	- DH - KR - ACP
27	ctg1 orf12401	7761084 - 7766768	Nrps-type 1 pks	beta-ketoacyl synthase	- KS - AT - DH - KR - ACP
27	ctg1 orf12404	7766858 - 7772904	Nrps-type 1 pks	acyl transferase	- ACP - KS - AT - AT - KR - ACP - Pks Docking
27	ctg1 orf12406	7773183 - 7785242	Nrps-type 1 pks	amino acid adenylation domain protein	- DH - ER - KR - ACP - C - A - PCP - KS - AT - KR - ACP - TE
27	ctg1 orf12408	7786858 - 7788315	Nrps-type 1 pks	beta-ketoacyl synthase	- PKS Docking - KS -
27	ctg1 orf12410	7788444 - 7791155	Nrps-type 1 pks	polyketide synthase	- DH - KR- ACP - PKS docking
27	ctg1 orf12413	7792132 - 7802340	Nrps-type 1 pks	polyketide synthase type I	- DH - ER - KR - ACP - KS - AT - DH - KR - ACP - KS
27	ctg1 orf12414	7803490 - 7805190	Nrps-type 1 pks	hypothetical protein	- KS -
27	ctg1 orf12427	7812209 - 7813234	Nrps-type 1 pks	cyclase	- Polyketide cyc2 - Polyketide cyc2 -
27	ctg1 orf12429	7813280 - 7813969	Nrps-type 1 pks	polyketide synthase	- AT -
27	ctg1 orf12430	7814158 - 7816257	Nrps-type 1 pks	peptide synthetase	- C -
28	ctg1 orf12460	7827256 - 7829040	Type 1 pks	malonyl CoA-acyl carrier protein transacylase	- KS -
28	ctg1 orf12461	7829037 - 7829843	Type 1 pks	probable 4'-phosphopantetheinyl transferase	- ACPS -
28	ctg1 orf12470	7834828 - 7836072	Type 1 pks	cytochrome P450	
28	ctg1 orf12474	7836209 - 7840762	Type 1 pks	polyketide synthase	- PKS Docking - KS - AT - DH -
28	ctg1 orf12476	7841138 - 7852612	Type 1 pks	acyl transferase	- KR - ACP - KS - AT - KR - ACP - KS - AT - DH - KR - ACP - PKS Docking
28	ctg1 orf12478	7852609 - 7857369	Type 1 pks	modular polyketide synthase	- PKS Docking - KS - AT - KR - ACP - PKS Docking
28	ctg1 orf12480	7857372 - 7866269	Type 1 pks	acyl transferase	- KS - AT - KR - ACP - KS - AT - DH -
28	ctg1 orf12481	7866329 - 7867897	Type 1 pks	polyketide synthase	- KR - ACP - PKS Docking
28	ctg1 orf12490	7871078 - 7888324	Type 1 pks	beta-ketoacyl synthase	- KS - AT - ACP - KS - AT - KR - ACP - KS - AT - KR - ACP - KS - AT - KR - ACP
28	ctg1 orf12498	7888361 - 7905634	Type 1 pks	polyketide synthase	- PKS Docking - KS - AT - ER - KR - ACP - KS - AT - KR - ACP - KS - AT - KR - ACP - KS - AT
28	ctg1 orf12501	7905706 - 7917783	Type 1 pks	beta-ketoacyl synthase	- DH - KR - ACP - KS - AT - DH - KR - ACP - KS - AT - KR - ACP
28	ctg1 orf12505	7918040 - 7938229	Type 1 pks	beta-ketoacyl synthase	- PKS Docking - KS - AT - KR - ACP - KS - AT - KR - ACP - KS - AT - KR - ACP - KS - AT - ER - KR - ACP
28	ctg1 orf12509	7938226 - 7947894	Type 1 pks	modular polyketide synthase	- PKS Docking - KS - AT - KR - ACP - KS - AT - KR - ACP
28	ctg1 orf12511	7947929 - 7950316	Type 1 pks	beta-ketoacyl synthase	- PKS Docking - KS - AT
28	ctg1 orf12512	7951108 - 7954743	Type 1 pks	beta-ketoacyl synthase	- DH - KR - ACP - KS
28	ctg1 orf12514	7956144 - 7958903	Type 1 pks	acyl transferase	- DH - KR - ACP - PKS Docking
28	ctg1 orf12516	7958908 - 7960086	Type 1 pks	beta-ketoacyl synthase	- PKS Docking - KS
28	ctg1 orf12519	7962380 - 7970317	Type 1 pks	modular polyketide synthase	- KR - ACP - KS - AT - KR - ACP - KS
28	ctg1 orf12520	7970860 - 7973193	Type 1 pks	acyl transferase	- AT - KR
28	ctg1 orf12522	7973230 - 7978743	Type 1 pks	beta-ketoacyl synthase	- PKS Docking - KS - AT - KR - ACP - TE
28	ctg1 orf12527	7981196 - 7982212	Type 1 pks	oxidoreductase	
28	ctg1 orf12528	7982338 - 7983594	Type 1 pks	glycosyltransferase	
28	ctg1 orf12534	7985723 - 7986832	Type 1 pks	aminotransferase	
28	ctg1 orf12536	7986863 - 7988059	Type 1 pks	cytochrome P450	
28	ctg1 orf12538	7988094 - 7988852	Type 1 pks	thioesterase	- TE -
28	ctg1 orf12539	7989132 - 7990388	Type 1 pks	dehydrogenase	- ER -
28	ctg1 orf12546	7990461 - 8015036	Type 1 pks	Request-URI Too Large	- KS - AT - KR - ACP - KS - AT - ER - KR - ACP - KS - AT - KR - ACP - KS - AT - KR - ACP - KS - AT - KR - ACP

