MASTERARBEIT/ MASTER THESIS

Titel der Masterarbeit / Title of the Master's Thesis

"Discovery of natural products from Actinobacteria via genome mining"

verfasst von / submitted by

Sonali Vaidya

angestrebter akademischer Grad / in partial fulfillment of the requirements for the degree of

Master of Science (MSc)

Wien 2021, Vienna 2021

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 606
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Drug Discovery and Development
Betreut von / Supervisor:	Univ.-Prof. Dr. Sergey B. Zotchev
Mitbetreut von / Co-Supervisor:	Dr. Olha Schneider

Table of Contents

I. List of abbreviations

ACP: Acyl carrier protein
AT: Acyl transferase AT
antiSMASH: Antibiotics and Secondary Metabolite Analysis Shell
BAC: Bacterial-artificial-chromosome
BGC: Biosynthetic Gene Cluster
BLAST: Basic Local Alignment Search Tool
bp: Base pair
CREs: cis-regulatory elements
DH: Dehydratase
dH_2O: distilled water
fDNA: Fosmid DNA
Helix-turn-helix HTH
HTS: High throughput screening
HGT: Horizontal gene transfer
KS: Ketosynthase
MRSA: Methicillin-resistant Staphylococcus aureus
MDR: Multidrug-resistance
NPs: Natural products
NPDD: Natural Product-driven Drug Discovery
NRP: Non-ribosomal peptide
NRPS: Non-ribosomal peptide synthase
Ori T: Origin of transfer
OSMAC: One strain many compounds
pDNA: Plasmid DNA
PKS: Polyketide synthase
PPT: 4-phosphopantetheine
SM: Secondary metabolite
ssDNA: Single-stranded salmon sperm DNA
TetR: Tetracycline repressor
TE: Thioesterase
TAR: Transformation-associated recombination TAR
TF: Transcriptional factor TF
WT: Wild Type

1. Zusammenfassung

Aufgrund der steigenden Zahlen von Erkrankungen, die von Antibiotika-resistenten Mikroorganismen verursacht werden, ist die Entwicklung der antimikrobieller Mittel von großer Bedeutung. Pflanzen und Mikroorganismen sind bereits bewiesene Quellen für natürliche bioaktive Stoffe. Mit der Fortschritten in der Weiterentwicklung von der Sequenzierungsmethoden wurde das verborgene Potenzial von Aktinobakterien als Produzenten der bioaktiven Naturstoffen aufgedeckt. Die Analyse der Genomsequenzierungsdaten zeigte, dass die Vertreter der Abteilung von Aktitobacteria eine Reihe von biosynthetischen Cluster tragen, deren Produkt während einer Kultivierung im Labor noch nie nachgewiesen werden konnte.

Im Rahmen dieses Projekts wurden Experimente mit zwei BGCs aus *Streptomyces bambergiensis* durchgeführt. *In-silico*-Analysen mit bioinformatischen Tools wie antiSMASH und BLAST ergaben, dass BGC 18 und BGC 20 für die Produktion vielversprechender neuen bioaktiven Verbindungen zuständig sein könnten. Als Ausgangspunk für diese Arbeit war ein Screening von *E. coli*-Klonen, die genomische DNA von *Streptomyces bambergiensis* tragen. Mit molekularbiologischen Techniken wurde die DNA aus den *E. coli*-Klonen extrahiert und anschließend für weitere Schritte kloniert. Um Zugang zu den angeblichen bioaktiven Verbindungen zu erhalten, wurden beide BGCs durch homologe Rekombination in Hefe mit Hilfe der TAR-Methode in den Shuttle-Vektor pCLY10 kloniert. Anschließend war eine heterologe Expression in verschiedenen *Streptomyces* heterologen Stämmen geplant. Basiert auf den Ergebnissen von dieser Arbeit, es konnten die Hinweise auf eine Instabilität von dem pCLY10 Vektorsystem erhaltet werden. Auch wenn die PCR-Analyse ergab den Nachweis von einer erfolgsreicher Assembly in Hefe, die Übertragung von dem cluster-tragenden Vektor zu *E. coli* und *Streptomyces* spp. war nicht erfolgsreich, und es müssen weitere Alternativen für eine Optimierung des Protokolls durchgeführt werden.

Der zweite Teil dieser Masterarbeit befasste sich mit den BGC 3 und 4 von *Actinoalloteichus fjordicus* DSM 46856, der aus dem Meeresschwamm *Geodia barretti* isoliert wurde, der im Trondheim-Fjord in Norwegen gesammelt wurde. Die BGCs wurden bereits erfolgreich in Hefe assembliert und das regulatorische SARP-Gen wurde in den Vektor pUWLoriT kloniert und weiter transferiert in die Stämme *Streptomyces coelicolor* M1154 und *Streptomyces albus* J1074. Zur Produktion von Sekundärmetaboliten wurde eine Fermentation in verschiedenen Medien durchgeführt. Der Extrakt aus diesen Kulturen wurde mittels analytischer HPLC und MS bewertet. Gleichzeitig wurde die Bioaktivität der Extrakte getestet, um möglicherweise Produktion von antibakteriellen Stoffen gegen verschiedene Testorganismen nachzuweisen.

Zusammenfassend lieferte diese Arbeit wichtige Hinweise auf eine Instabilität des breitverwendbaren pCLY10-Vektors für die heterologe Expression der biosynthetischen Genclusters. Auch wenn im Rahmen dieses Projekt eine erfolgsreiche Produktion der gewünschten Stoffen nicht erzielt werden konnte, bitten experimentell erzeugte Ergebnisse ein Basis für weitere Forschung an der Verbesserung der biosynthetischen Gencluster-Expression und entsprechend Naturstoffbiosynthesis.

2. Abstract

The discovery of novel antimicrobials is a pressing priority to address the rising problem of pathogens causing infections in humans and becoming resistant due to the frequent use of currently available antibiotics. Plants and microorganisms have already been proven as sources of natural bioactive compounds. Advances in genome sequencing have unravelled the presence of silent Biosynthetic Gene Clusters (BGCs) in microorganisms, especially Actinobacteria and their activation could lead to the discovery of bioactive secondary metabolites which potentially may become new drug candidates.

During this project, experiments were performed on two BGCs from *Streptomyces bambergiensis*. *In-silico* analysis using bioinformatics tools like antiSMASH and BLAST suggested that BGC 18 and BGC 20 encode for potentially novel natural products. After initial pooled-PCR-based library screening of *E. coli* clones harbouring *Streptomyces bambergiensis* genome fragments, clones encoding BGC 18 and 20 were identified. Molecular biology techniques were used to extract the DNA from *E. coli* clones and subsequently genetically manipulate it. To gain an access to the putative bioactive compounds, both BGCs were assembled into the shuttle vector pCLY10 utilizing homologous recombination in yeast using the TAR method. Subsequently, heterologous expression was planned in different engineered *Streptomyces* host strains. However, stability of the cluster-carrying vector in *E. coli* and *Streptomyces* spp. during horizontal gene transfer was compromised and more experiments need to be performed to ensure successful conjugation of BGC into engineered *Streptomyces* host.
The second part of this master thesis work involved working on BGC 3 and 4 of *Actinoalloteichus fjordicus* DSM 46856 which was isolated from a marine sponge *Geodia barretti* collected at the Trondheim fjord in Norway. The BGCs were already successfully assembled in yeast in the previous work and regulatory gene SARP was cloned into multi-copy pUWLoriT vector during this project. The goal was to carry out conjugation in *Streptomyces coelicolor* M1154 [1] and *Streptomyces albus* J1074 [2]. For the production of secondary metabolites, fermentations in different media were carried out. The extract of secondary metabolites from these cultures was evaluated by analytical HPLC and LC-MS. Simultaneously, disc diffusion assays were performed to possibly demonstrate their activity against various test organisms.

We got a research foundation despite of not being able to meet our ultimate goal. It can be assumed that the system used to clone BGC has indefinite stability and that alternative shuttle vectors for cloning large BGCs is therefore necessary.

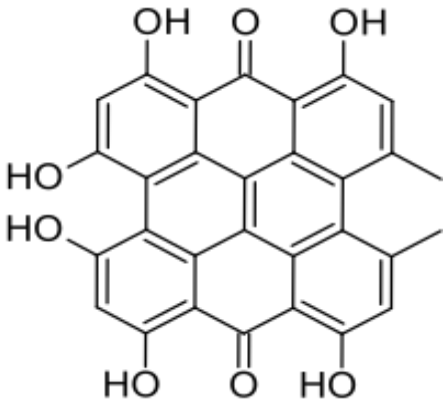
3. Introduction

3.1. Development and use of natural products as drugs

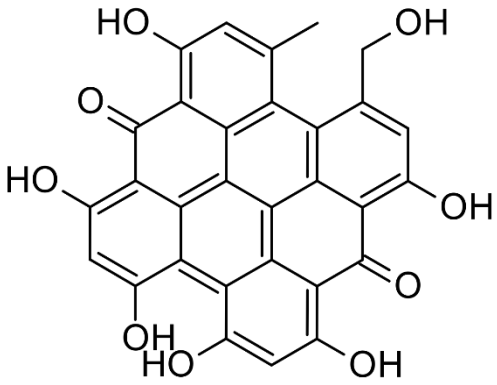
Natural products and derivatives of natural products have been a promising source for the development of new and potential drugs. [3] Plants along with other organisms such as marine and terrestrial animals and microorganisms produce very important drug candidates. Some metabolites from these sources have been proven to be bioactive. [4]

Traditional medicine from ancient civilizations and the wisdom described in Ayurveda, which is an alternative medicine system from India, and Chinese folk medicine also highlight the importance of natural products in therapeutics. This traditional knowledge along with new technologies such as automated separation techniques, high-throughput screening, and combinatorial chemistry will greatly facilitate natural products research and revolutionize the process of drug discovery. [5]

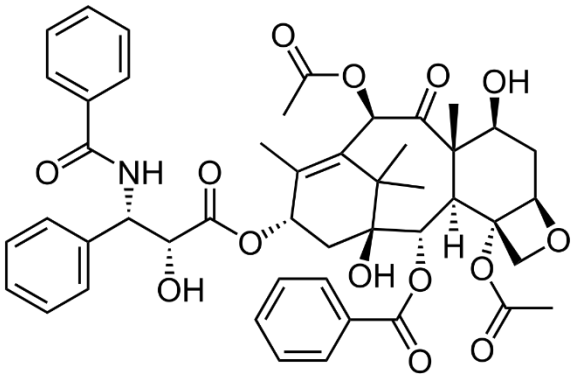
Many drugs derived from natural products have been used for the treatment of diseases in humans and animals. Paclitaxel (Taxol) which was first isolated from the bark of the Pacific yew tree *Taxus brevifolia* (Taxaceae) is among the anticancer drugs that were discovered from natural sources. Activity against a variety of retroviruses, including HIV, two compounds isolated from *Hypericum perforatum* (Guttiferae) are hypericin and pseudohypericin. These compounds inhibit the release of reverse transcriptase by stabilizing the structure of the HIV capsid and thus preventing the uncoating process. A great example of serendipitous discovery is the discovery of the antibiotic penicillin in the laboratory from the fungus *Penicillium notatum.* [6, 7]



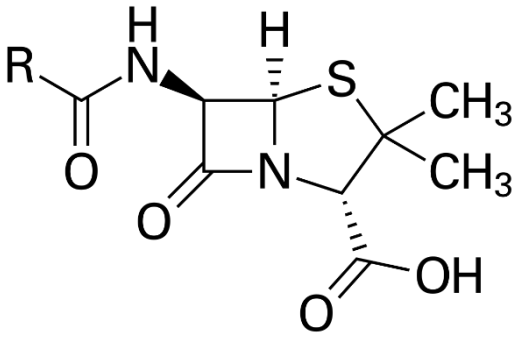
Hypericin



Pseudoypericin



Paclitaxel



Penicillin

Even though the natural products are still promising sources for the discovery of novel therapeutic agents, which when combined with synthetic chemistry and genomics offer the potential to discover novel structures that can lead to effective agents in a variety of human diseases, the industry is spending fewer and fewer resources on their discovery and development. [8] Although there is excellent work going on in this field and will continue to identify novel and potent agents, their development may be hindered due to the current funding systems. There is a decrease in the number of groups actively working on natural products as drug leads and the governmental funding for this research has been effectively ceased in most of the developed countries. [9, 10]

3.2. History of antibiotics

The use of antibiotic-producing microbes stretches back to approximately 2000 years ago, when people from ancient civilizations in Serbia, China, Greece, and Egypt, used poultices from mouldy bread to treat open wounds. The Eber's papyrus from 1550 BC is the oldest preserved medical text that includes moldy bread and soil as remedies. [11]

A 1,000-year-old treatment for eye infections: Anglo-Saxon remedy that contained onion, garlic, and part of a cow's stomach, was shown to kill MRSA (methicillin-resistant *Staphylococcus aureus*) [12]

The beginning of the development of anti-infective drugs and the fundamental perception of chemotherapy is accredited to Paul Ehrlich. He developed the synthetic arsenic-based pro-drugs salvarsan (salvation arsenic) and neo-salvarsan about 100 years ago to treat *Treponema pallidum*, the causative agent of syphilis. This can be considered as one of the first systematic screens for drug discovery using a library of synthetic compounds. This was encouraged after Ehrlich's work on dyes that specifically stained bacterial cells. [11] Figure 1 gives an overview of the timeline and history of antibiotics.

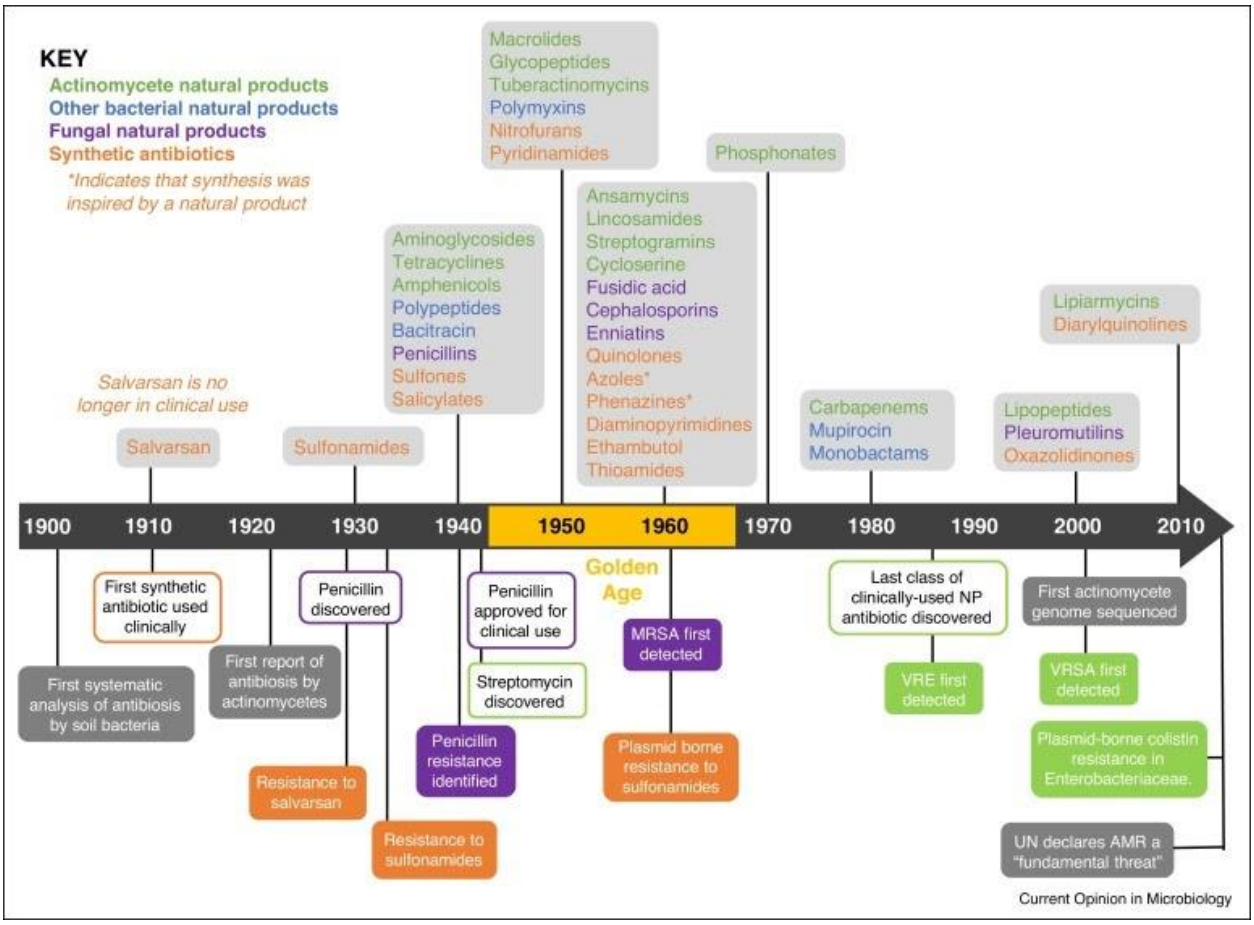


Figure 1: Timeline for the discovery of antibiotics [11]

Gerhard Domagk was responsible for the discovery of Sulphonamide and the sulphonamide prodrug Prontosil. Sulfonamides were the first truly effective, broad-spectrum antimicrobials approved for clinical use. This was superseded with the discovery of penicillin, which was discovered serendipitously in a contaminated petri-dish by Alexander Fleming in 1928. Later researchers worked on the purification and development of other beta-lactam derivatives. This enabled the development of semi-synthetic derivatives which would bypass penicillin resistance. Therefore, can be regarded as an important breakthrough.[11, 13, 14]

Louis Pasteur and Robert Koch coined the term “antibiosis” and proposed that microbes could secrete material to kill other bacteria. This was described in 1877, well before the discovery of penicillin. [15]

Later, in 1942, an American microbiologist Selman Waksman projected a systematic study of microbes as producers of antimicrobial compounds. Waksman defined an antibiotic as ‘a compound made by a microbe to destroy other microbes’ and was instrumental in identifying soil-dwelling filamentous Actinomycetales (‘actinomycetes’) as prolific producers of antimicrobial

compounds. Waksman discovered numerous antibiotics made by soil-dwelling actinomycetes, including neomycin and streptomycin, the first agent active against tuberculosis. [11, 15]

This work accelerated the process of discovery of antibiotics. During the period from 1950s to 1960s, most of the important classes of antibiotics, for example aminoglycosides, tetracyclines, macrolides were discovered. Among which, many drugs are still in clinical use and hence it is called the Golden Age of antibiotics. [11]

3.3. Need for the discovery of novel antimicrobials

The biggest problem with natural product research is the rediscovery of already isolated and characterized SMs. Researchers also face challenges like fermentation of producer strains and low yields of SM. Also, the incompatibility between existing NPs' libraries and the high throughput screening (HTS) methods, are becoming more and more important since the late 20th century. Moreover, pharmaceutical companies lost interest and abandoned NP research because the focus during this time was more on synthetic libraries of small molecules. [16]

Secondly, even at the initial stages of antibiotic discovery, resistant strains were often encountered as soon as new antibiotics were identified. Fig.2 shows a case of penicillin and its derivatives, where it can be seen that even before this antibiotic was used to treat bacterial infections, resistant strains producing the bacterial penicillinase (which destroys the beta-lactam- ring) were discovered. The detection of transferable plasmids, which harbour genes encoding antibiotic resistance, confirmed this problem. Entire populations of bacteria turn out to be resistant to certain antibiotics because of transferable plasmids. This finding revealed the problem of antibiotic resistance and how hard will it be to counteract this. [17]

Numerous bacterial species that are responsible for infections in humans become multidrug-resistant (MDR) after being treated with antibiotics. Due to multiple mutations, some bacteria are now resistant to many antimicrobial agents. These so-called "superbugs" are often found in hospitals. Currently, antibiotic resistance is among the major caused behind the unsuccessful treatment of infections. [16]

Methicillin, semi-synthetic penicillin, was the first antibiotic designed for resistant bacterial strains and was used against resistant bacteria producing penicillinase. Just 3 years after its first use in 1960, the first methicillin-resistant *Staphylococcus aureus* (MRSA), also known as the "hospital superbug", was found. [17] This problem of resistant Gram-positive and Gram-negative bacteria continues to grow and will eventually increase the already high requirement for novel antibiotics. [18]

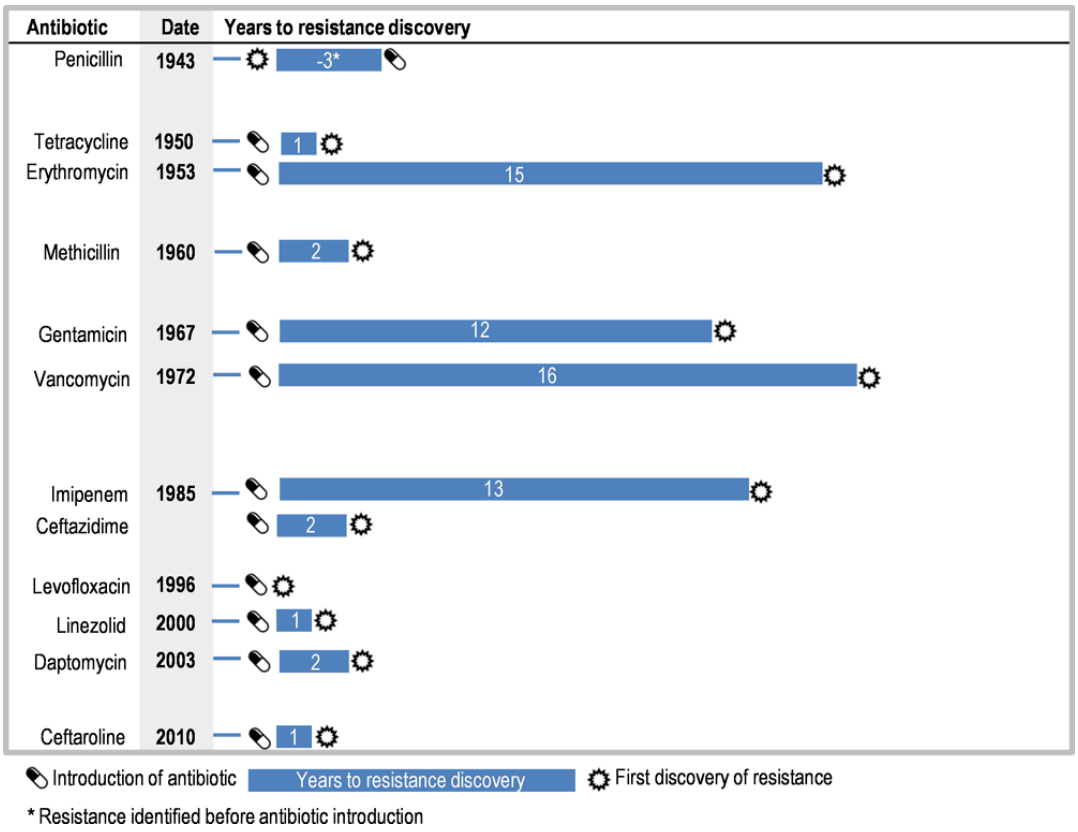


Figure 2: Discovery of most important antibiotic resistances from 1943 to 2010 [18]

Microbial NPs are structurally diverse and have a broad range of biological activities. These molecules have high molecular weight and most of them have stereo-specific carbon centers. This makes their chemical synthesis challenging. Even though lately, there has been a shift in the drug development process from natural products to totally synthetic drugs, most of these compounds still have either a pharmacophore from NPs, or are somehow derived from NPs. Thus, it can be said that NPs becoming more important in medicine. Thanks to advances in bacterial genomics and metagenomics as well as in the understanding of NP biosynthesis, bioinformatics, and other tools, science is possibly steering towards a new "*Golden Age*" of antibiotics. [8, 19–21]

3.4. Microorganisms as a source of drugs

Microorganisms are important producers of bioactive compounds. Research has shown that approximately 25% of the anticancer drugs which were in clinical trials in 2012, derived from microorganisms (Fig. 3). Doxorubicin, daunorubicin, actinomycin D, bleomycin, mitomycin C, etc. are approved anticancer drugs, produced by microorganisms. Another important category of NP-based medicines is antibiotics. Approximately, 70% of all compounds with antibacterial activity originate from NP.