Cluser Nr	Designation	Location	Type	Function	Of Composition
29	ctg1 orf12721	8114690 - 8115259	Nrps-bacteriocin-terpene	N-acetyltransferase GCN5	
29	ctg1 orf12747	8130591 - 8133956	Nrps-bacteriocin-terpene	amino acid adenylation protein	- A - PCP - E -
29	ctg1 orf12752	8137013 - 8137936	Nrps-bacteriocin-terpene	2-oxoacid ferredoxin oxidoreductase subunit beta	
29	ctg1 orf12769	8149120 - 8150412	Nrps-bacteriocin-terpene	beta-ketoacyl synthase	- AT -
29	ctg1 orf12770	8150444 - 8151634	Nrps-bacteriocin-terpene	alpha-amylase, partial	
29	ctg1 orf12771	8151671 - 8152798	Nrps-bacteriocin-terpene	dimodular nonribosomal peptide synthetase	
29	ctg1 orf12778	8155030 - 8156127	Nrps-bacteriocin-terpene	beta-ketoacyl synthase	- KS -
29	ctg1 orf12783	8158233 - 8158742	Nrps-bacteriocin-terpene	polyketide synthase	- AT -
29	ctg1 orf12784	8158765 - 8159253	Nrps-bacteriocin-terpene	polyketide synthase	- AT -
29	ctg1 orf12793	8162958 - 8163446	Nrps-bacteriocin-terpene	acetyltransferase	
29	ctg1 orf12794	8163583 - 8164068	Nrps-bacteriocin-terpene	laidlomycin polyketide synthase	- KS -
29	ctg1 orf12797	8164593 - 8165030	Nrps-bacteriocin-terpene	trehalose synthase	
29	ctg1 orf12803	8166586 - 8167008	Nrps-bacteriocin-terpene	amino acid adenylation domain-containing protein	- C -
29	ctg1 orf12809	8168903 - 8169058	Nrps-bacteriocin-terpene	polyketide synthase	- KS -
29	ctg1 orf12810	8169051 - 8169380	Nrps-bacteriocin-terpene	modular type I polyketide synthase	
29	ctg1 orf12812	8169400 - 8169729	Nrps-bacteriocin-terpene	modular type I polyketide synthase	
29	ctg1 orf12815	8170444 - 8170764	Nrps-bacteriocin-terpene	acyl transferase	
29	ctg1 orf12819	8171429 - 8171719	Nrps-bacteriocin-terpene	beta-ketoacyl synthase	
29	ctg1 orf12820	8171751 - 8172041	Nrps-bacteriocin-terpene	beta-ketoacyl synthase	
29	ctg1 orf12823	8172947 - 8173207	Nrps-bacteriocin-terpene	polyketide synthase	
29	ctg1 orf12829	8174386 - 8174571	Nrps-bacteriocin-terpene	polyketide synthase	
29	ctg1 orf12831	8175242 - 8175403	Nrps-bacteriocin-terpene	acyl transferase	
29	ctg1 orf12832	8175636 - 8175827	Nrps-bacteriocin-terpene	acyl transferase	
29	ctg1 orf12841	8178162 - 8178326	Nrps-bacteriocin-terpene	polyketide synthase	
29	ctg1 orf12844	8180013 - 8180132	Nrps-bacteriocin-terpene	1-hexene synthase	
29	ctg1 orf12846	8180585 - 8180719	Nrps-bacteriocin-terpene	acyl transferase	
29	ctg1 orf12847	8180720 - 8180860	Nrps-bacteriocin-terpene	beta-ketoacyl synthase	