As demonstrated in Fig. 3, approximately 51% of the drugs from NPs approved by the FDA, originate from bacteria, while 46% derive from fungi and 3% from plants. This highlights the importance of microorganisms in drug discovery. [22] Antibiotics are not only used for curing bacterial infections as a primary indication but are also used as co-medication, in conjunction with anticancer agents, or during organ transplantations.

In general, antibiotics can be categorized into the following major groups according to their different modes of action:

1. Inhibitors of protein synthesis by acting on the 30S or 50S subunit of the ribosomes e.g., aminoglycosides like kanamycin.
2. Inhibitors of cell wall synthesis e.g., penicillin.
3. Agents affecting the nucleotide synthesis e.g., sulphonamide antibiotics.
4. Agents preventing DNA replication by inhibiting DNA gyrase e.g., quinolones.
5. Inhibitors of bacterial DNA transcription e.g., the antituberculosis drug rifampicin

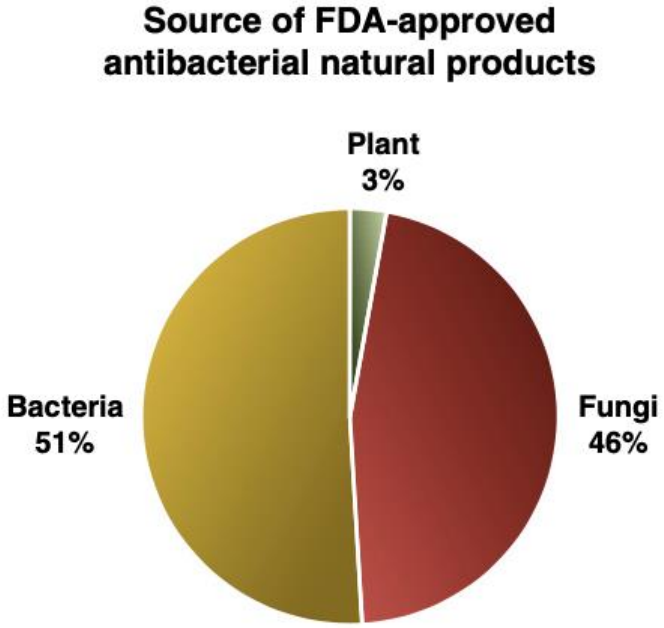
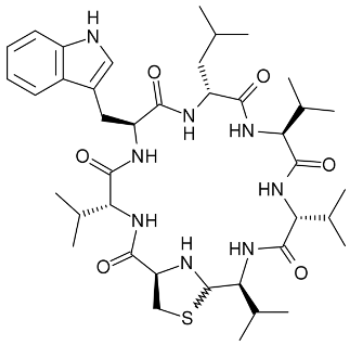
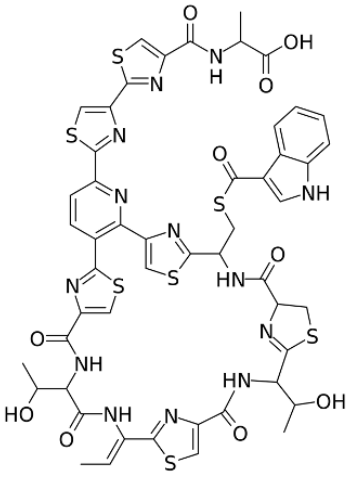


Figure 3: FDA approved drugs from natural products [22]

Efforts are being made in the development of new methods for isolating rare microbial species and cultivating previously uncharacterized bacteria from different environments. New approaches to investigating underexplored bacterial phyla showed interesting outcomes. [23] By prospecting biodiversity of human microbiota antibiotic lugdunin, which is a cyclic peptide isolated from *Staphylococcus lugdunensis*, was shown to be effective against Gram-positive bacteria. The second example is lactocillin, a thiopeptide antibiotic isolated from *Lactobacillus gasseri* and effective against vaginal pathogenic bacteria. [24]



Lugdunin



Lactocillin

Figure 4: Structures of Lugdunin and Lactocillin

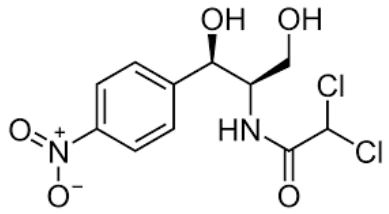
3.4.1. Actinomycetes as a source of antimicrobial secondary metabolites

Waksman's revolutionary work identified the genus *Streptomyces* as prolific producers of NPs. [11] These are compounds not required for the normal growth, development, or reproduction of an organism in the laboratory. Many *Streptomyces*-derived NPs are active against bacteria, fungi, viruses, nematodes, and insects and they have also been developed as anti-cancer and immunosuppressant drugs. [10]

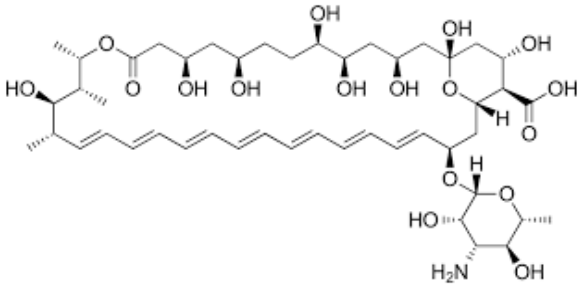
The majority of NPs derived from bacteria are produced by the family of *Actinomycetaceae*, which are filamentous Gram-positive bacteria. SMs from microorganisms, are structurally diverse because of their different natural habitat and have a greater chemical diversity than NPs from plants and animals. [20, 25] Discovery of bacterial metabolites started with the detection of "Actinomycin", an antibiotic isolated from *Actinomycetes* in 1940. [16] In the 1950-1970s, many antibiotics were discovered by isolating and screening bacteria producing various secondary metabolites. Prominent examples of discovered antibiotics were vancomycin and erythromycin, which were antibacterial agents, nystatin and amphotericin B used as antifungal agents, as well as anticancer drugs like doxorubicin and bleomycin. [13, 26]

- Antibiotics- vancomycin, erythromycin, tetracycline, chloramphenicol, gentamicin
- Antifungal- amphotericin B, nystatin
- Antitumor- doxorubicin, bleomycin
- Antiparasitic- ivermectin
- Immunosuppressant- rapamycin

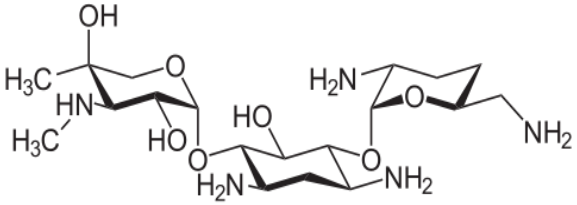
Several theories have been proposed to explain why soil microbes make so many bioactive NPs. The most likely explanation is that they have several functions, as chemical weapons as a defense against competitors in the soil either for protection or predation, as signaling molecules, or to mediate interactions with eukaryotic hosts such as insects and plants. This is consistent with evidence that *Streptomyces* species and other filamentous actinomycetes evolved roughly 440 million years ago, around the same time when plants first colonized the land. [24]



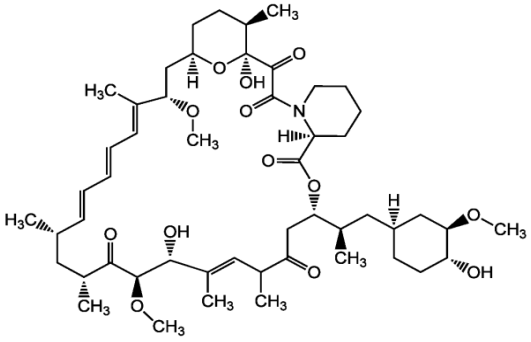
Chloramphenicol



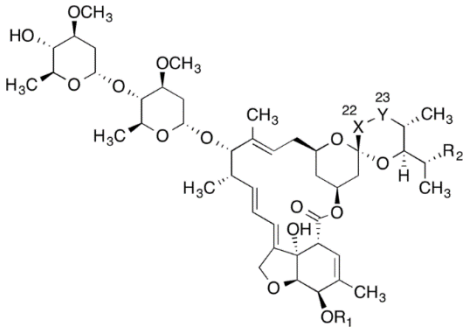
Amphotericin B



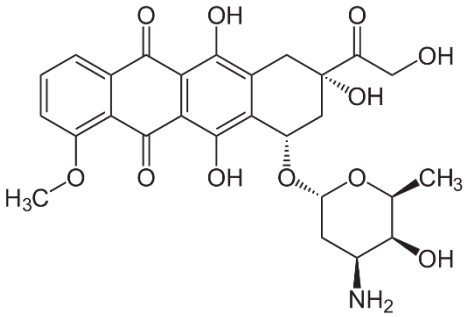
Gentamicin



Rapamycin



Ivermectin



Doxorubicin

3.5. Biosynthetic gene clusters (BGCs)

NPs from microorganisms are produced through metabolic pathways encoded by genes that are co-localized in the genomes of producing strains, forming the so-called biosynthetic gene clusters, BGCs. [27] The figure below shows a schematic overview of a typical biosynthetic gene cluster.

The central part of a BGC is called a "*core*" and contains genes encoding enzymes responsible for the biosynthesis of the antibiotic scaffold. Each NP group has its typical scaffold. Genes for scaffold modification are located near the core and these genes encode the enzymes responsible for scaffold modification, e.g., hydroxylases, methyltransferases, acyltransferases, halogenases, glycosyltransferases, etc. Additionally, BGCs contain genes encoding transporters, resistance proteins for self-protection, and regulatory proteins, that are necessary to control the transcription of the genes and thus for the regulation of the SM production. NRPS and PKS are the most famous examples of multi-modular antibiotic biosynthesizing core enzymes. [27–29]

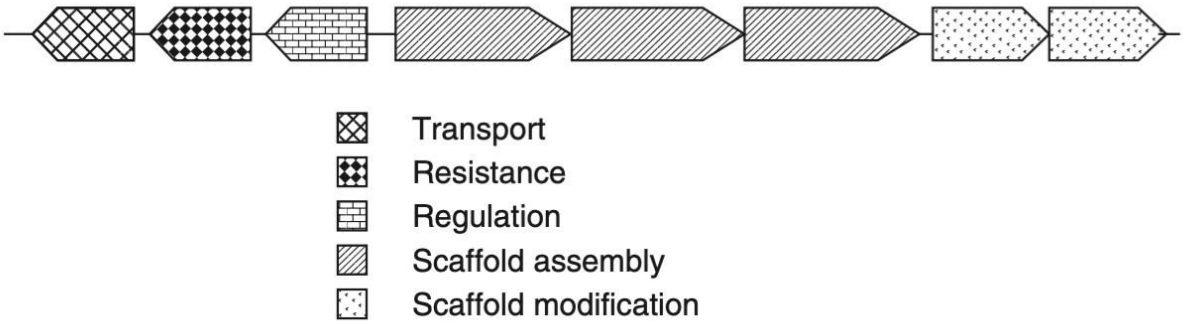


Figure 5: Illustration of a biosynthetic gene cluster [27]

3.5.1. Enzymes responsible for the biosynthesis of SMs

The two most common multi-modular enzymes NRPS and PKS type I, are among the largest and most complex enzymes in nature that are involved in the biosynthesis of SMs. In these biosynthetic systems, protein domains are organized into units called 'modules', which operate in assembly line fashion to construct polymeric chains and tailor their functionalities. Products of PKSs and NRPSs include several blockbuster medicines, therefore, it is important to understand the functioning of these enzymes in detail. In principle, the PKSs build polyketide chains from simple acyl-CoA building blocks, and the NRPSs assemble peptides from activated amino acids. [30]

3.5.1.1. Non-ribosomal peptide-synthetases (NRPSs)

These are responsible for the synthesis of non-ribosomal peptides, NRPs, without the need for the cell ribosomal machinery. NRPs are broadly distributed among microorganisms such as bacteria and fungi. NRPs were also isolated from higher eukaryotes which derive from their bacterial and fungal symbionts. [31]

Figure 6 depicts a typical NRPS with three modules **(A).** They can be categorized into an initiation module (blue), an elongation module (green), and a termination module (orange). All these modules work in coordination with each other. During the elongation step, a specific monomer (an amino acid) is identified, activated, and then finally fused into the growing peptide chain. This process is facilitated by enzymatic domains with specific functions within the module. [30, 31]

The activation of the NRPS-modules by the addition of a 4'- phosphopantetheine (PPT) prosthetic group to each T-domain **(B)** begins the process of NRP biosynthesis. The adenylation (A) domain selects and activates a specific amino acid by adding AMP and finally links it to the PPT prosthetic group of the downstream thiolation (T) domain **(C).** This marks the start of the catalytic cycle. Further, the substrate is attached to the T domain and passed onto the downstream condensation (C) domain. This catalytic domain assists in the formation of a peptide bond between this donor substrate (linked to the T domain of the same module) and the acceptor substrate (attached to the downstream T domain) **(D).** Later, the growing peptide is hydrolyzed from the upstream T domain and is now attached to the downstream T domain **(E).** At this point, either another elongation step follows or – in case the peptide is complete – the thioesterase (TE) domain of the termination module releases, and often cyclizes the product **(F)**. [30, 31]

Some important therapeutic agents synthesized by NRPS are- daptomycin, vancomycin, and cephalosporin C.

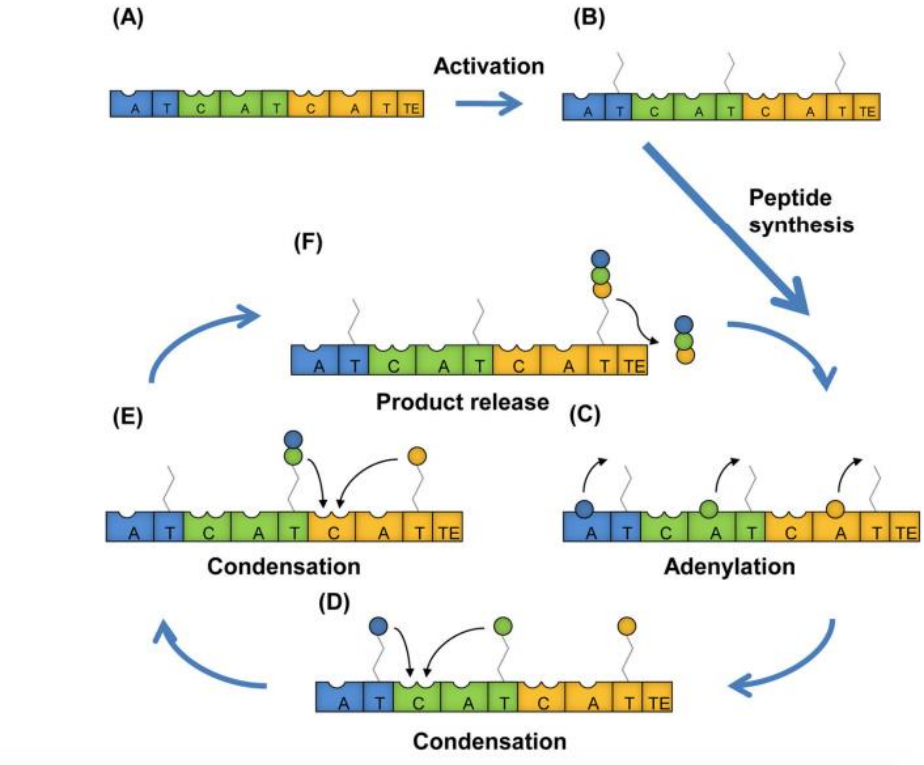


Figure 6: Illustration of NRPS modular system. [31]

3.5.1.2. Polyketide synthases (PKS)

The polyketide synthases represent another group of multi-domain enzymes responsible for the biosynthesis of polyketides using simple activated carboxylic acids as building blocks. PKs have been grouped into three classes: type I, II, and III PKS. Type I PKSs are large multi-domain enzymes which can function in successive or iterative way. These enzymes shuttle substrates between functional domains and acyl carrier proteins (ACPs). Modular type I PKSs have organized functional modules, each responsible for individual decarboxylation of acyl-CoA followed by condensation steps in polyketide formation. The iterative type I PKSs has functional domains within a single module and each of these domains carries out the steps in polyketide synthesis in a repetitive manner. [32]

Type II PKSs are multi-enzyme complexes that can be dissociable. Here, each protein represents a single and independent enzymatic domain iteratively used during polyketide synthesis. Another type of iterative PKSs is type III. These enzyme complexes do not require an ACP for the elongation of the polyketide chain. [33]

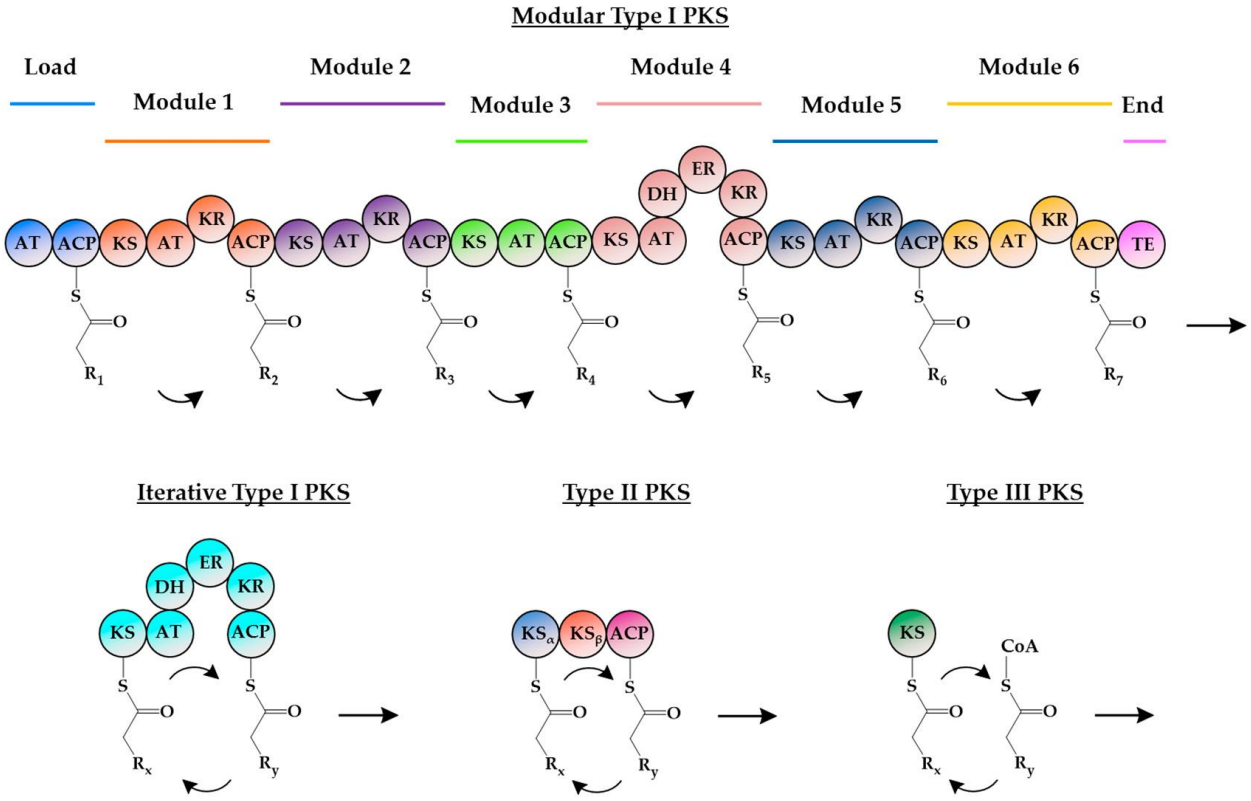


Figure 7: Illustration of types of PKSs. [32]

Some examples among the group of polyketides biosynthesized by PKSs are tetracycline, erythromycin, daunorubicin, rapamycin, lovastatin, and many others. [33]

The initiation module of type I PKS consists of an acyltransferase (AT) domain which starts building of a polyketide chain by selecting a precursor acyl-CoA and transferring it to the acyl carrier protein (ACP) domain within the same module, which was previously modified with 4-phosphopantetheine.

The elongation module consists at least of a ketoacyl synthase (KS) domain for the decarboxylative Claisen condensation between the growing polyketide chain and the extender unit, an AT domain responsible for the selection and activation of an extender monomer and an ACP domain. Modifications of the incorporated extender unit can be achieved by reduction steps specified by the dehydratase (DH) domain, a ketoreductase (KR) domain, or an enoyl reductase (ER) domain which may, or may not be present in the extender module.

The termination module additionally contains a thioesterase (TE) domain responsible for the release and cyclization of the final product [30, 34]. The scaffold of erythromycin A, a macrolide antibiotic, is produced by type I PKS. Its biosynthesis can be divided into two major steps: the formation of the polyketide backbone (macrolactone ring) and its modification by "tailoring" enzymes. Propionyl-CoA is the starter unit, recognized by the loading module. In the first phase, three different deoxyerythronolide B synthases (DEBS 1-3) each consisting of two modules, convert all methylmalonyl CoA extender units into the final product named 6-deoxyerythronolide B. After this point, additional tailoring reactions as the addition of deoxysugar groups, hydroxylation, and methylation of the polyketide core take place. [32, 33]

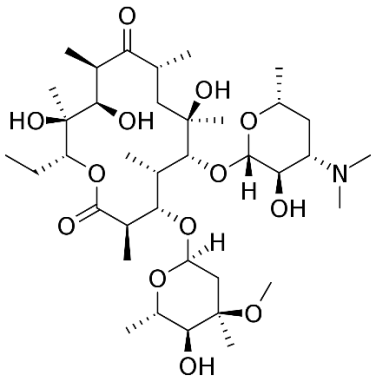


Figure 8: Structure of erythromycin A synthesized using type I PKS

As the biosynthetic pathway of type I PKS is well-studied, the engineering of these enzymes offers the possibility to make predictable changes in the macrolactone ring, and consequently, novel bioactive compounds can be obtained. [27]

The type II PKS consists of similar catalytic domains as type I PKS, but each is represented by one individual protein, and working iteratively as a complex. In difference to type I PKS, type II PKS consists of two KS domains KSa (responsible for the Claisen condensation) and KSb (controlling the polyketide length) and one ACP domain. Reductive processes and other

modifications can eventually take place after the biosynthesis of the whole scaffold. The resulting products of type II PKS are often polycyclic aromatic structures as e.g., chlortetracycline.

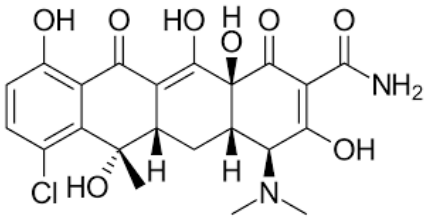


Figure 9: Structure of chlortetracycline, an example of type II PKS

Type III PKS are members of the chalcone synthase (CHS) and stilbene synthase (STS) family of condensing enzymes. They function as homodimers, working iteratively. Type III PKS uses free CoA thioesters as substrate and produces a polyketide backbone by condensation, working independently from the ACP domain. This type of PKS uses also malonyl-CoA as extender units. The type III PKS is responsible for the production of the polyphenolic stilbenoid, resveratrol. [32]

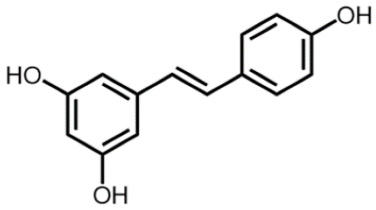


Figure 10: Structure of resveratrol synthesized from type III PKS

3.5.2. Genome mining

The genome mining approach is for example useful to find secondary metabolites of bacteria that might be used for drug discovery in a much more targeted manner. [35] This approach is a combination of bioinformatics for identification of cryptic BGCs in the bacterial genomes, chemoinformatics to predict the type of NPs they code for, genetic engineering that can help to activate these BGCs, and analytical chemistry to identify the true product of this BGC. [24, 36, 37]

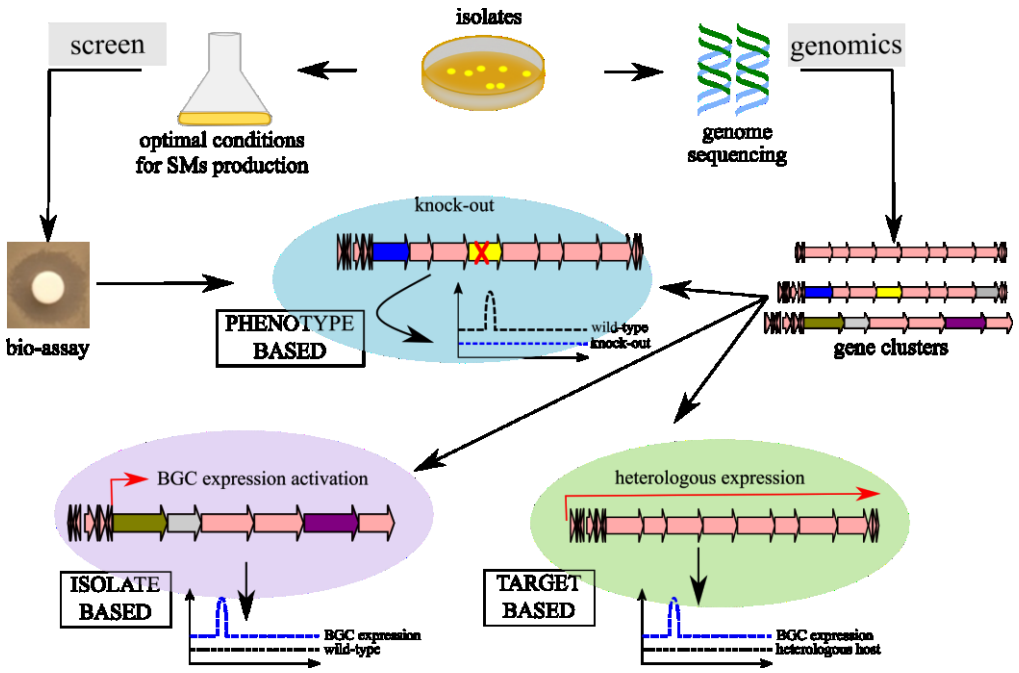


Figure 11: Different approaches in genome mining [21]

Genome mining can be roughly divided into three different types, based on the approaches used (Figure 11). In the phenotype-based approach, the isolated microbial cultures are screened and the bioactivity of extracts is analyzed via bio-assays. Once the desired bio-activity is detected, prediction of BGC encoding the bioactive compound is made with the help of genome sequencing data. Studies are done to verify the relationship between BGC and the bioactive compound by deactivating the core biosynthetic gene and testing the resulting mutant for bioactivity. If this mutant lacks bioactivity, it can be surmised that the deactivated gene is the one responsible for the production of bioactive SM. [21, 24, 25]

Another approach used in genome mining is target-based. Here, the choice of a specific chemical class of NP is determined from analysis of BGC with software antiSMASH [38] and it is the starting point. The genome sequencing data are analyzed to find the chemical class of interest and the BGC responsible for its biosynthesis. This identified BGC is then expressed in a well-known or engineered heterologous host, or activated in the original host via manipulation of regulatory genes. The metabolite profile of a recombinant strain can be analyzed for example via HPLC and if the expression was successful and the produced compound could be detected via an analytical method, the isolation and characterization of a target molecule can be carried out. [21, 24]

An alternative approach presented in Figure 11 is based upon the uniqueness of the bacterial isolate that may be a promising source of novel metabolites. In this technique, the genome of a unique isolate is sequenced and the data analyzed to detect putative new BGCs. The expression of chosen BGC could be activated or up-regulated by genetic manipulation of transcriptional regulators.

Additionally, a comparative study of the metabolomic profile of extracts from the recombinant strain and the wild type is performed. [21, 24]

3.5.2.1. Biosynthetic gene cluster activation in a native host

As mentioned earlier, the majority of the BGCs are silent under standard laboratory conditions and they need to be activated to enhance the production of NPs. One approach for the discovery of bioactive compounds from cryptic BGCs is the manipulation of transcription factor-encoding genes that control SM production. The information from the bioinformatics analysis can help to control and predict the nature of these pathway-specific genes. Typically, the pathway-specific regulators are located together with genes for resistance, biosynthesis, and secretion in BGCs. Therefore, the expression of many such BGCs can be induced by overexpression of activator genes or deletion of repressor genes. [39, 40]

The transcription factors acting as repressors prevent the expression of a specific gene or sets of genes in various ways. For instance, they compete with an activator for binding at the promotor region or they bind directly at the promotor region and prevent access of the RNA polymerase, which leads to an inhibition of transcription initiation. They can also bind downstream of the promotor and thus prevent the elongation of transcription. [41]

An important example for a pathway-specific repressor in *Streptomyces* is the tetracycline repressor family (TetR). Some TetR-like proteins contain a DNA-binding domain at their N-terminus and a ligand-binding domain at the C-terminus, which can undergo a conformational change upon interaction with a ligand. [40]

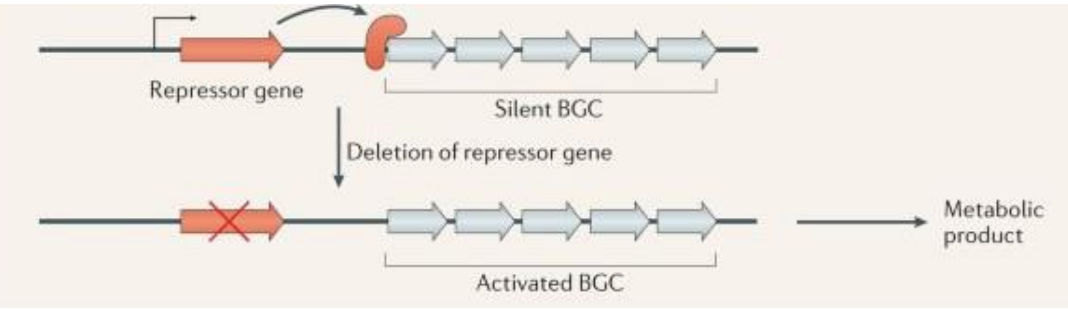


Figure 12: An illustration of the inactivation of a repressor gene in a silent BGC. [39]

Some transcription factors act as activators of silent BGCs and work by stabilizing the initial RNA polymerase-promotor complex or accelerate the transition to an open complex after the RNA polymerase has bound to the promotor. [41]

After the gene encoding an activator is identified, its amplification is carried out via PCR and it is subsequently cloned into a vector under control of a strong constitutive promotor. The resultant

recombinant plasmid is then integrated into the genome of the host strain to overexpress the regulatory gene and thus turn on SMs production. [42]

The SARP proteins are common pathway-specific positive regulator family among actinomycetes. [43] Their N-terminus contains a helix-turn-helix domain to bind the promotor region of the gene of interest and has a transcriptional activator domain on the C-terminus. Most SARP-family members act at the end of the signal transduction cascade and lead to an up-regulation of biosynthetic genes expression in the specific cluster. [43] The SARP-encoding genes are present in different BGCs, like type I and type II PKS clusters, non-ribosomal peptide synthetase clusters, etc. [1, 41] Other types of positive regulators found in *Streptomyces* can be LuxR, OmpR, etc.

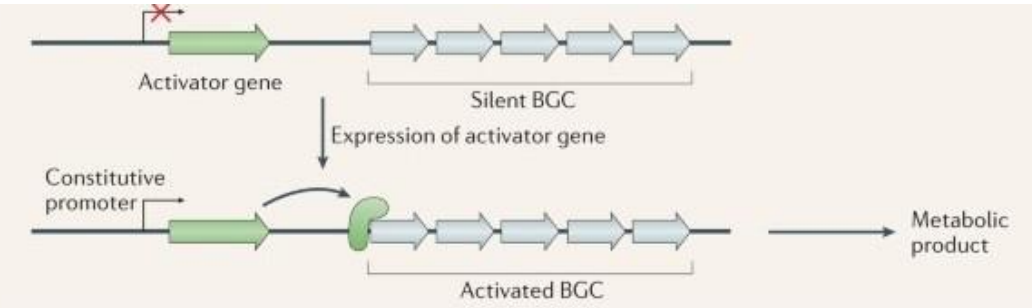


Figure 13: An illustration of overexpression of an activator gene in a silent BGC [39]

3.5.2.2. Heterologous expression

Heterologous expression is another approach to activate and express silent BGCs of actinomycetes in engineered host strains. It can be regarded as a BGC-specific approach of genome mining for the discovery of new bioactive compounds. Although the production of secondary metabolites is not optimal using this approach, it is a promising technique that uses organisms that are rapidly growing and easy to modify compared to the original producer strain. [43–45]

The crucial elements for a successful heterologous expression of BGCs include a suitable stable vector carrying the BGC of interest and a suitable heterologous host amenable to genetic manipulation to enable the BGC´s expression. Even though appropriate host organisms and vectors may be used, this does not guarantee a successful expression of heterologous BGCs. [45, 46]

A typical workflow for the heterologous expression is as follows: [39, 46]

1. Isolation of the DNA from native producer and cloning into suitable vectors (usually cosmids, fosmids, or bacterial artificial chromosomes (BAC)) to generate genomic libraries.
2. Screening of the generated genomic library to identify clones containing BGCs of interest or their parts.

3. Reconstruction of the BGCs through homologous recombination in yeast using a shuttle vector.
4. Introduction of the entire BGCs into a heterologous host, coupled with genetic manipulation for a better expression.
5. Optimization of the compound production from the recombinant strains by optimizing growth conditions.

For capturing large BGCs (up to 200 kb in size) BAC vectors are used. Smaller DNA inserts with around 45 kb can be carried by cosmids and 100-300 kb DNA fragments are harboured by P1-derived artificial chromosome (PAC) vectors. These vectors may thus carry an entire BGC. For successful and sustainable heterologous expression in the host strain, a vector shall preferentially encode an integrase for stable integration into the host genome. [47]

For the selection of a suitable host for heterologous expression, it is necessary to know the genetic and physiological properties of native and host strain producers. Besides, a better expression of the BGC can be reached when heterologous organisms and native producers are closely related. If this is not the case, an exchange of promotors can provide for increased expression and thus higher compound yields. [46]

3.5.2.3. Combined approaches

Even though bacteria are prolific producers of new NPs, some of them are difficult to cultivate on an industrial scale. The problem of low growth rate and therefore slow accumulation of biomass to have sustained high yielding production persists. Hence, heterologous expression and production in other bacterial hosts can be a promising alternative. Combined approaches have been developed wherein, heterologous expression is coupled with overexpression of activators or replacement of regulatory elements e.g., promoters.

E. coli and *Streptomyces* are regarded as suitable heterologous hosts. *E. coli* is simple to cultivate and its toolbox for gene manipulation is already well-developed. However, some SMs cannot be produced in this bacterium as it lacks certain precursors or sigma factors for the recognition of foreign promotors. Also, the problems of protein folding or toxicity of the final compounds make the process complicated. [24]

To overcome these hurdles, bacteria of the genus *Streptomyces* were developed as heterologous hosts. Their capacity of producing precursors, performing the post-translational modifications, and availability of resistance genes to counteract toxic metabolites are advantageous when compared with *E. coli*. Because *Streptomyces* synthesize often their own SMs which can lead to competition with heterologous pathways and complications with the purification of NPs, such endogenous pathways have to be eliminated via deletion of the corresponding BGCs. In the well-known host *Streptomyces coelicolor* M145, four BGCs were deleted to enhance the precursors' pool and to create a simplified metabolic profile. [39]

The combination of several methods of genome mining is becoming more common, as exemplified for the strain *Streptomyces albus* J1074, which is often used for heterologous expression. Five BGCs in this strain's genome were manipulated by either placing a strong constitutive promotor in front of the biosynthetic genes, overexpressing pathway-specific regulators, or via gene deletion. This combined approach has resulted in the isolation of new polycyclic tetramate macrolactams, an activation of candicidin and antimycin biosynthesis, and identification of paulomycins BGC. [39]

3.5.3. Cloning techniques

3.5.3.1. In vitro

Cloning of large BGCs, some of which can be over 100 kb in size makes heterologous expression challenging. Many cloning techniques have been developed to enable the process of heterologous expression. *CATCH* - Cas9-assisted targeting of chromosome – is a single-step *in vitro* cloning technique for large DNA fragments up to 100 kb. This method is using RNA-guided Cas9 nuclease, which cuts bacterial chromosomes at specific target sequences in low melting agarose gel. [48] The resulting digested fragments can then be ligated into a cloning vector, which has a 30 bp overlap at both ends with the target DNA by *Gibson assembly*. The recombinant plasmid thus created, is subsequently electroporated into a cloning host. Gibson assembly is a single-pot, isothermal strategy to assemble overlapping DNA fragments by using a 5´exonuclease, a DNA polymerase, and a DNA ligase. [49]

Successful cloning of 36 kb jadomycin BGC from *Streptomyces venezuelae* and the 32 kb chlortetracycline BGC from *Streptomyces aureofaciens*, was accomplished using this approach. [47, 48]

3.5.3.2. In vivo

An example of *in vivo* cloning technique is the TAR (Transformation Associated Recombination) system which utilizes natural recombination between homologous DNA fragments in *Saccharomyces cerevisiae* upon transformation. [50] This method is well suited for direct cloning of BGCs with a size of up to 150 kb. [47]

DNA representing flanking fragments of the BGC of interest can be amplified via PCR and cloned into a shuttle vector, generating a "capture" vector. Transformation of such linearized vector into yeast cells together with overlapping DNA fragments representing a BGC results in the assembly of complete BGC in the vector. The resulting construct can then be introduced into an engineered host strain for heterologous expression. Taromycin, an NRPS compound of BGC with size 67 kb was successfully expressed in heterologous host *Streptomyces coelicolor* M1146 using vector pCAPOl with the TAR cloning method. [51] Demethylchlortetracycline, a Type II PKS is another example of NP produced by TAR cloning followed by heterologous expression in *Streptomyces aureofaciens* DM-1. [52, 53] Figure 14 shows an outline of the TAR cloning system. [44, 50]

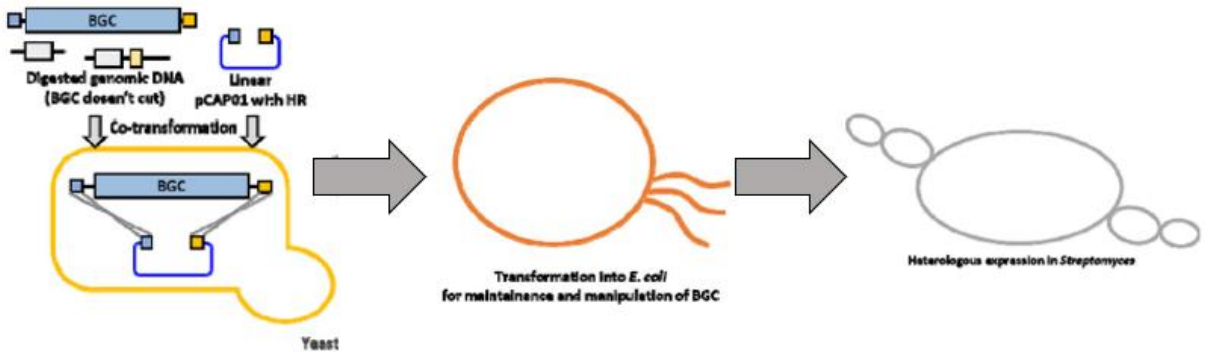


Figure 14: TAR cloning system [50]

Another *in vivo* cloning technique for BGCs is phage-mediated homologous recombination. It is based on the bacteriophage *λ Red system*. This system can catalyze recombination between short homology regions and works even without RecA, which is an important protein for homologous recombination in *E. coli*. λ bacteriophage encodes its recombination functions and contains genes, *exo*, and *beta*. *Exo* degrades dsDNA from 5´end and exposes 3´overhanging bases which can be bound by *beta*. In presence of the complimentary strand, *beta* anneals the two strands to generate recombinants. Another gene, *gam*, ensures full recombination potential and inactivates the RecBCD enzyme, which can degrade inward DNA. [54] Novobiocin, a DNA gyrase inhibitor and bacillomycin, an anti-fungal agent, were successfully produced by cloning the BGC by Red/ET recombination. [55]

3.6. Implications of Genomics and Bioinformatics in drug discovery

Genomics recovers information from individual whole genomes. This, when combined with transcriptomics, proteomics, lipidomics, and metabolomics enable robust high-throughput screening and discovery of novel bioactive compounds. These together encompass “Omics”, which is a combinatorial approach to high-throughput analysis of bioactive compounds and study all biological processes. Advancements in the next generation sequencing (NGS) technologies have opened an avenue for pan-genome analysis of microbes and made the gene level data of bacterial strains accessible. [56]

Recent studies have combined molecular biology with analytical chemistry methods for discovery of new drug leads from novel bacterial strains. This further emphasizes the need for advances in bioinformatics and chemoinformatics as vital instruments in the discovery process. The bioinformatics-based genome annotation could facilitate the identification of biosynthetic gene clusters (BGCs) and offer insights into the ability of various microbial strains for the compounds with potential bioactivity. Bioinformatics has also accelerated the detection of genes for secondary metabolite biosynthesis, and the types of compounds they specify. [56, 57]

The well-established online bioinformatics/chemoinformatics tools include the antibiotics and Secondary Metabolites Analysis SHell (anti-SMASH) [38], the antiSMASH database, Prediction Informatics for Secondary Metabolomes (PRISM), Global Alignment for Natural-products Cheminformatics (GARLIC), etc. Such tools with embedded databases offer a platform for analyses of SMs. They also enable alignment of gene cluster levels to their nearest relatives from a database containing other known gene clusters, mining of microbial genes and prediction of secondary metabolites BGCs, comparative alignment of different synthases' assembly lines, and so on. Through the identification of pathways, these databases allow screening and appropriate selection of strains with potential for drug discovery. [56, 58]

antiSMASH [38] is an open-source online tool that identifies regions in the microbial genomes that encode enzymes responsible for the biosynthesis of secondary metabolites, e.g., PKS, NRPS, terpenes, aminoglycosides, aminocoumarins, beta-lactams, siderophores, etc. It assists in the prediction, analysis, and detection of BGCs in bacterial genome sequences by comparing identified regions from the gene cluster to their closest matches from the MIBiG database [59], which contains all previously characterized BGCs. [24, 29, 38]

4. Aim of the work

Natural products and their derivatives are a promising source for the discovery and development of new and potential drug candidates. The traditional knowledge along with new technologies in genome sequencing, high-throughput screening, and combinatorial chemistry have greatly facilitated natural products research and revolutionized the process of drug discovery.