8.2 Appendix 2

Annotation of all regulatory genes found in in silico mining of gene clusters in *S. bambergiensis.*

Cluster	**Type of Cluster**	**Type of Regulator**
Cluster 1	**Nrps**	
R1	smCOG1136	GenTr Family Transcriptional Regulator
R2	smCOG1058	ArsR Family Transcriptional Regulator
R3	smCOG1016	LuxR Family DNA-Binding Response Regulator
R4	smCOG1149	LuxR Family Transcriptional Regulator
R5	smCOG1041	Transcriptional Regulator SARP Family
Cluster 3	**Other**	
R1	smCOG1041	Transcriptional Regulator SARP Family
R2	smCOG1008	Response Regulator
Cluster 4	**Nrps-t1pks**	
R1	smCOG1008	Response Regulator
Cluster 5	**Lantipeptide**	
R1	smCOG1195	IclR Family Transcriptional Regulator
R2	smCOG1016	LuxR Family DNA-Binding Response Regulator
Cluster 7	**Type 2 pks**	
R1	smCOG1071	GntR Family Transcriptional Regulator
R2	smCOG1058	ArsR Family Transcriptional Regulator
R3	smCOG1078	LacI Family Transcriptional Regulator
R4	smCOG1135	MarR Family Transcriptional Regulator
R5	smCOG1260	AraC Family Transcriptional Regulator
R6	smCOG1030	Serine / Threonine Protein Kinase
Cluster 9	**Type 1 pks**	
R1	smCOG1041	Transcriptional Regulator SARP Family
Cluster 12	**Nrps-type 1 pks**	
R1	smCOG1136	GenTr Family Transcriptional Regulator
Cluster 14	**Type 1 pks**	
R1	smCOG1078	LacI Family Transcriptional Regulator
Cluster 19	**Type 3 pks**	
R1	smCOG1041	Transcriptional Regulator SARP Family
R2	smCOG1135	MarR Family Transcriptional Regulator
R3	smCOG1041	Transcriptional Regulator SARP Family
R4	smCOG1284	Transcriptional Regulator MarR Family
R5	smCOG1041	Transcriptional Regulator SARP Family
R6	smCOG1041	Transcriptional Regulator SARP Family
R7	smCOG1014	LysR Family Transcriptional Regulator
R8	smCOG1195	IclR Family Transcriptional Regulator
Cluster 20	**Nrps-type 1 pks**	
R1	smCOG1197	Autoinducer - Binding Transcriptional Regulator
Cluster 22	**Nrps**	

Cluster	Type of Cluster	Type of Regulator
R1	smCOG1078	LacI Family Transcriptional Regulator
Cluster 24	**Bacteriocin - t1pks**	
R1	smCOG1078	LacI Family Transcriptional Regulator
Cluster 25	**Nrps**	
R1	smCOG1078	LacI Family Transcriptional Regulator
R2	smCOG1016	LuxR Family DNA-Binding Response Regulator
R3	smCOG1016	LuxR Family DNA-Binding Response Regulator
Cluster 27	**Nrps-type 1 pks**	
R1	smCOG1041	Transcriptional Regulator SARP Family
R2	smCOG1041	Transcriptional Regulator SARP Family
Cluster 28	**Type 1 pks**	
R1	smCOG1149	LuxR Family Transcriptional Regulator
Cluster 29	**Nrps-bacteriocin-terpene**	
R1	smCOG1016	LuxR Family DNA-Binding Response Regulator