The main objective of this master thesis work was to discover novel bioactive compounds from microorganisms using synthetic biology approaches. Target-based bioprospecting which involves a combination of genomics, bioinformatics, molecular biology, and analytical techniques was used to study and synthesize secondary metabolites.

Actinomycetes are prolific producers of secondary metabolites. Microbial natural products and secondary metabolites are structurally diverse. These have different applications in the treatment of various diseases. For example, antibiotics, antitumor agents, antifungals, immunosuppressants, antiparasitic agents, and so on. Two bacteria from the phylum Actinobacteria: *Streptomyces bambergiensis* and *Actinoalloteichus fjordicus,* were chosen to study, analyze and synthesize putative novel secondary metabolites with bioactivity.

Series of experiments involving pooled PCR-based screening of the genome library, extraction of fDNA from the identified clones, carrying out restriction digestion, cloning, ligation, transformation, homologous recombination in yeast, conjugation in desired *Streptomyces* host for heterologous expression, and subsequent fermentation of recombinant *Streptomyces* strains for production of secondary metabolites were planned and carried out. Lastly, the aim was to analyze the extracts using HPLC and evaluate the bioactivity using disc-diffusion assay.

5. Materials

5.1. Media

5.1.1. Media used for growing *Escherichia coli*

Luria-Bertani medium (LB)

Bacto™ Tryptone	10 g
Difco Bacto Yeast extract	5 g
NaCl	10 g
Distilled water	Makeup to 1000 mL

L-Agar (LA)

Bacto™ Tryptone	10 g
Difco Bacto Yeast extract	5 g
NaCl	10 g
Agar	10 g
Distilled water	Makeup to 1000 mL

5.1.2. Media used for growing *Saccharomyces cerevisiae*

YPD medium

Difco Bacto Yeast extract	10 g
Bacto™ Peptone	20 g
Glucose	20 g
Distilled water	Makeup to 1000 mL

YPD agar

Difco Bacto Yeast extract	10 g
Bacto™ Peptone	20 g

Glucose	20 g
Agar	15 g
Distilled water	Makeup to 1000 mL

Y1376 medium (Yeast Synthetic Drop-out medium)

Sigma Yeast Synthetic Drop-out Medium Supplements	1.62 g
Yeast Nitrogen Base without Amino Acids	6.7 g
Distilled water	Makeup to 960 mL
After autoclaving add	
Glucose Solution (50% w/v)	40 mL

Y1376 agar

Sigma Yeast Synthetic Drop-out Medium Supplements	1.62 g
Yeast Nitrogen Base without Amino Acids	6.7 g
Agar	20 g
Distilled water	Makeup to 960 mL
After autoclaving add	
Glucose Solution (50% w/v)	40 L

5.1.2.1. Media used for growing *Streptomyces*

2X YT Medium

Bacto™ Tryptone	16 g
Difco Bacto Yeast extract	10 g

NaCl	5 g
Distilled water	Makeup to 1000 mL

Soy flour Mannitol Agar

Soy flour	20 g
D-Mannitol	20 g
Agar	20 g
Hot tap water	Makeup to 1000 mL

5.1.3. Fermentation media and media used for the production of secondary metabolites

TSB medium

Oxoid Tryptone Soya Broth powder (CM129)	30 g
Distilled water	Makeup to 1000 mL

YEME

Yeast extract	3 g
Bacto Peptone	5 g
Malt Extract	3 g
Glucose	10 g
Sucrose	340 g
Add 2.5 M $MgCl_2.6H_2O$ after autoclaving	2 mL
Distilled water	Makeup to 1000 mL

SM17 medium

Glucose	2 g
Glycerol	40 g
Soluble starch	2 g
Soya flour	5 g
Peptone (Oxoid L37)	5 g
Yeast extract	5 g
NaCl	5 g
$CaCO_3$	2 g
Tap water	Makeup to 1000 mL

MYM medium

Maltose	4 g
Yeast extract	4 g
Malt extract	10 g
Distilled water	500 mL
Tap water	500 mL
Adjust pH to 7.3	
Trace elements, after autoclaving	2 mL

Trace elements for MYM medium

$ZnCl_2$	40 mg
$FeCl_3.6H_2O$	200 mg
$CuCl_2.2H_2O$	10 mg

$MnCl_2.4H_2O$	10 mg
$Na_2B_4O_7.10H_2O$	10 mg
$(NH_4)6Mo_7O_{24}.4H_2O$	10 mg
Distilled water	up to 1000 mL

PM4-1 Medium

Glucose	15 g
Soy meal	15 g
Corn steep solids	5 g
$CaCO_3$	2 g
Distilled water	up to 1000 mL
TMS 1 Trace elements	6 mL

Trace Elements for PM4-1

$FeSO_4.7H_2O$	5000 mg
$CuSO_4.5H_2O$	390 mg
$ZnSO_4.7H_2O$	440 mg
$MnSO_4.H_2O$	150 mg
$Na_2MoO_4.2H_2O$	10 mg
$CoCl_2.6H_2O$	20 mg
HCl	50 mL/L
Distilled water	up to 1000 mL

5.2. Stock solutions

5.2.1. Antibiotic stock solutions

Table 1: Antibiotic stock solutions used for Streptomyces and Escherichia coli

Antibiotic	Stock concentration (mg/ml)	Used concentration *E. coli* (µl/ml)	Used concentration *Streptomyces* (µl/ml)	Solvent	Weight for preparation of 5 ml stock (mg)
Ampicillin (Amp)	100	100	-	H_2O	500
Apramycin (Am)	100	100	50	H_2O	500
Chloramphenicol (Cml)	25	25	-	Ethanol abs.	125
Kanamycin (Kan)	25	25	-	H_2O	125
Nalidixic acid (Nal)	30	-	30	0.1M NaOH	150
Thiostrepton (Thio)	30	30	30	DMSO	150

5.2.2. Solution for preparation of cell suspensions

Glycerol 20%

Glycerol	200 mL
Distilled water	Makeup to 1000 mL

5.2.3. Solution for preparation of competent cells

$CaCl_2$ (50mM)

$CaCl_2$.2H2O	14.7 g

Distilled water	Makeup to 2000 mL

5.2.4. Solutions used for pooled-PCR library screening

Sol I

Tris HCl (50 mM), pH 8	0.5 mL
EDTA (10 mM)	0.2 mL
Glucose (50 mM)	1 mL
Distilled water	Makeup to 10 mL

Sol II

NaOH (0.2M)	1.6 g
SDS (1% w/v)	2 g
Distilled water	Makeup to 200 mL

Sol III

KOAc (3M), pH 4.8	58.88 g
Distilled water	Makeup to 200 mL

L-arabinose (1M)

L-arabinose	3.00 g
Distilled water	Makeup to 20 mL

Tris HCl (10 mM), pH 8

Tris HCl (1M)	20 μL
Distilled water	Makeup to 2 mL

5.2.5. Solution used for yeast transformation

Lithium acetate solution (1M)

Lithium acetate dihydrate	1.02 g
Distilled water	Makeup to 10 mL

PEG solution (50%)

Polyethylene glycol (MW 3350)	100 mL
Distilled water	100 mL

5.3. Material for gel electrophoresis

TBE buffer (10x)

Tris base	12.1 g
Boric acid	6.18 g
EDTA	0.74 g
Distilled water	1000 mL

Agarose gel

Depending on the purpose, different concentrations of agarose gel (0.5 %, 0.8 %, and 2 %) were used during this work. Depending on the required concentration, Agarose-BioReagent for molecular biology by Sigma-Aldrich was dissolved in a 1X TBE buffer and GelRed Nucleic Acid Stain (10,000X) was added (5 µl/100 ml).

5.4. Strains, plasmids, and primers used during the project

5.4.1. Strains

Table 2: Microbial strains

Strain	**Type/Characteristics**	**Source**
Escherichia coli XL1-Blue MR	Engineered host strain used for transformation	[60]
Escherichia coli DH5 α	Engineered host strain used for transformation	-
Escherichia coli ET12567	Carries the vector pUZ8002 promoting mobilization function during conjugation	[61]
BAC-Optimized Replicator™ v2.0	Electrocompetent *E. coli* Cells	[62]
Escherichia coli EPI300	Electrocompetent *E. coli* strain, with a mutant trfA gene, which leads to a high replication rate	[63]
Streptomyces coelicolor M1154	Has four deleted native BGC and additional point mutations	[1]
Streptomyces albus J1074	Engineered host cell for heterologous expression, with a valine- and isoleucine-auxotrophic phenotype and is defective in the SalI (SalGI) restriction-modification system	[2]
Saccharomyces cerevisiae BY4742	his3Δ1selectable marker for recombinant plasmids	[64]

5.4.2. Plasmids

Table 3: Plasmid vectors used during the project

Plasmid vector	**Characteristics**
pCLY10	Shuttle vector, maintained in *E. coli, S. cerevisiae,* and *Streptomyces.* Resistance: LEU and Am^R

pCLY10_LR_C18	Shuttle vector pCLY10 with $Am^{R,}$ modified with flanking sequences for BGC 18
pCLY10_LR_C20	Shuttle vector pCLY10 with $Am^{R,}$ modified with flanking sequences for BGC 20
pSV_C18	Shuttle vector pCLY10 with BGC 18 from *S. bambergiensis*
pSV_C20	Shuttle vector pCLY10 with BGC 20 from *S. bamambergiensis*
pUWLoriT	Shuttle vector maintained in *E. coli* and *Streptomyces.* Has $Amp^{R,}$ $Thio^{R,}$ and ermE promoter region
pUWLoriT_pSS3_SARP	Shuttle vector maintained in *E. coli* and *Streptomyces.* Has $Amp^{R,}$ $Thio^{R,}$ and ermE promoter region for overexpression of SARP regulator gene from BGC3 in *Actinoalloteichus fjordicus*
pUWLoriT_pSS3_SARP	Shuttle vector maintained in *E. coli* and *Streptomyces.* Has $Amp^{R,}$ $Thio^{R,}$ and ermE promoter region for overexpression of SARP regulator gene from BGC4 in *Actinoalloteichus fjordicus*

5.4.3. Primers

Table 4: Primers for pooled PCR library screening of S. bambergiensis BGC 18 and BGC 20

Primer	Sequence	Size of the product
SBam_C18_F1	GACGAGGTCGTGTTCTGGAAG	705 bp
SBam_C18_R1	ATACCGGAAGGTGTGACGAG	
SBam_C18_F2	ATGTCATCGCGCTCAATGCC	475 bp
SBam_C18_R2	GAGACCCTCGCTCAGAAATC	

SBam_C18_F3	CTTCGCCGAGGTCCAATTGC	917 bp
SBam_C18_R3	GGACATCGACGTGTCCGTAG	
SBam_C20_F1	CGATGCTGCCGTTCTTCTAC	759 bp
SBam_C20_R1	CATGGGCTCAGCCTGACTTC	
SBam_C20_F2	AGTACCGGGAGTACCGGAAG	848 bp
SBam_C20_R2	ACGACACGTTCGTGGAACTC	
SBam_C20_set2bF	CACGGAGTTCGGAGTCTGAG	763 bp
SBam_C20_set2bR	GTGCCTCCTCACTGGAGTTG	
SBam_C20_F3	CCGCCGAGTAGTTCAGGTTC	790 bp
SBam_C20_R3	GTTCGACATCCGGGTCAACG	

Table 5: Primers to produce sticky ends on vector pCLY10 for cloning S. bambergiensis BGC18 and BGC20

BGC18_SBAM_ApaI	AATGGGCCCGCGTCACCGTCCCGTAGTCC	531 bp
BGC18_SBAM_PmeI	CGAGTTTAAACTCGGCGGEGCAAGGGATCTAC	
BGC18_SBAM_PmeIb	CGAGTTTAAACCGACGCCTGACGAGTCAGC	649 bp
BGC18_SBAM_HindIII	GCTAGAAGCTTGCGTGGTGCGCTGTTGGTTG	

BGC20_SBAM_ApaI	AATGGGCCCGTGACCAGGCAGAACTTTCG	631 bp
BGC20_SBAM_PmeI	CGAGTTTAAACATCTTGCGCCGGTAGGAGTC	
BGC20_SBAM_PmeIb	CGAGTTTAAACGACGGCCTCACCTCCGACAC	689 bp
BGC20_SBAM_HindIII	GCTAGAAGCTTTGTGCAGCATCGTGGTCATC	

5.5. Kits

1. To isolate plasmid DNA:
   a) Wizard® Plus SV Minipreps DNA Purification System (PROMEGA)
   b) Monarch® Plasmid Miniprep Kit
1. To purify PCR products: DNA Clean & ConcentratorTM-5 (Zymo Research)
2. For PCR
   a) Q5® High-Fidelity 2X Master Mix (New England BioLabs)
3. Monarch® DNA Gel Extraction Kit
4. QiaeX II- Qiagen Gel Extraction Kit for isolation of larger fragments

6. Methods

6.1. *In silico* methods

6.1.1. antiSMASH and BLAST

antiSMASH is an open-source bioinformatic tool that can be used to predict, analyze and align secondary metabolite gene clusters in bacterial genomes. [38]
antiSMASH has access to BLAST, which is an alignment tool used to compare the amino acid sequence of BGC to sequences in an open-source database of NCBI (National Centre for Biotechnology Information). In addition, antiSMASH is supported by other tools (HMMer 3, Muscle 3, FastTree, PySVG, and Jquery SVG), which enables comparison between already known gene clusters encoding for secondary metabolites, e.g., polyketides (PK), nonribosomal peptides (NRP), etc. within a query sequence. Thus, similarities between sequences can be predicted.

6.1.2. Primer design

Primers were designed using the software Clone Manager 10. [65] All primers were ordered from Eurofins Genomics. Forward and reverse primers were designed for the pooled-PCR-based screening of the genome library. To perform this polymerase chain reaction (PCR), three pairs of primers were designed for BGC18 and another three for BGC20. To create sticky ends in a shuttle vector, for homologous recombination, suitable primers for both the BGCs were also designed.
Altogether, all the primers used during this work have a length between 1 and 30 base pairs (bp). The annealing temperature was set according to the recommendation by software Clone Manager, usually between 60 °C and 62 °C. To avoid false binding, the temperature shouldn't differ more than 4 °C from this range.

6.2. Microbiological methods

6.2.1. Preparation of media

Different types of media were used to carry out experiments during this project. All the ingredients are listed above in chapter 4. The media were then autoclaved at 121 °C for 20 min. The pH was adjusted before autoclaving, when necessary.

6.2.2. Preparation of antibiotic stock solutions

The required quantity of antibiotics was weighed and dissolved in the appropriate solvent, as listed in table 1 in chapter 4. Later, all the antibiotic solutions were filtered sterilized, except chloramphenicol and thiostrepton, which have 96 % ethanol and DMSO as a solvent, respectively. All antibiotics were stored at -20 °C.

6.2.3. Cultivation of *Streptomyces* and *Escherichia coli*

6.2.3.1. Cultivation of *Streptomyces*

For the cultivation of *Streptomyces* strains, 30 mL of the corresponding media were initially poured into Petri dishes. They were left to dry for about 30 min and then *Streptomyces* strains were either

taken from freshly grown colonies or a glycerol stock and spread on the plate, using a sterile loop. The plates were incubated at 28 °C for at least one week until they sporulated.

6.2.3.2. Cultivation of *E. coli*

6.2.3.2.1. Cultivation of *E. coli* in liquid media

Individual colonies of *E. coli* were picked with a sterile loop or a sterile wooden stick from agar plates and transferred into cultivation tubes containing 3 mL of LB medium mixed with required antibiotics if needed. The cells were cultivated overnight at 37 °C in a shaker incubator at 200 rpm.

6.2.3.2.2. Cultivation of *E. coli* on solid media

30 mL of LB agar with required antibiotics was poured onto sterile petri dishes. After drying for approximately 30 min either single colonies after transformation were picked and spread onto the agar plate with a sterile loop, or 100 µL of spore suspension was pipetted onto the agar plate and spread with a sterile loop. The agar plates were incubated overnight at 37 °C.

6.2.4. Storage of strains

Selected clones were preserved as 20% glycerol stock for future use. The selected clone was picked from the respective LA plate and inoculated in liquid culture supplied with a selective marker. The culture was incubated and the cells were allowed to grow for approximately 18 hours. The next day, 1 mL of *E. coli* overnight culture was centrifuged. The supernatant was discarded and the pellet was resuspended in 1 mL glycerol 20 % under a sterile laminar airflow hood. It was then transferred to cryogenic tubes and stored at -80 °C.

For the storage of *Streptomyces* strains, 5 mL of glycerol 20 % were pipetted onto the edge of an agar plate, having fully grown and well-sporulating cells. With the tip of the pipet, the strains were carefully scratched from the surface. The suspension was transferred to a sterile syringe with cotton wool as a filter and then into a sterile cryogenic tube. Aliquots of approximately 1 mL were stored in cryogenic tubes at -80 °C.

6.2.5. Preparation of competent *Escherichia coli* cells

6.2.5.1. Preparation of competent *E. coli* cells by $CaCl_2$-method

The competent *E. coli* cells are prepared using CaCl2-method according to the following protocol: 1 mL overnight culture is transferred into 100 mL of fresh LB medium in a 200 mL flask. The culture is incubated in a shaker at 200 rpm at 37 °C. The OD_{600} of the culture is to be monitored intermittently after approximately 3 hours of growth, using a UV/Vis spectrophotometer. When an OD600 is 0.4-0.5, chill it on ice. The culture is then transferred into sterile Falcon™ tubes. Cells are collected by centrifugation at 6000 rpm for 8 min at 4 °C and discard the supernatant. Resuspend the cells in 20 mL of ice-cold 50 mM CaCl2. Incubate on ice for 20 min. Centrifuge the cells at 6000 rpm for 8 min at 4 °C. To store the cells as frozen stock for a long period, resuspend the cells in 2.5 mL of ice-cold 50 mM CaCl2 containing 10 % of glycerol. Dispense the

competent cells into aliquots of 100-200 µL into 1.5 Eppendorf® tubes and freeze in liquid nitrogen for later use. Store these aliquots at -80 °C.

6.2.5.2. Preparation of competent *E. coli* cells for electroporation

Electroporation was used to transform pSV_C18 and pSV_C20 into *E. coli* EPI300 cells following the protocol:

E. coli XL-Blue cells were inoculated in 10 mL LB medium and incubated at 200 rpm overnight at 37 °C. 0.5 mL of the resulting overnight culture was inoculated into a 50 mL LB medium. The culture was incubated at 37 °C, shaking at 200 rpm for 3-4 h until it reached an OD600 of approximately 0.4-0.6. Once the required OD was reached the cells were recovered in a 50 mL Falcon™ tube by centrifugation at 4000 rpm for 5 min at 4 °C. The supernatant was decanted, and the pellet was resuspended by gently mixing it in 40 mL ice-cold sterile 10% glycerol with the following centrifugation. The pellet was resuspended in 30 mL ice-cold 10% glycerol, recovered again via centrifugation, and finally resuspended in the remaining volume of 10 % glycerol. Aliquots of 50 µL were prepared and used for transformation.

6.2.6. Polymerase chain reaction (PCR)

The primers that are designed using clone manager are ordered from Eurofins genomics and are diluted in a given volume of fresh autoclaved water, vortex it to mix it well. The primer concentration is 100 pmol/ µL. For the PCR reaction, a dilution of 1:10 was made to obtain a final concentration of 10pmol/ µL which is needed for the PCR. All primers were stored at -20 °C. For this project, two types of polymerases were used- Q5® Master Mix and Hot Start Taq DNA Polymerase (NEB).

The following tables show the setup for the PCR and the thermocycling conditions, which were optimized after multiple reactions with wild-type *S. bambergiensis* genomic DNA and extracted fosmids as templates.

Table 6: PCR with Q5® Master Mix

Reagents	**Volume (µL)**
Q5 High-Fidelity 2X Master Mix	12.5
10 µM Forward Primer	1.5
10 µM Reverse Primer	1.5
Template DNA	1
Nuclease-Free Water	Up to 25

Table 7: PCR with Taq DNA Polymerase

Reagents	**Volume (µL)**
Taq Standard Reaction Buffer	2.5
dNTPs	0.5
10 µM Forward Primer	1
10 µM Reverse Primer	1
DMSO	1
Template DNA	1
Taq DNA polymerase	0.2
Nuclease-Free Water	Up to 25

Table 8: Thermocycling conditions for Q5® Master Mix, 25 cycles

Step	**Temperature**	**Time**
Initial Denaturation	98 °C	3 min
Denaturation	98 °C	30 sec
Annealing	Different for respective primers used. Usually between 58 to 62°C	30 sec
Extension	72 °C	30 sec/kb
Final Extension	72 °C	7 min
Hold	4 °C	

Table 9: Thermocycling conditions for Taq DNA Polymerase, 30 cycles

Step	Temperature	Time
Initial Denaturation	95 °C	3 min
Denaturation	95 °C	30 sec
Annealing	Different for respective primers used. Usually between 58 to 62°C	30 sec
Extension	68 °C	1 min/kb
Final Extension	68 °C	7 min
Hold	4 °C	

6.2.7. Pooled-PCR library screening

To perform the library screening, 19 96-well-plates (microtiter plates) that harbor the genome of *Streptomyces bambergiensis* were used. The genomic DNA library created by Dr. Olha Schneider was used for this project. All 72 agar plates were prepared according to the protocol of replica plating. The gDNA library plates were allowed to thaw on ice before performing the experiments. The stamp was flamed two times after dipping into 96 % ethanol, dried, and cooled down before performing replica plating. Once it was cooled, it was dipped into the left side (48 wells) of the microtiter plate and then stamped on a LA + Cml (12,5 µg/mL) petri dish. The stamp was dipped again into the left side of the microtiter plate and a second LA + Cml (12,5 µg/mL) Petri dish was stamped and used as a backup. The same procedure was performed with the right side of the microtiter plate (48 wells). Plates were incubated overnight at 37 °C and the next day the fosmid DNA (fDNA) was isolated according to the protocol given below. The backup plates were stored at 4 °C in the cold room.

6.2.8. Isolation of fosmid DNA from *E. coli* grown on Petri dishes

Initially, fDNA of *E. coli* was isolated for an agar plate with 48 clones. Then, during the library screening, only the fDNA of *E. coli* clones from a plate that showed a band on the gel after performing a PCR was isolated. This isolation process was carried out according to the following protocol:

2 mL sterile LB medium was pipetted on each petri dish. Afterward, the colonies were washed down with the tip of a pipette or with an inoculation loop. The Petri dishes were held at an angle

and the cells were suspended in the LB medium on the agar plate. The suspension was then transferred into 2 mL Eppendorf® tubes. Centrifuge at 13000 rpm for 5 min, discard the supernatant. Resuspend the pellet in 200 µL of ice-cold Sol I mixed with RNAse (10 µL RNAse/10 mL Sol I) just before adding it to the cell pellet. Mix it well by using a vortex. To this add 400 µL Sol II stored at room temperature and invert the tubes, till the cloudy solution becomes clear. Add 300 µL of ice-cold Sol III and invert the tubes five times, till a white precipitate is seen. Centrifuge the tubes for 5 min at 13200 rpm. Transfer the clear supernatant to a fresh 2 mL Eppendorf® tube. Add 1 mL ice-cold isopropanol and invert the tubes to mix well and check for a smear. Incubate the tubes on ice for approximately 10 min. Centrifuge the tubes at 13200 rpm for 10 min. Wash the pellet with 200 µL ice-cold 70 % ethanol. Centrifuge again at 13200 rpm for another 5 min. Remove the ethanol using a pipette and let the pellet air dry for approximately 10 min. Rehydrate the pellet with 100 µL of dH2O and store at -20 °C.

6.2.9. Isolation and purification of fosmid DNA from *E. coli* liquid cultures

A large colony was inoculated in 20 ml LB medium containing chloramphenicol (12.5µg/ml) and L-arabinose (2 µL/ml). The incubation was carried out overnight at 37°C in a shaker incubator at 200 rpm. The next day 20 ml liquid culture was centrifuged at 13000 rpm for 1 minute at 4°C. After 800 µl of ice-cold Sol I was added, the pellet was resuspended using a vortex. The suspension was equally divided among 4 Eppendorf® tubes (2 ml). Immediately 400 µL of Sol II was added and tubes were inverted 10 times. Quickly 300 µL of ice-cold Sol III was added and mixed by inverting the tubes 5 times. The tubes were spun at full speed in a microcentrifuge for 5 minutes at 4°C. The clear supernatant was transferred to a fresh Eppendorf® tube and 300 µl of Protein Precipitation solution was added, mixed by inverting, and incubated on ice for 10 minutes. The tubes were spun at full speed in a microcentrifuge for 5 minutes at 4°C. The clear supernatant was transferred to a fresh Eppendorf® tube and 2µl RNase was added to each tube. After incubating for 1 hour at 37°C, the tubes were cooled down on the ice for 5 minutes. 0.8 times the volume of 2-propanol was added and mixed by inverting until no streaks were visible. The tubes were incubated on ice for 10 minutes. They were spun at full speed for 15 minutes at 4°C and the pellet was washed with 500 µL of 70% ethanol. The tubes were spun again at full speed for 15 minutes at 4°C, the liquid was drained away and left at room temperature for approximately 10 minutes for the pellet to dry. The pellet was resuspended in 100 µL 10 mM Tris/HCl (pH 8) and stored at 4°C.

6.2.10. Restriction digestion

During this project, various restriction endonucleases were used to cut the DNA at a specific position, to carry out the following experiments:

A. to generate the pCLY10-based capture vector
B. to analyze the recombinant plasmids

After sequencing the isolated fDNA, restriction digestion with specific restriction enzymes of the respective fosmids was performed. This restriction was performed to subsequently enable the

yeast-based DNA assembling technique as described in the following chapters. The reactions were set up according to the protocol stated by the manufacturer.

For the positive regulator genes *ompR* and *pucR*, restriction digestion of the carrying vector pUWLoriT and the respective positive regulator genes was carried out with the same restriction enzymes. Subsequently, after some intermediate steps, independent ligation reactions of the restricted vector with the *pucR*-gene and *ompR*-gene were performed to amplify these specific parts of the wild-type DNA to stimulate the transcription, thus increasing the biosynthesis of SM.

Table 10: Restriction digestion reaction

Reagent	**Vol (µL)**
Fosmid DNA (fDNA)	30
Buffer	5
Restriction enzyme	1.5
dH_2O	Up to 50

The reaction was incubated in a water bath at an appropriate temperature according to the activity of the enzyme.

Table 11: List of enzymes used

Nsi I
Mre I
Dra I
EcoR V
Psi I
Sbf II
Acl I
PspX I
EcoR V
Pme I

Apa I
Hind III
Age I

Table 12: Analytical digestion reaction

Reagent	**Vol (µL)**
pDNA/fDNA	3
Buffer	1.5
dH2O	10.5
Enzyme	0.5

6.2.11. Gel electrophoresis

During the project, agarose gel electrophoresis was used for different purposes because it is a convenient method to separate DNA fragments based on size. Gel electrophoresis was used to recover DNA fragments of the desired size that were previously restricted with restriction endonucleases. It was also used for DNA purification after performing PCR and to verify successful ligation.

6.2.12. Recovery of digested fosmid DNA fragments from 0.5 % agarose gel

After the gel electrophoresis of restricted fDNA on an 0.5 % agarose gel, the DNA was extracted from the gel with the following protocol:

A gel slice containing the DNA was excised and chopped using a scalpel and put into a 2 mL Eppendorf® tube. The volume of the solvents used is based on the weight of the cut gel. One volume of buffer-equilibrated phenol was added to the tube with a gel slice. Vortex for 10 seconds till a cloudy mixture is seen and freeze the tubes at -70 °C for 20 min. The frozen tubes were centrifuged for 15 min at room temperature at 13600 rpm. The upper phase containing the DNA was collected and placed into a clean 2 mL Eppendorf® tube. 1 volume of Phenol/Chloroform and 0.1 volume of 3M sodium acetate were added to this. The contents of the tube were mixed gently and centrifuged for 15 min at room temperature at 13600 rpm. The upper phase was collected again and transferred into a clean 1.5 mL Eppendorf® tube. To precipitate the DNA, 1 volume of ice-cold isopropanol was added, and the tube was gently inverted. Centrifugation for 15 min at room temperature at 13600 rpm was performed. The supernatant was removed, the pellet was washed with 0.7 volume of ice-cold 70 % ethanol, the tube was centrifuged for 15 min at room temperature at 13600 rpm and air-dried until the remaining ethanol was completely evaporated.

Ultimately, 40 µL of rehydration buffer from the Wizard® Plus SV Minipreps DNA Purification System Kit (PROMEGA) was added, and the DNA was dissolved overnight in a fridge at 4° C. Alternatively, QiaeX II- Qiagen Gel Extraction kit was used for isolation of larger DNA fragments within the range of 40 bp to 50 kb.

To verify the success of DNA extraction, gel electrophoresis with 2 µL of fDNA using 0.5 % agarose gel, at 100V for 40 min was carried out.

6.2.13. Recovery of plasmid DNA from 0.8% agarose gel

After performing restriction digestion of vectors with specific endonucleases and performing gel electrophoresis using 0.8 % agarose gel, the recovery of the corresponding DNA was performed following the protocol of the Monarch DNA Extraction Kit. DNA fragments of size 50 bp to 10 kb can be obtained after extraction.

6.2.14. Purification of a PCR product

PCR products used for ligation were purified using the protocol of DNA Clean & Concentrator™-5 Kit.

6.2.15. Ligation

DNA ligation is a common molecular biology technique to combine a vector and the desired fragment. The ligation was performed using the T4 DNA Ligase Buffer (10X) and the T4 DNA-Ligase from New England Biolabs. The reaction mixture was vortexed and centrifuged, then it was either incubated at 16 °C overnight or at room temperature for 10 min. After ligation, the resulting product was transformed into competent *E. coli* cells.

Table 13: Ligation reaction

Reagent	**Reaction 1**	**Reaction 2**	**Negative Control**
pCLY10 vector	1 µL	2 µL	2 µL
Insert 1	7.5 µL	7.5 µL	-
Insert 2	7.5 µL	7.5 µL	-
T4 DNA Ligation buffer	2 µL	2 µL	2 µL
T4 DNA ligase	1 µL	1 µL	1 µL
dH_2O	1 µL	-	15 µL

6.2.16. Transformation into *E. coli*

During the project, depending on the purpose for future applications of the construct, several competent *E. coli* strains (DH5α, XL-Blue, ET12567, and BAC-Optimized Electrocompetent Cells) were used. To the 100 µL of competent cells (described in chapter 5.2.5.1) stored at -80 °C and thawed on ice for about 30 min,1 µL of pDNA (plasmid DNA) or 20 µl of ligation reaction were added. The cells were incubated on ice for 30 min. Later, cells were subjected to heat shock at 42 °C for 45 sec. The tubes were incubated again on the ice. After 2-3 min, 1 mL of LB medium was added, under laminar airflow. Cells were allowed to regenerate at 37 °C for 1 hour and finally spread on an agar plate with appropriate antibiotics. The plates were stored in an incubator at 37 °C overnight.

6.2.17. Isolation of plasmid DNA from *E. coli*

The single colonies from a transformation plate were randomly chosen and inoculated in 2 mL LB medium with the appropriate antibiotic, at 37°C for approximately 16-18 hours. Plasmid isolation from liquid culture was carried out the next day following the protocol of Wizard® Plus SV Minipreps DNA Purification System (PROMEGA). The plasmid DNA thus isolated was stored at -20 °C.

6.3. Yeast-based DNA assembling technique

6.3.1. Yeast Transformation

Plate *Saccharomyces cerevisiae* BY4742 on YPD agar plate and incubate overnight at 30 °C. The next day, inoculate a single colony from the YPD agar plate in 5 mL of YPD liquid medium. Allow the yeast to grow overnight to saturation in rich (YPD) medium at 30°C. Back-dilute to A_{600} nm ~0.25 i.e. ~ 0.4mL in 10 mL of fresh YPD medium and let it grow at 30°C with shaking for roughly 4-5 h until A_{600} nm ~1.0. Monitor the OD at 600nm using a UV/vis spectrophotometer.

Preparation of Transformation master mix: Thaw the SS DNA (single-stranded salmon sperm DNA in water; 10 mg/mL) at room temperature. Heat it at 100°C on a heating block for 10 min. Incubate it for 10 min on ice.

Table 14: Transformation master mix

Components per 1 reaction	**Volume**
50% PEG (MW 3350)	240 µL
1M Lithium acetate	60 µL
SS DNA (boiled, cooled on ice)	50 µL

Once an OD ~1.0 is achieved, centrifuge 10 mL culture of YPD with BY4742 cells, for 5 minutes in a 15 ml- falcon tube at 6000 rpm and discard the supernatant. Resuspend the pellet gently using a pipette tip in 500 µl of sterile water and transfer it to a fresh, Eppendorf® tube. Spin the tubes at 6,000 rpm for 1 minute. Decant off the water. Later, resuspend the pellets gently with 400 µl of sterile 100 mM lithium acetate solution. Spin at 6,000 rpm for 1 minute. Remove the lithium acetate solution with a pipette tip and keep the pellet.

Table 15: Assembly reaction

Reagent	**Volume (µl)**
Linear pCLY10_LR	2
BGC fragments	5 (for each fragment)
dH2O	Up to 36

For negative control	
linear pCLY10_LR	2
dH_2O	Up to 36
For positive control	
pCLY10	2
dH_2O	Up to 36

Linearized fDNA and shuttle vector pCLY10 with inserted flanks that are generated using restriction endonucleases were introduced into BY4742 yeast cells for homologous recombination. For which, the following steps were carried out:

Add 350 µL of Master Mix to each tube of yeast on top of the pellet – do not mix. Add Assembly reaction DNA Mix. Vortex for 10-20 seconds, to ensure proper mixing of all the added solutions. Incubate at 42°C in a water bath for 45 minutes. Spin tubes at 6,000 rpm for 1 minute. Discard the viscous supernatant. Add 500 µl of synthetic complete drop-out medium (Y1376) and gently resuspend the yeast pellet with a pipette. Plate 500 µl of resuspended sample on Y1376 large agar plates and incubate at 30°C for 2 days.

6.3.2. Colony PCR of *Saccharomyces cerevisiae* transformants

To verify the success of yeast transformation, colony PCR of *Saccharomyces cerevisiae* transformants was carried out, according to the following protocol:

Single colonies of *Saccharomyces cerevisiae* grown on big Y1376 agar plates were picked randomly with a sterile wooden stick and inoculated on a new Y1376 agar plate. The plate was incubated at 28 °C overnight. The next day, 40 x 1.5 mL Eppendorf® tubes were labeled and filled with 20 µL NaOH (20 mM). A part of each colony was scraped off the agar plate with a sterile micropipette tip and transferred into respective 1.5 mL Eppendorf® tubes. The tubes were incubated at 95 °C on a heat block for 45 min. Then they were centrifuged at 13600 rpm for 10 min. 2 µL supernatant from each tube was used as a DNA template for colony PCR.

PCR was performed on three primer sets for respective BGCs using Taq polymerase. Agarose gel electrophoresis was carried out to verify whether the entire BGC was incorporated in the vector or not.

6.3.3. Isolation of plasmid DNA from yeast

The clones of *Saccharomyces cerevisiae* that gave signals in the agarose gel after PCR were picked with a wooden stick and inoculated in 5 mL Yeast Synthetic Drop-out medium-Y1376. The colonies were incubated at 30°C, shaking at 200 rpm overnight. The plasmid DNA from the yeast was isolated according to the following protocol: 2.5 mL of yeast overnight culture was centrifuged at 6000 rpm. 50 µL of previously flicked Zymolyase® was added to the tube before gently mixing the content. The cell suspension was incubated at 37 °C for 45 min. At the end of the incubation, the yeast cell suspension was centrifuged at 10000 rpm for 5 min and the supernatant was discarded. Later, for plasmid extraction, components of the Wizard® Plus SV Minipreps DNA Purification System kit (PROMEGA) were used. 250 µL Cell Resuspension Solution was added to the pellet for resuspension, then a 500 µL Cell Lysis Solution was added. The tube was inverted 4 times and incubated for 5 min at room temperature.10 µL of Alkaline Protease Solution were added and the tube was inverted 4 times before incubating it for 5 min at room temperature. After adding 700 µL of Neutralization Solution, the tube was immediately inverted 4 times. The lysate was centrifuged at 13600 rpm for 10 min at room temperature. The clear lysate was then transferred to the prepared spin column by decanting and centrifuging for 1 min at maximum speed. The spin column was washed two times with 750 µL of Column Wash Solution. Ultimately, the plasmid DNA was eluted in 60 µL Nuclease-Free Water, previously preheated at 50 °C, by incubating at 37 °C for 15 min. The isolated plasmid DNA was stored at -20 °C.

6.3.4. Transformation of pDNA from yeast via electroporation into the *E. coli*

After the yeast-based DNA assembling technique, plasmids pSV_C18_8, pSV_C18_ 15, pSV_C18_ 22, pSV_C18_23 were obtained. Each of these plasmids had to be transformed into EPI300 *E. coli* cells. The transformation was performed according to the following protocol:
Electrocompetent EPI300 *E. coli* cells, the plasmid DNA about to transfer, sterile LB medium and 0.2 cm sterile electroporation cuvettes were kept on ice before starting with the electroporation. 2 µL of DNA were transferred onto the wall of the electroporation cuvette, right after 50 µL of competent cells were added and gently mixed. For electroporation, a Gene Pulser Xcell Electroporation System (BioRad) was set to 200 W, 25 µF, and 2.5 kV; the expected time constant was in a range of 4.5-4.9 ms. Immediately 1 mL of ice-cold LB medium was added to the cells. The cells were incubated at 37 °C shaking at 200 rpm for 1 h in a 1.5 mL tube. After the incubation, the whole cell lysate was spread on an LB agar plate containing apramycin as a selective marker. The agar plate was incubated at 37 °C overnight.

6.4. Conjugation into *Streptomyces* strains

During the project, pDNA was transformed into ET12567 *E. coli* cells, and conjugation of plasmids from *E. coli* ET12567 to *Streptomyces albus* J1074 and *Streptomyces coelicolor* M154 was performed. The conjugation was performed according to the following protocol:
Selected clones of *E. coli* ET12567 transformants carrying the plasmid of interest were densely spread on an LA plate containing three different antibiotics for selection: kanamycin and chloramphenicol for keeping the pUZ8002 plasmid in ET12567 cells, and apramycin (for pCLY10-derivatives) or ampicillin (for pUWLoriT-derivatives). The plate was incubated at 37 °C overnight. The next day, the best-grown colony of ET12567 cells was chosen, swiped from the agar plate with a sterile loop, and suspended into 500 µL 2XYT medium in 1.5 mL Eppendorf® tubes.

Later, 50 µL of *Streptomyces* spore suspension was thawed on ice and was mixed with 350 µL of 2XYT medium in a 1.5 mL Eppendorf® tube and incubated for 7 min at 50 °C on a heat block. The reaction tubes were allowed to cool down at room temperature for approximately 10 min. 100 µL of ET12567 cell suspension was added to 400 µL of heat-shocked *Streptomyces* spores, the tubes were inverted gently and then centrifuged for 1 min at 5000 rpm. Approximately 300 µL of supernatant was removed and the pellet was resuspended gently in the remaining supernatant. The suspension was spread on SFM agar plates containing a 1M MgCl2 solution. The plates were incubated at 28 °C for 18-20 h. The next day, an antibiotic solution was prepared in 1 ml dH2O, with 30 µL nalidixic acid (Nal), 15 µL of apramycin, and 30 µl of thiostrepton. 1 mL of this solution was used to spread and cover the agar plate, gently using a sterile spreader. The plates were allowed to dry under the sterile laminar airflow bench and incubated at 28 °C for approximately 2 days until the growth of clones was visible During this step, recombinant *Streptomyces* can be selected since Nal inhibits the growth of *E. coli* cells and the transconjugants harbor an apramycin resistance gene.

The transconjugants were picked with a sterile wooden stick and transferred onto an SFM agar plate containing 15 µL apramycin, 30 µl thiostrepton, and 30 µL Nal. The plate was incubated at 28 °C for 2-3 days. As soon as the growth of colonies was sufficient, the most visible ones were picked and transferred onto SFM agar plates containing 15 µL apramycin and 30 µl thiostrepton until they sporulated. After sufficient sporulation, a spore suspension of the different strains was prepared and stored at -80 °C.

6.5. Fermentation

Most of the production of SM occurs during the growth of *Streptomyces* in a complex fermentation medium. The fermentation was performed in liquid media according to the following protocol:
The first step was the preparation of pre-culture for each recombinant Streptomyces strain and also for their wild type, which was used as a control for all comparative studies. For *S. albus,* 15 mL TSB medium containing thiostrepton and apramycin was inoculated with 50 µL spore suspension. For *S. coelicolor*, 15 mL of YEME medium containing thiostrepton and apramycin was inoculated with 50 µL spore suspension, in a 250 mL sterile baffled flask. The flasks were incubated at 28 °C overnight, shaking at 200 rpm.

The main culture was prepared after enough growth was seen in the pre-culture. For *S.albus*, it was after 72 hours. 25 mL of each of the fermentation media- MYM, SM17, and PM4-1, were inoculated with 2.5 mL of pre-culture in 250 mL sterile baffled flasks. The flasks were incubated for 7 days at 28 °C, shaking at 200 rpm. After 7 days, the respective cultures were transferred into 50 mL falcon tubes and subjected to freeze-drying.

6.6. Extraction and analysis

50 ml of 100 % methanol were added to the freeze-dried culture and the methanol/pellet mixture was allowed to mix for 2 h at room temperature at approximately 100 rpm. The methanolic extracts were centrifuged and the supernatant was transferred to an Eppendorf tube and concentrated using a Speed Vac. The resultant pellet was dissolved in 100 % methanol to adjust the volume up to 500 µL. The methanol extracts were stored at 4 °C and analyzed using HPLC and disc diffusion tests.

6.6.1. Bioassay

Methanolic extracts were tested for antimicrobial activity via disc diffusion. Three microorganisms listed in the table were tested to verify the presence of potential antimicrobial compounds. 30 µl of methanolic extract were applied onto a paper disc (6 mm Whatman filter paper discs) and dried for 20 min at RT. In the meantime, 300 µl cell suspension of different strains was plated on an appropriate agar plate and spread with a sterile spatula. When the plates and the paper disks were dry, the discs were carefully placed on the agar plates. Depending on the ideal growth temperature, the plates were placed into an incubator overnight and the next day the growth was observed.

Table 16: Test organisms used for bioassay

Strain	Type of Microorganism	Temperature	Medium
Pseudomonas putida KT 2440	Gram-negative bacteria	28 °C	LA
Micrococcus luteus DSMZ 1790	Gram-positive bacteria	28 °C	LA
Saccharomyces cerevisiae	Fungi	28 °C	YPD

6.6.2. HPLC

New SMs in the fermented recombinant strain extracts were analyzed using HPLC by comparison of chromatograms with the control strains. A C18 column suitable for liquid chromatography with column dimensions of 250 x 4.6 mm, Luna® from Phenomenex, suitable for particle size 5µm and particle size 100 Å, was used as a stationary phase. An HPLC unit from Shimadzu Corporation was used to carry out this analysis. The detection was carried out using a light-scattering detector (LSD). The mobile phases used were dH_2O with 0.1% Formic acid (Solution A) and acetonitrile (Solution B) was used to equilibrate the column, with a flow rate of 0.5 mL/min. For an appropriate separation, a wavelength of 190 nm to 800 nm was used. The column temperature was maintained at 25 °C.

Table 17: Mobile phase gradient used during analysis

Solvent B %	Time
5-95%	0 to 45 min
95%	45.01 to 55 min
95-5%	55.01 to 65 min

7. Results

7.1. *Streptomyces bambergiensis*

7.1.1. *In silico* Genome analysis

Genome mining facilitates the understanding of genomic sequences and BGCs in bacteria responsible for the production of secondary metabolites. Using an open-source *in silico* bioinformatics tool antiSMASH, BGCs within the genome of *S. bambergiensis* were identified and studied. BGCs that are responsible for the biosynthesis of various polyketides and non-ribosomal peptides, as well as some terpenes, butyrolactones, siderophores, lantipeptides, and other secondary metabolites were identified. antiSMASH predicted that the *S. bambergiensis* carries at least 27 BGCs as listed in Table 1, 24 are located on the chromosome and 3 on the plasmid. This tool gives a detailed overview of analysis about the comparison between BGCs in *S. bambergiensis* and those described in other bacteria with the percentage similarity. Two BGCs, BGC18 and BGC20, which might be responsible for the biosynthesis of new natural products were chosen as targets for this project.

Table 18: In silico analysis of genome sequencing data from S. bambergiensis by antiSMASH 6.0 with detected BGCs and their homologs.

Region	Type of BGC	Size of the BGC	Similar to known BGC	Cluster type	% Similarity
1	T1PKS, NRPS	41 kb	primycin	Polyketide (PK)	5
2	terpene	18 kb	lavendiol	Polyketide	6
3	NRPS	77 kb	streptolydigin	NRP + PK: Modular type 1 + Saccharide (hybrid/tailoring)	7
4	terpene	25 kb	hopene	Terpene	76
5	NRPS, T1PKS	55 kb	himastatin	NRP	12
6	T3PKS, terpene	58 kb	isorenieratene	Terpene	100

7	siderophore	12 kb			
8	terpene	20 kb	geosmin	Terpene	100
9	RiPP-like	11 kb	-	-	-
10	T1PKS, NRPS-like	64 kb	enduracidin	NRP	4
11	terpene	21 kb	albaflavenone	Terpene	100
12	NRPS, T1PKS	48 kb	SGR PTMs	NRP + Polyketide	100
13	siderophore	11 kb	desferrioxamine	Other	66
14	melanin	10 kb	istamycin	Saccharide	4
15	T1PKS, redox cofactor	53 kb	tetronasin	Polyketide	3
16	ectoine	10 kb	ectoine	Other	100
17	T2PKS	72 kb	spore pigment	Polyketide	83
18	NRPS	70 kb	telomycin	NRP	14
19	Lanthipeptide Class I	26 kb	meridamycin	Lanthipeptide	40
20	T1PKS, NRPS, beta lactone	55 kb	chalcomycin A	Polyketide	4
21	Lassopeptide	22 kb	-	-	-
22	NRPS	42 kb	stenothricin	NRP: cyclic depsipeptide	13

23	Phosphoglycolipid,nucleoside	35 kb	telcomycin	Other	66
24	NRPS	41 kb	rhizomide A/B/C	NRP	100
25	PKS NRPS/PKS	216 kb 98 kb	stambomycin unknown	T1PKS Hybrid cluster	64 -
26	NRPS/trans-AT PKS	11 kb	griseoviridin / viridogrisein	butyrolactone	5
27	NRPS	68 kb	unknown	peptide	-

7.1.2. BGC18 and BGC20

BGC18 has a size of 70 kb and is an NRPS type of cluster, with four genes that most probably contribute to the core molecule biosynthesis and comprise 7 NRPS modules. An illustration from antiSMASH depicting gene organization in BGC18 is shown in figure 16. Among other genes, few surrounding the core biosynthetic genes were predicted to encode transposase proteins and one (*orf* 42), indicated in blue, was predicted to encode an ABC transporter ATP-binding protein. The predicted functions of the genes and their homologs are listed in Tables 23 and 24 in Appendix I.

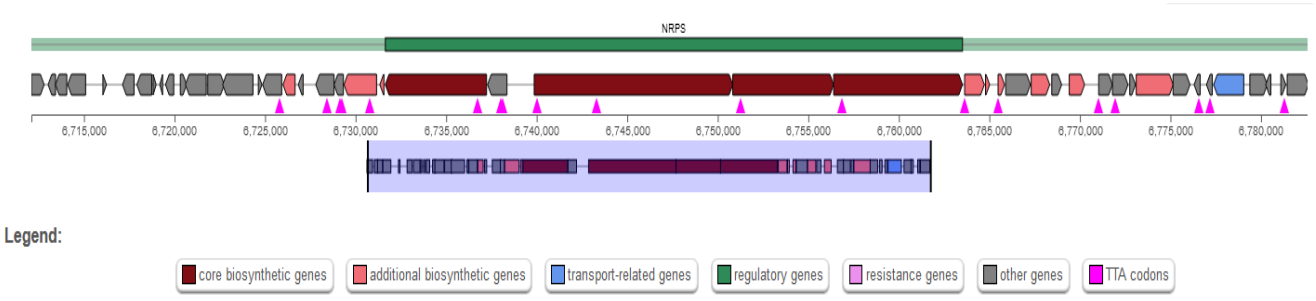


Figure 16: Biosynthetic gene cluster 18

BGC20 spans 55 kb and is a hybrid cluster with PKS, NRPS, and beta lactone biosynthetic genes. The core biosynthetic region has one amino acid adenylation domain as part of the NRPS/PKS modules for biosynthesis (*orf* 20), as shown in figure 17. The other one is a gene for pyruvate carboxylase (*orf* 29). Additionally, it has three regulatory genes, two of which belong to the *tetR* family of transcriptional repressors. The other one is a regulatory protein from the OmpR family,

which is a DNA-binding response regulator, with REC and winged-helix-turn-helix (wHTH) domain. The other regulatory gene, PucR C-terminal HTH domain (*orf* 31*)*, was not detected by antiSMASH and was identified manually using BLAST. *pucR* and *ompR* regulatory genes can be considered for overexpression to enhance SM biosynthesis.

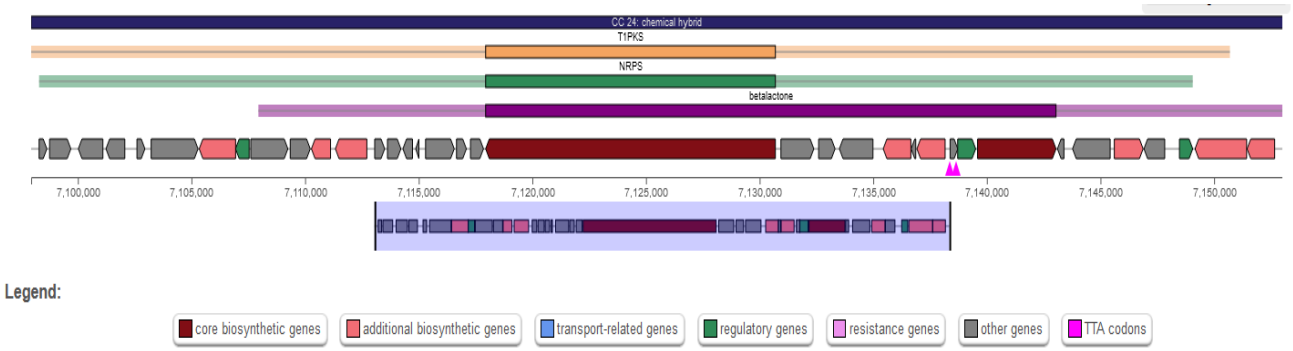


Figure 167: Biosynthetic gene cluster 20

7.1.3. Genome library screening

For this project, the genome library of *Streptomyces bambergiensis*, consisting of 1854 *E. coli* clones harbouring the entire genome created by Dr. Olha Schneider, was used. PCR-based screening of the library was performed to identify clones containing BGC 18 and 20. Primers were designed specifically to amplify desired genes and were tested for their functionality and specificity. The gDNA of *Streptomyces bambergiensis* was used as a control DNA template for optimizing PCR conditions. Taq DNA polymerase was used for these experiments.

Table 19: Expected sizes of PCR amplified product

Primer sets	Size
BGC 18_P1	705 bp
BGC 18_P2	475 bp
BGC 18_P3	917 bp
BGC 20_P1	759 bp
BGC 20_P2	763 bp
BGC 20_P3	790 bp

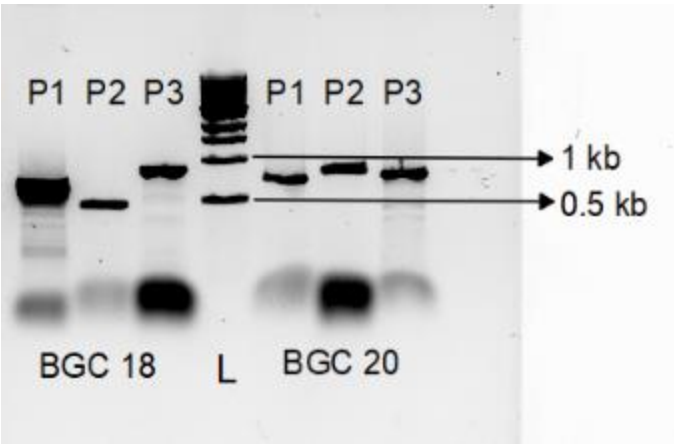


Figure 18: Agarose gel to verify the functionality of primer sets for BGC18 and 20. *Expected sizes are listed in table 19.*

Replica plating technique was used to plate clones from a 96-well plate onto an agar plate with a suitable antibiotic as a selection marker. Every 96-well plate was divided into two parts, "part a" and "part b", consisting of 48 *E. coli* clones each. In the first PCR run, this pool of 48 clones ("part a") of every 96-well-plate was screened to detect parts of the respective BGC. Subsequently, the same procedure was performed for "part b" of all 12 96-well-plates.

For BGC 18, plates 14a, 16b, 9b, 19b, and 15b, gave positive signals during pooled-PCR screening. To narrow down the screening process, a pool consisting of 48 different *E. coli* clones was divided into smaller pools containing only 6 different *E. coli* clones each. Following this, individual clones were identified. In the case of BGC 20, the entire cluster was harboured by three *E. coli* clones on plates 8b, 10b, and 16a. DNA sequencing was performed using a service from Eurofins Genomics. These results confirmed the clones containing genes for BGC 18 and 20. The table below summarizes the results. Once all clones harbouring parts of the BGCs were detected, the respective fosmid DNA (fDNA) was isolated according to the protocol discussed in chapter 5.

Table 20: An overview of identified BGC fosmids, primer set, plate, and clone number, which gave a positive PCR signal.

BGC	Fosmid	Primer set	Plate and clone
18	pFOS1_C18	1	14_E5
	pFOS2_C18	1	16_A6
	pFOS3_C18	2	9_F12
	pFOS4_C18	2	19_H11
	pFOS5_C18	3	15_G8

20	pFOS1_C20	1	8_F7
	pFOS2_C20	2b	10_D8
	pFOS3_C20	3	16_G5

7.1.4. Preparation of BGC fragments for assembly in yeast

After analyzing the sequencing results, overlapping BGC fragments were found. Using Clone Manager, the location of the specific sites for restriction enzymes that can be used to cut the fosmid was identified.

The PCR-based library screening showed that the genes of BGC 18 were harboured by five *E. coli* clones, while BGC 20 was divided into three *E. coli* clones. After the isolation of the fDNA from every single clone, the DNA was digested with respective restriction enzymes as listed in the table below. The resulting fragments were used for the yeast-based DNA assembly. The expected sizes of DNA fragments used for assembly are marked in bold.

Table 21: Restriction enzymes used for digestion of fDNA with expected product sizes. Fragments, which were used for assembly reaction are marked in bold.

Fosmid DNA	Restriction enzyme	Sizes (bp)
BGC 18		
pFOS1_C18	*Mre* I	330,651,75,2120,1875,633,8274,6743,**11039**,6015
pFOS2_C18	*EcoR* V-HF	2613,1835,**11876**,13169,229,5674
pFOS3_C18	*Psi* I	196,8651,**36595**
pFOS4_C18	*Nsi* I	11330,**25788**
pFOS5_C18	*Dra* I and *EcoR* V	3345,276,229,1215,1611,339,696,**16467**,20186

BGC 20		
pFOS1_C20	*Sbf* II, *Acl* I	1574,2892,14265,**27880**
pFOS2_C20	*Psp* XI	10636,**31747**
pFOS3_C20	*Mfe* I	9256,**41469**

Agarose gel electrophoresis of all the digested fosmids mentioned above was performed, to separate the fragments, as seen in the figure below. (See Appendix I for images with well-separated bands). BGC fragments were recovered from the agarose gel using a Qiagen DNA extraction kit.

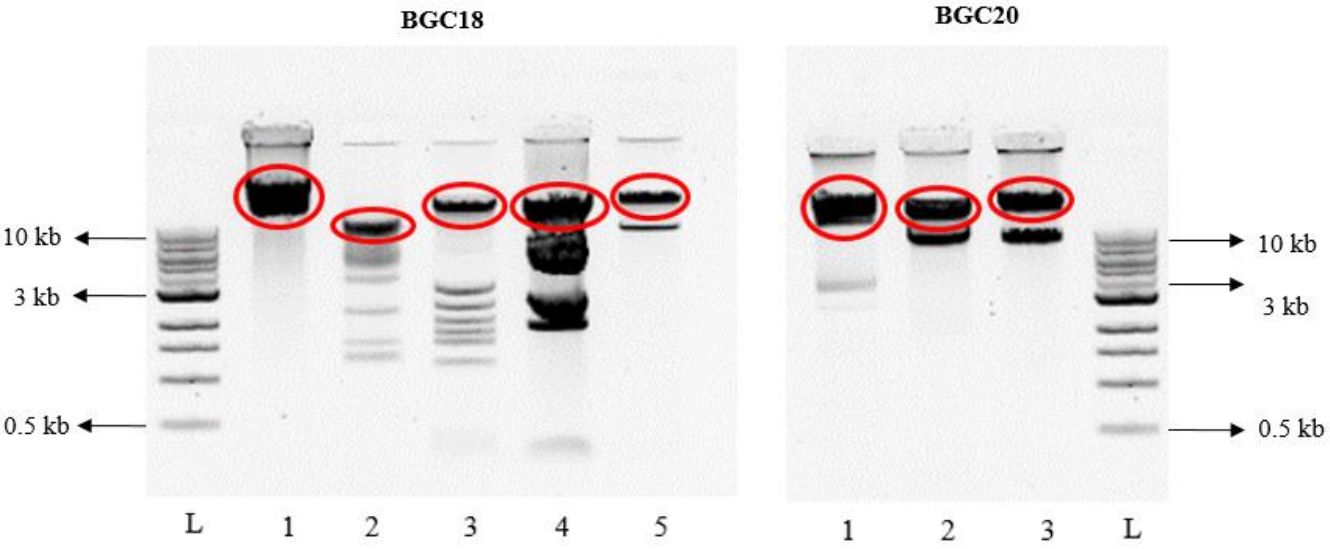


Figure 19: DNA fragments by restriction digestion, for yeast-based DNA assembly. *In 19a 1: pFOS1_C18, 2: pFOS2_C18, 3: pFOS3_C18, 4: pFOS4_C18 and 5: pFOS5_C18 for BGC18. In 19b 1: pFOS1_C20, 2: pFOS2_C20, 3: pFOS3_C20 for BGC 20 and 1 kb ladder (L) was used as a marker; expected sizes are listed in Table 21.*

7.1.5. Preparation of the pCLY10-based capture vectors

To enable the successful assembly of BGC fragments in yeast, the pCLY10 vector was modified by cloning flanks or the capture arms specific for the BGC flanks. These DNA fragments enable homologous recombination between the DNA fragments representing BGC and the cut, linearized vector, specific to respective BGCs. The shuttle vector pCLY10 can be maintained in *Saccharomyces*, *E. coli*, and *Streptomyces* strains, since it carries selection marker genes for *Saccharomyces (leu2)*, which provides the selection against leucine auxotrophy and the resistance genes for antibiotic apramycin (acc), which is active against both, Gram-positive and Gram-

negative bacteria (Figure 5). pCLY10 is replicative in *S. cerevisiae* and *E. coli* strains and can be transferred into the *Streptomyces* strains via conjugative transformation, where it integrates into the chromosome.

As shown in figure 5, the vector pCLY10 contains seven different elements:
1. A yeast centromere/replication origin- ***CEN6/ARS4***; guarantees replication in yeast
2. A yeast selectable marker, namely ***LEU2***, to select the clones after the transformation into yeast.
3. For the replication of the vector in *E. coli*, the origin of the replication gene ***ori15A*** is necessary.
4. for the selection in *E. coli* and *Streptomyces*, ***aac(3)IV***, an apramycin resistance gene.
5. ***oriT***, the origin of transfer, enables the conjugation of the plasmid into a recipient cell.
6. ***attP-site*** and ***VWBint***, encoding for the phage integrase, enable the site-specific integration of the vector into attP- sites in the genomes of *Streptomyces*

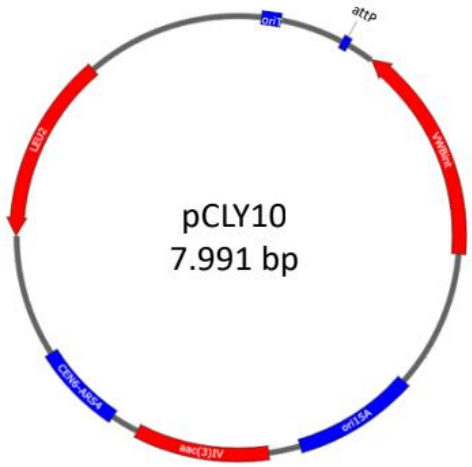


Figure 170: pCLY10 vector map. *The vector consists of the following seven elements: oriT, attP, VWBint, ori15A, aac(3)IV, CEN6-ARS4, and LEU2. The shuttle vector pCLY10 has a size of 7.991 bp*

7.1.5.1. Modification of pCLY10 for BGC18

For assembly of BGC18 in *Saccharomyces cerevisiae*, pCLY10 was modified to have BGC-specific capture arms- pCLY10_LR_C18. The capture arms, i.e., the flanking sequences with the sizes 531 bp and 649 bp, were amplified using PCR. The PCR products (LRI_C18 + LRII_C18) were separated by gel electrophoresis, isolated from the agarose gel, purified, digested with appropriate restriction endonucleases and ligated into pCLY10 digested with *Apa* I and *Hind* III. To verify the construct, analytical digestion of pCLY10_LR_C18 with *Apa* I and *Hind* III was performed.

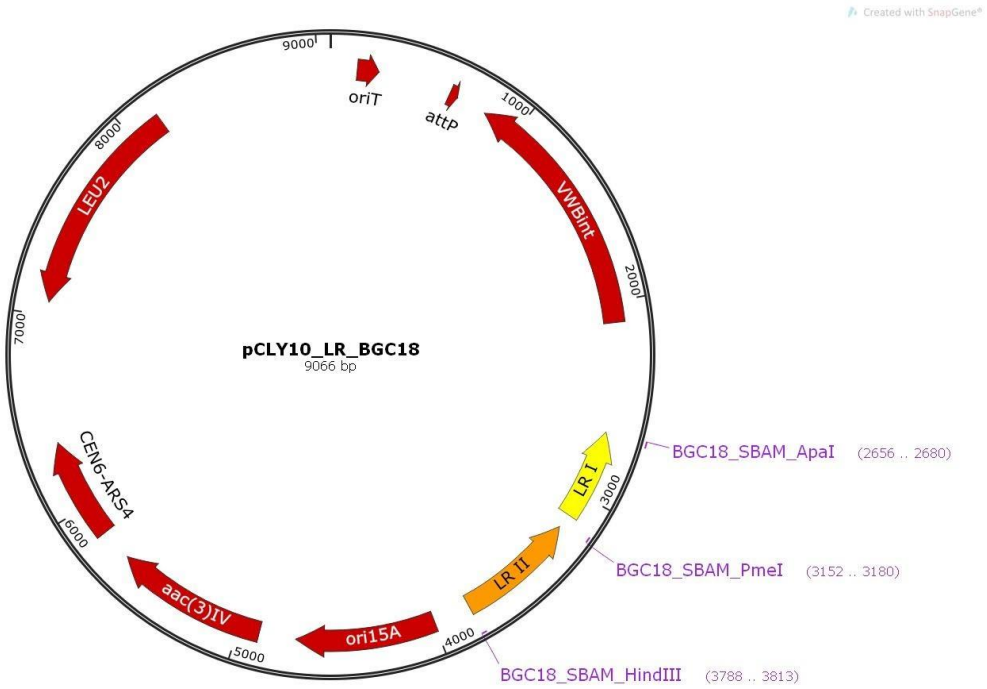


Figure 21: pCLY10_LR_C18 with flanking sequences. *The yellow flanking sequence is LR I of size 531 bp and orange flanking sequences LR II with size 649 bp.*

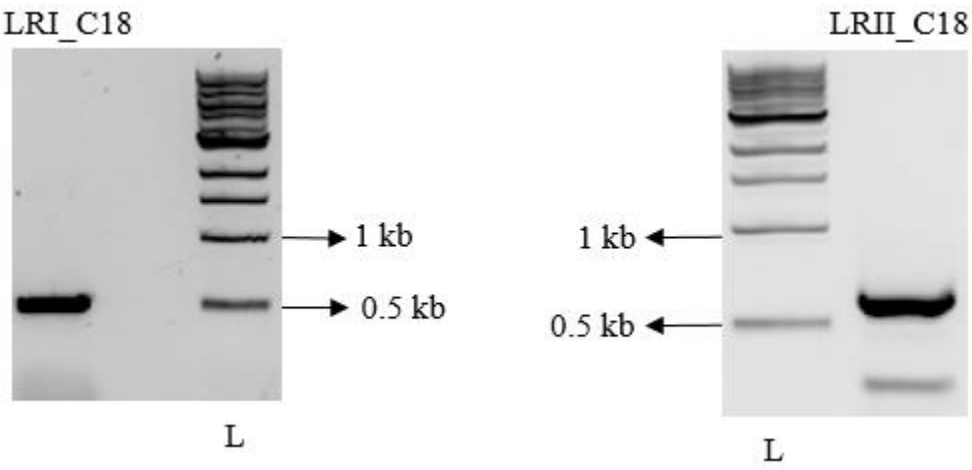


Figure 22: Agarose gel with PCR amplified product. *LRI with size 531 bp and LRII with size 649 bp for BGC18 amplified using PCR*

Later, the modified vector was digested with *Pme* I. *Pme* I has only one cutting site in the pCLY10 vector, thus the vector becomes linear after this restriction digestion. This enables homologous recombination between the capture arms of the vector and the homologous ends of the BGC DNA fragments.

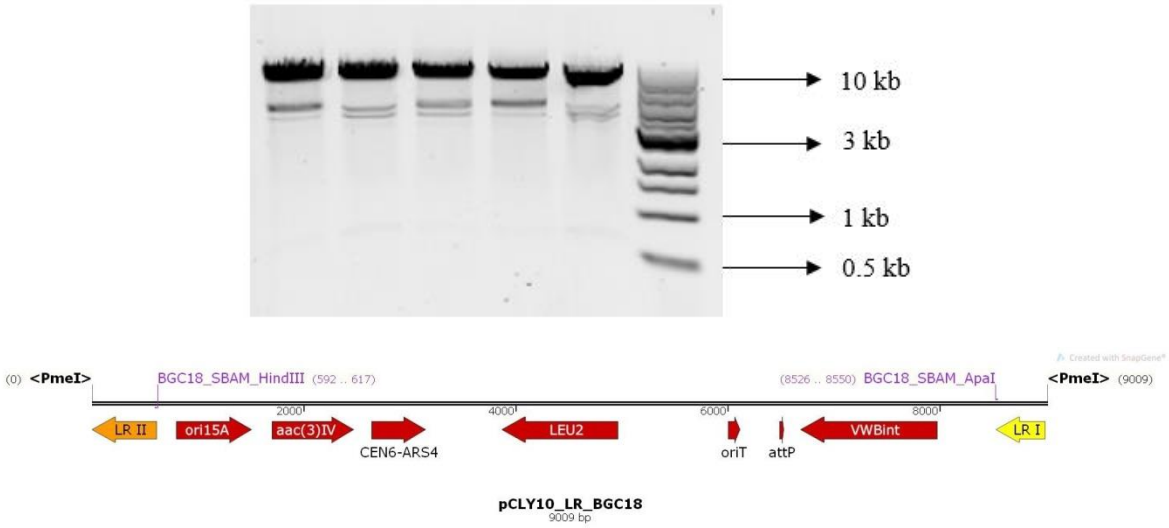


Figure 23: Pme I linearized pCLY10_LR_BGC18 vector

7.1.5.2. Modification of pCLY10 for BGC20

The flanking sequences for BGC 20 are of size 631 bp and 689 bp. These were amplified using PCR, and the PCR products (LRI_C20 + LRII_C20) were separated by gel electrophoresis, isolated from the agarose gel, purified, digested with restriction endonucleases and ligated into pCLY10 digested with *Apa* I and *Hind* III. To verify the construct, an analytical digestion of pCLY10_LR_C20 with *Apa* I and *Hind* III was performed.

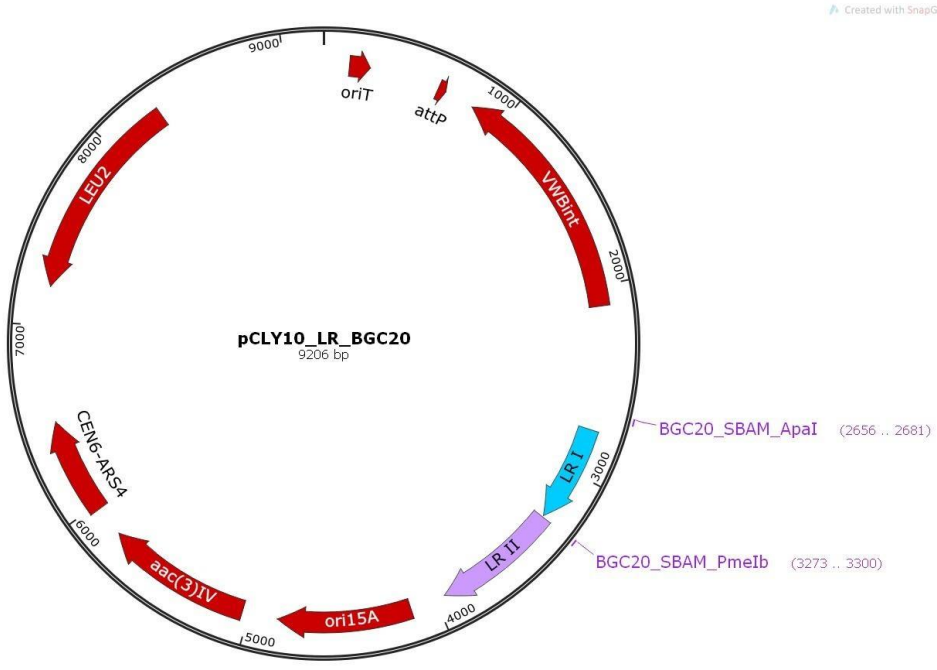


***Figure 24: pCLY10_LR_BGC20 with flanking sequences**. The blue flanking sequence is LR I of size 631 bp and violet flanking sequences LR II with size 681 bp.*

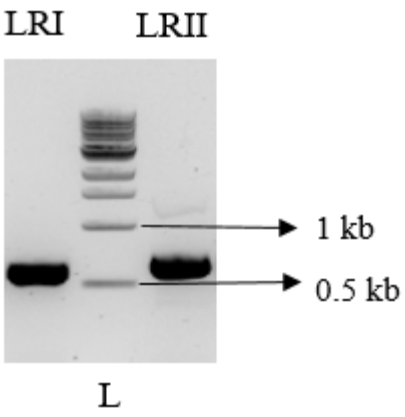
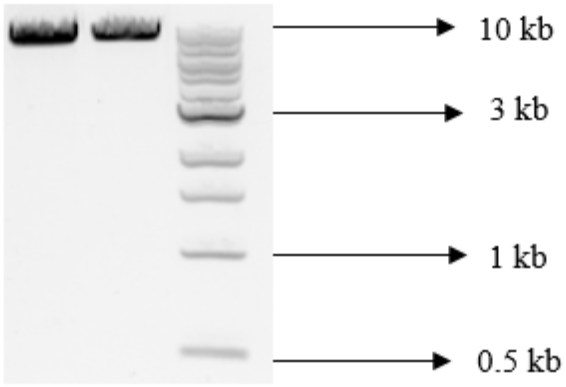


Figure 25: Agarose gel with PCR amplified product*. LRI with size 631 bp and LRII with size 689 bp for BGC20 amplified using PCR*

The modified vector was then digested with *Pme* I, making it linear after this restriction digestion. This enables homologous recombination between the capture arms of the vector and the homologous ends of the BGC DNA fragments.



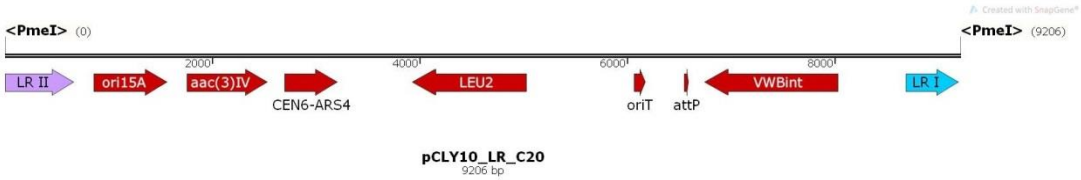


Figure 26: Pme I linearized pCLY10_LR_BGC20 vector

7.1.6. Yeast-based DNA assembly

The complete BGC must be assembled from the fragments led before an attempt on heterologous expression. In vivo cloning technique in yeast was used to assemble large BGCs. Linearized pCLY10-based capture vectors and fragments of fosmid DNAs covering BGCs were introduced into *Saccharomyces cerevisiae* cells to generate an assembled, circular plasmid DNA via homologous recombination.

Colony PCR-based screening was performed with all the above-mentioned primer sets and already optimized PCR conditions, to verify the presence on pSV_C18 and pSV_C20 plasmids that should carry complete BGC 18 and 20, respectively. The *Saccharomyces cerevisiae* BY4742 clones that generated the PCR product can be considered to have the entire BGC cloned into the vector.

Subsequently, the pSV_C18 and pSV_C20 plasmids were isolated from *S. cerevisiae* and transformed into competent *E. coli* EPI300 via electroporation to increase plasmid copy number and allow isolation of pDNA sufficient for analysis and further manipulation.

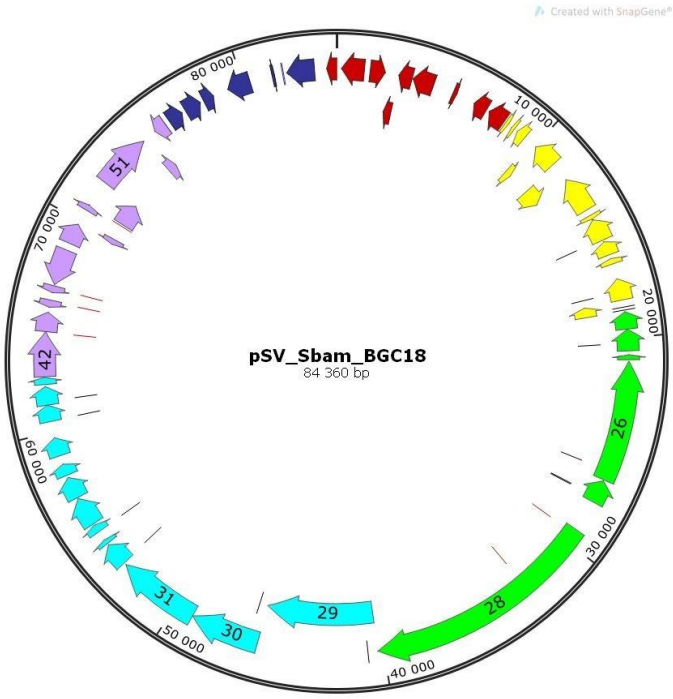


Figure 27: Expected plasmid pSV_C18 of 84.36 kb generated after carrying out assembly in yeast. *BGC fragments are illustrated as follows: pFOS1 in red, pFOS2 in yellow, pFOS3 in green, pFOS4 in cyan, pFOS5 in violet, and the pCLY10 vector genes in blue.*

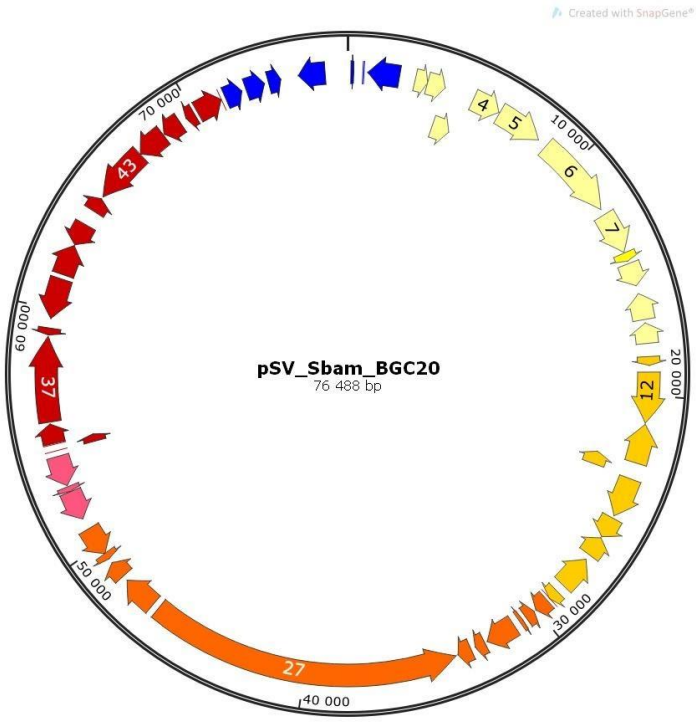


Figure 28: Expected plasmid pSV_C20 of 76.486 kb generated after carrying out assembly in yeast*. BGC fragments are illustrated as follows: pFOS1 in light yellow, the overlap between pFOS1 and pFOS2 in dark yellow, pFOS2 in orange, the overlap between pFOS2 and pFOS3 in pink, pFOS3 in red, and pCLY10 vector genes in blue.*

A PCR-based screening of all yeast clones bearing assembled pSV plasmids gave positive results when tested with primer sets specific for both BGCs. However, the pSV plasmids proved unstable when introduced into E. coli EPI300 cells. Thus, concurrent steps of transformation into conjugative E. coli strain ET12567 and heterologous expression were not possible due to this instability.

7.2. *Actinoalloteichus fjordicus*

7.2.1. Activation of BGC 3 and 4 in *Actinoalloteichus fjordicus* DSM 46856

For this part of the project, a rare actinobacterial strain *Actinoalloteichus fjordicus* DSM 46856, which was isolated from a marine sponge Geodia barretti collected at the Trondheim fjord, Norway was used to study and synthesize putative SMs. *In-silico* analysis of genome sequencing data showed that *A. fjordicus* DSM 46856 carries 24 BGCs. Two BGCs, i.e., BGC3 and BGC4, which might potentially specify biosynthesis of ladderane and an NRPS compound respectively, were cloned into vectors after successful homologous recombination in yeast, by MSc. Stefanie Schalko.

These vectors were used for expression of BGCs and evaluation of SMs it might produce after heterologous expression. Since no functional expression of the clusters was demonstrated during

the previous work, overexpression of SARP regulators for activation of BGCs was considered during this project.

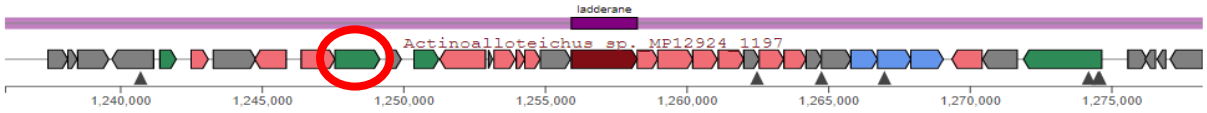


Figure 29: Biosynthetic gene cluster 3; *highlighted in red is the SARP regulatory gene targeted for overexpression.*

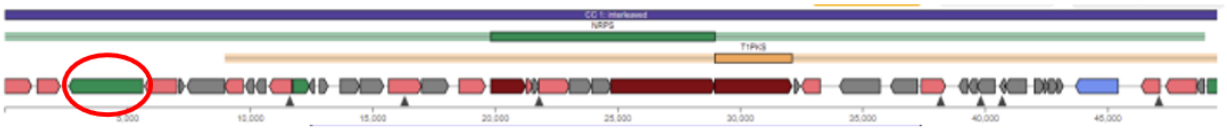


Figure 30: Biosynthetic gene cluster 4; *highlighted in red is the SARP regulatory gene targeted for overexpression.*

SARP regulators from BGC 3 and 4 were cloned into the multi-copy vector pUWLoriT and transformed into *E. coli* DH5α cells. These cloning experiments were carried out by Miriam Kuzman.

The plasmids were extracted and analytical restriction digestion was carried out with *Hind* III and *Age* I, shown in figure 29, using three plasmids isolated from *E. coli* clones that contain pSS3_SARP and pSS4_SARP. All the clones showed three bands after agarose gel electrophoresis. Clones 1 and 3 were selected for further experiments. Both the vectors were transformed into *E. coli* ET12567 to subsequently perform a conjugation into *Streptomyces coelicolor* M1154 and *Streptomyces albus* J1074 strains harbouring BGCs 3 and 4.

1 L 2

Figure 31: Analytical digestion of pDNA*; 1: pSS4_SARP (Age I/Hind III 2555,3319,4654), 2: pSS3_SARP (Age I/Hind III, 1116,3319,4636); L is the 1kb DNA ladder*

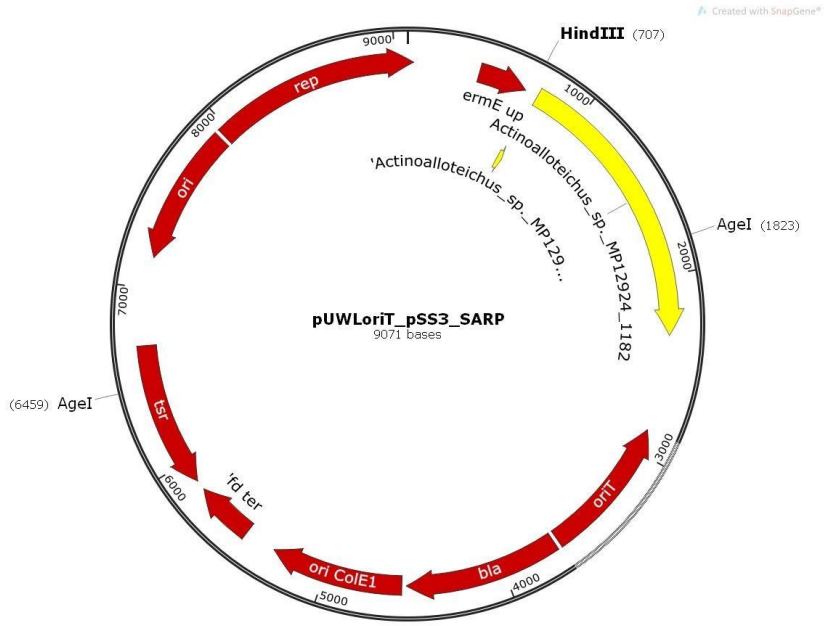


Figure 32: puWLoriT vector with Actinoalloteichus fjordicus BGC3 SARP*; gene in yellow indicate the SARP genes*

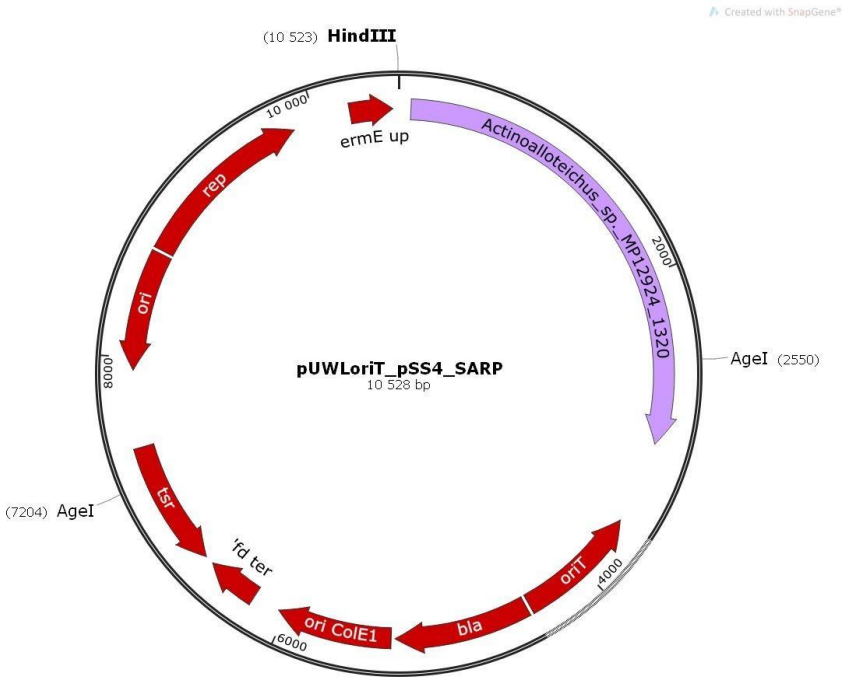


Figure 33: puWLoriT vector with Actinoalloteichus fjordicus BGC4 SARP*; gene in violet indicate the SARP genes*

7.2.2. Analysis of the extracts

Streptomyces albus J1074 and *Streptomyces coelicolor* M1154 harbouring BGCs 3 and 4 along with the plasmids for overexpression of respective SARP genes were cultivated in different fermentation media. The extracts were analyzed by HPLC and tested for bioactivity using a disc

diffusion assay. The strains with unmodified pCLY10 vectors with empty pUWLoriT and strains with BGC were used as controls.

7.2.2.1. Bioassay

Methanolic extracts after fermentation of recombinant *S. albus* were tested for bioactivity using the disc diffusion assay with three test organisms as mentioned in the methods section. Overall, 46 samples as listed in the tables below, 15 samples for each medium i.e., MYM, SM17, and PM4-1 along with 100% methanol as blank, were analyzed.

Table 22: Results from bioassay; '+' indicates a zone of inhibition was observed against the test organism and '-' indicates no bioactivity against the test organism; S.a. for Streptomyces albus J1074; S.c. for Streptomyces coelicolor M1154.

Sample Number	**Sample**	**Bioactivity against *Micrococcus luteus***	**Bioactivity against *Saccharomyces cerevisiae***
SM17 Medium			
1	S.a.pSS3+oriT_a	+	+
2	S.a.pSS3+oriT_b	+	+
3	S.a.pCLY10+oriT_a	+	+
4	S.a.pCLY10+oriT_b	+	+
5	S.a.pCLY10+SARP4_a	+	-
6	S.a.pCLY10+SARP4_b	+	+
7	S.a.pSS4+SARP4_a	+	+
8	S.a.pSS4+SARP4_b	+	+
9	S.a.pSS4+oriT_a	+	+
10	S.a.pSS4+oriT_b	+	-
11	S.a.pCLY10+SARP3_a	+	-
12	S.a.pCLY10+SARP3_b	+	-
13	S.a.pSS3+SARP3_a	+	+

14	S.a.pSS3+SARP3_b	+	+
15	Control Medium	+	-
PM4-1 Medium			
16	S.a.pSS3+oriT_a	+	+
17	S.a.pSS3+oriT_b	+	-
18	S.a.pCLY10+oriT_a	+	-
19	S.a.pCLY10+oriT_b	+	-
20	S.a.pCLY10+SARP4_a	+	+
21	S.a.pCLY10+SARP4_b	+	-
22	S.a.pSS4+SARP4_a	+	+
23	S.a.pSS4+SARP4_b	+	+
24	S.a.pSS4+oriT_a	+	-
25	S.a.pSS4+oriT_b	+	-
26	S.a.pCLY10+SARP3_a	+	-
27	S.a.pCLY10+SARP3_b	+	+
28	S.a.pSS3+SARP3_a	+	-
29	S.a.pSS3+SARP3_b	+	+
30	Control Medium	+	-
MYM Medium			
31	S.a.pSS3+oriT_a	+	+
32	S.a.pSS3+oriT_b	+	+
33	S.a.pCLY10+oriT_a	+	+

34	S.a.pCLY10+oriT_b	+	+
35	S.a.pCLY10+SARP4_a	+	+
36	S.a.pCLY10+SARP4_b	+	+
37	S.a.pSS4+SARP4_a	+	+
38	S.a.pSS4+SARP4_b	+	+
39	S.a.pSS4+oriT_a	+	+
40	S.a.pSS4+oriT_b	+	+
41	S.a.pCLY10+SARP3_a	+	+
42	S.a.pCLY10+SARP3_b	+	-
43	S.a.pSS3+SARP3_a	+	-
44	S.a.pSS3+SARP3_b	+	+
45	Control Medium	+	-
46	100% Methanol	-	-

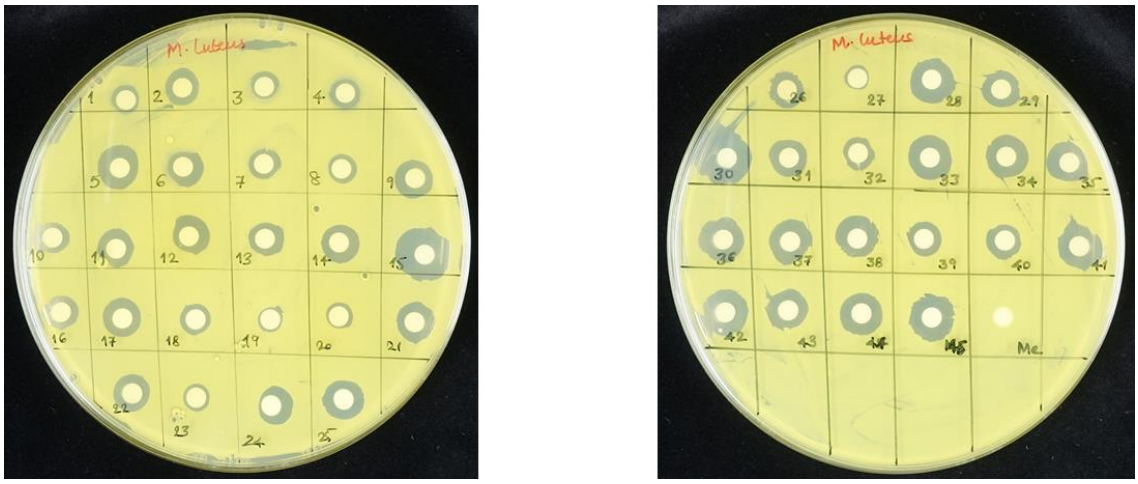


Figure 34: Bioassay with Micrococcus luteus

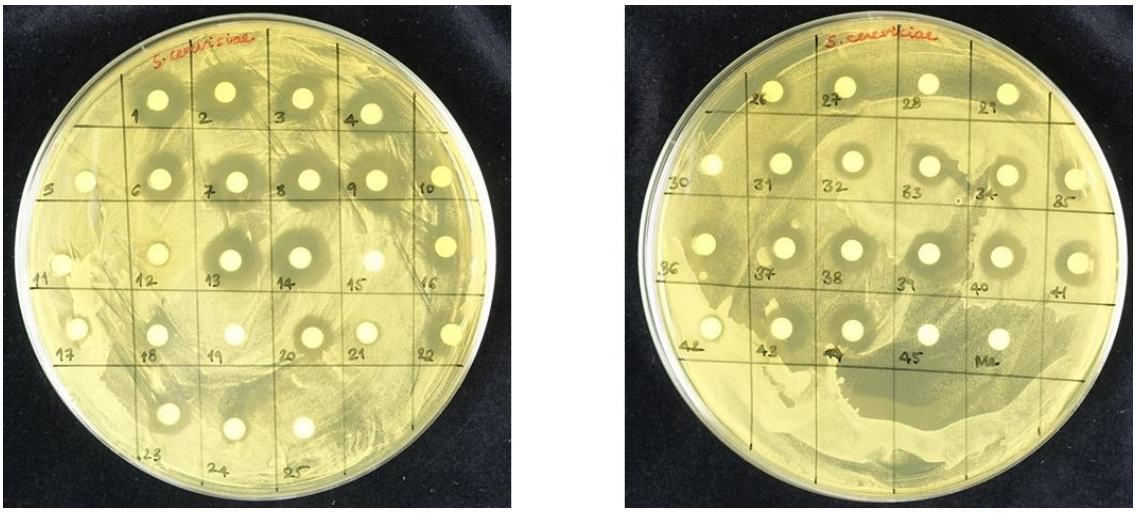


Figure 35: Bioassay with Saccharomyces cerevisiae

The results show a zone of inhibition for some of the extracts that were tested for activity against the aforementioned test organisms. However, it is evident that a zone of inhibition is seen for control samples i.e., vectors without SARP, *i.e.*, samples 1,2,3,4,9 and 10 in SM17 medium, 16,17,18,19,24 and 25 in PM4-1 medium and 31,32,33,34,39 and 40 in MYM medium.

7.2.2.2. High-Performance Liquid Chromatography (HPLC)

Metabolite profiles of the methanolic extracts were obtained using HPLC. Additional peaks were detected in chromatograms of extracts with overexpressed SARP compared to the control extracts. Figure 36 shows the chromatograms of *S.albus* recombinant strains which were fermented in SM17 medium, the culture was freeze-dried and extracted with 100 % methanol.

The extracts contain pSS3 with SARP in violet, pSS4 with SARP in green, and control strains without SARPs pCLY10+puWLoriT in blue and pCLY10 vectors with SARP 4 in red. In figure 37, chromatograms of *S.albus* recombinant strains fermented in PM4-1 medium can be examined, where pSS4 with SARP in green and control strains without SARPs pCLY10+puWLoriT in blue and pCLY10 vectors with SARP 4 is illustrated in red colour.

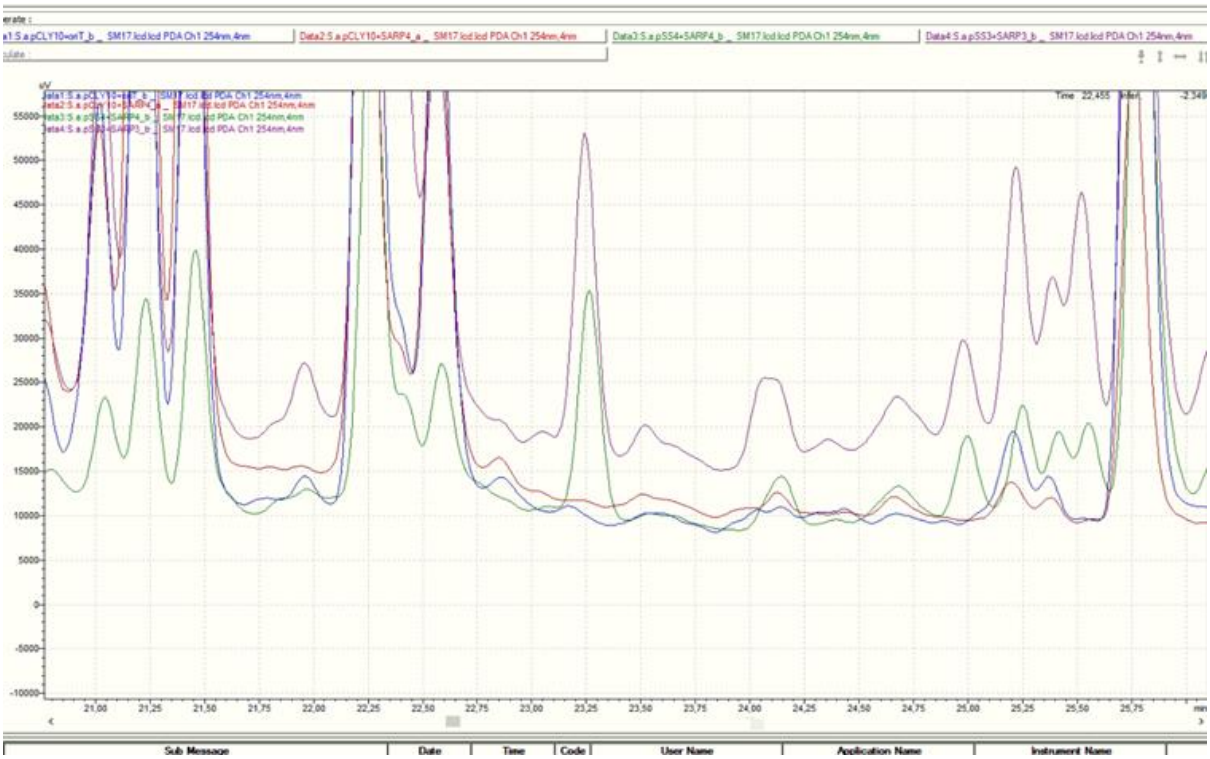


Figure 36: Chromatogram of the extracts from recombinant Streptomyces albus J1074 harboring BGC3 and BGC4 with overexpressed SARP in SM17 media, at wavelength 254 nm; *violet lines indicate pSS3 with SARP, green lines indicate pSS4 with SARP and control strains without SARPs pCLY10+puWLoriT are in blue and pCLY10 vectors with SARP 4 in red. Y-axis is the adsorption intensity in aU, the x-axis is the retention time in min.*

As it can be seen from the chromatogram in figure 36, additional peaks around 23.25 min and 25 min retention time can be seen in samples with pSS3 with SARP and pSS4 with SARP, respectively. These peaks are not seen in the control samples at given retention times. These experiments were carried out in duplicates, with 2 different mutants; the chromatogram below illustrates one set of samples as an example.

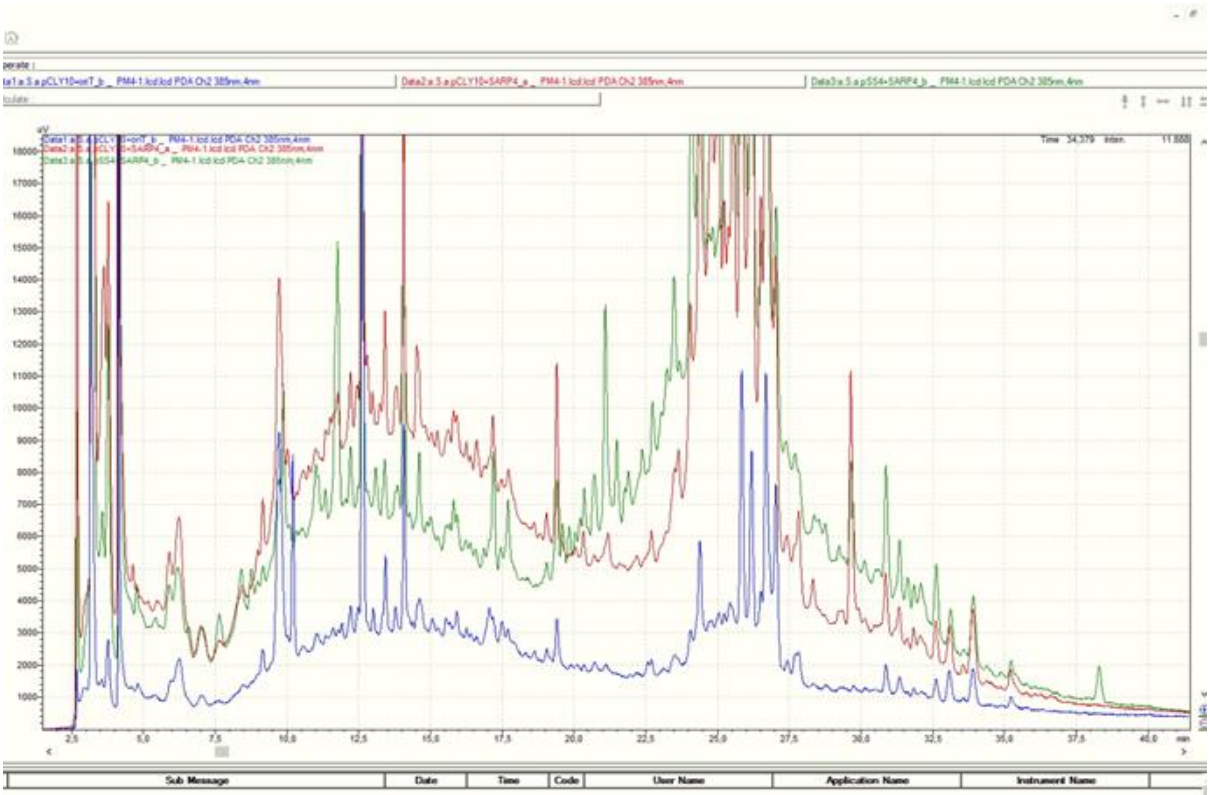


Figure 37: Chromatogram of the extracts from recombinant Streptomyces albus J1074 . harbouring BGC3 and BGC4 with overexpressed SARP in PM4-1 media, at wavelength 385 nm. *The green lines indicate pSS4 with SARP, control strains without SARPs, pCLY10+puWLoriT in blue and pCLY10 vectors with SARP 4 are illustrated in red. Y-axis is the adsorption intensity in aU, the x-axis is the retention time in min.*

In figure 37, three additional peaks in samples with clusters containing SARP 4 at retention times 11-, 21- and 38-min can be seen. Samples without SARP do not show these peaks at the same retention times. These additional peaks can be attributed to the presence of new compounds present in the samples originating from recombinant strains overexpressing SARP.

8. Discussion

Streptomyces bambergiensis is a bacterium species of the genus of *Streptomyces* that has been isolated from soil. *Streptomyces bambergiensis* produces a phosphoglycolipid antibiotic called moenomycin [66–68]. According to the sequencing data, the genome from this organism consists of a chromosome and a plasmid. Altogether 27 BGCs could be detected *in-silico*.

This project aimed to discover new NPs from two unique BGCs i.e., BGC 18 and 20, from *Streptomyces bambergiensis,* and heterologously express them in other *Streptomyces* host strains. The *in-silico* tool, antiSMASH predicted that BGC 18 encodes an NRPS compound and has only 14% similarity with already known cluster encoding telomycin, which is a cyclic depsipeptide antibiotic active against Gram-positive bacteria. [69] BGC 20 which is a hybrid cluster, with betalactone, T1PKS and NRPS domains.

BGC 18 has TTA codons and research has suggested that these codons have a role in developmental regulation. [70] Particularly, in all sequenced streptomycetes genomes to date, leucyl codon TTA is the rarest one. $tRNA_{LeuUAA}$ of streptomycetes, encoded by the *bldA* gene, usually appears in mature form only after the onset of morphological differentiation and activation of secondary metabolism. [71] Additionally, the transposases encoding genes surrounding the BGC 18. These enzymes are responsible for movement of transposons, the mobile genetic elements. These genes can be hypothesized to have a role in horizontal gene transfer across species. [72] A gene encoding a transporter from the ABC family is also present in the BGC 18 (*orf* 42). The ATP-binding cassette transporter is responsible for pumping out SMs across cellular membranes. [73]

Homologous recombination to assemble the respective BGC fragments was carried out in yeast. PCR-based screening of all the yeast clones containing assembled pSV plasmids gave positive signals when tested with respective primer sets specific for both the BGCs. However, it was observed that the pSV plasmids were probably unstable when introduced into EPI300 *E. coli* cells [63]. These cells can carry large inserts and be transformed with large plasmids - up to at least 145 kb plasmid DNA. Therefore, BAC-optimized *E. coli* [62] electrocompetent cells were used to help in the production of multi-copy plasmids. These cells allow cloning of large-insert DNA, BAC, fosmid, or plasmid clones. BAC vectors can maintain inserts of up to 350 kb, utilizing the single copy origin of replication from the *E. coli* F plasmid. In addition, these vectors are extremely useful for cloning smaller unstable and recalcitrant DNAs (10-20 kb), which appear to be more stable at lower copy numbers. [62] Unfortunately, this did not give positive results. It can therefore be concluded that the possible instability of pSV plasmids in *E. coli* cells may be due to the size of the BGCs that are incorporated in the pCLY10 vector. This phenomenon was previously observed for several other BGCs (unpublished data, according to internal communication).

Several self-replicating plasmids can be used to carry the inserts with large-size DNA fragments. pStreptoBAC V and pSBAC are examples of BAC-based shuttle vectors, which can take up big BGCs and be used for the expression of clusters encoding for tautomycetin (pSBAC) and daptomycin (pStreptoBAC V). Secondly, other shuttle vectors like *E. coli-Streptomyces* artificial chromosomes (ESACS) can be considered. [47, 50]

The pSV plasmids have a low copy number in the yeast cells. Therefore, they were introduced in EPI 300 *E. coli* or BAC optimized *E. coli* cells, where they can be maintained as multi-copy vectors. Later, to enable conjugation, these plasmids are transformed into *E. coli* ET12567 cells. The problem observed during this project was the instability of pSV plasmids when introduced into *E. coli* cells.

Since many BGCs are silent and not expressed under laboratory conditions, several strategies have been used to activate them and increase the possibility of the production of the target compound. An approach to enhance SM production by overexpressing the regulatory genes in BGC20 was proposed and initial cloning experiments for the same were carried out. Another possibility could be the inactivation of *tetR* regulators in BGC20. These are known to act as repressors for SMs production in *Streptomyces*. Research has shown the successful production of heterologously expressed compounds in *Streptomyces* strains after targeting the regulatory genes within the BGCs. [39–41]

The second part of this master thesis work involved experiments with overexpression of SARP regulators in BGC3 and BGC4 of *Actinoalloteichus fjordicus.* SARP-overexpression plasmids were conjugated in *Streptomyces albus* and *Streptomyces coelicolor* strains harbouring BGCs 3 and 4. The recombinant strains were cultivated in different media to study and analyze optimum growth and production of SMs. The methanolic extracts were analyzed for the presence of putative antimicrobial compounds using HPLC, MS, and disc- diffusion bioassay.

The zone of inhibition that can be seen for numerous samples in the above pictures can be attributed to the antibiotic thiostrepton that was added to the growth media for stable maintenance of the pUWLoriT-based vectors. Therefore, the observed bioactivity against Gram-positive bacteria seen in the extracts with SARP regulators most likely can be attributed to thiostrepton. However, activity against *Saccaromyces cerevisiae* cannot be so easily explained, since thiostrepton is not active against yeast. It is possible that such activity is due to candicidin, a polyene macrolide whose BGC is present in *S. albus* J1074 but usually not expressed. The use of other test organisms like *Escherichia coli* DH5F, *Erwinia persicina* DSMZ 19328, *Bacillus subtilis* DSMZ 10, *Kocuria rhizophila* DSMZ 348, *Enterococcus mundtii* DSMZ 4840, *Staphylococcus carnosus* DSMZ 20501, could be tried to further evaluate bioactivity.

The HPLC analysis of the extracts show some significant additional peaks in samples containing extracts from recombinant strains with overexpressed SARP. These additional peaks can be attributed to the presence of new compounds present in these samples. However, further analysis using Mass spectrometry, needs to be performed to confirm the same.

9. Conclusion and Outlook

In this project, homologous recombination in yeast and subsequent integration of *S. bambergiensis* BGC 18 and 20 in the pCLY10 vector was successful. However, the vectors were unstable, most likely due to the large size of the clusters. Therefore, the use of different vectors like BAC-based shuttle vectors, pStreptoBAC V, and pSBAC or ESACS can be proposed.

The activation of BGC 3 and 4 of *Actinoalloteichus fjordicus* by targeting the SARP regulatory genes for overexpression was achieved and it was followed by heterologous expression in *Streptomyces albus* and *Streptomyces coelicolor*. The peaks detected in the HPLC chromatograms of the extracts can be attributed to the presence of novel SMs. Further analysis can be done using mass spectrometry to understand the exact nature of the compounds that give peaks in the chromatograms.

Certain additional strategies to enhance the production of SMs can be considered for further experiments, for example, the use of different fermentation media or different heterologous hosts.

Acknowledgments

It is precisely said that no one who achieves success does so without acknowledging the help of others. I deem it a great pleasure and opportunity to dedicate the following few lines to people who have enabled me to complete this wonderful journey of getting a Master's degree in Drug Discovery and Development at the University of Vienna.

I am thankful to Univ.- Prof. Dr. Sergey B. Zotchev for giving me this wonderful opportunity to work in his research group on such an interesting topic and providing the best of facilities to carry out my research work. I am also grateful to the University for providing quality learning. It has offered me a wonderful environment and opportunity to learn and grow.

I am very fortunate for having been guided by Dr. Olha Schneider. I would treasure everything she has taught me. I take this opportunity to thank her for the great skills and knowledge she has imparted to me.

I have been very fortunate to learn so much from all the members of the Pharmaceutical Biotechnology research group. I would like to express my immense gratitude and respectful regards to them for their moral support throughout my master thesis work.

I couldn't have imagined my stay here in Vienna, without Rujuta, Marilena, and Nadja. Thanks for making ordinary days extraordinary; there has never been a dull moment with them. Thank you for making "Michaelerstrasse 11" a home away from home for me.

I thank Ankita, Umang, and all my friends in India, who have been my confidants and have been there always regardless of different time zones and distinct stages of life.

I would like to express my thanks to Janhavi, Vineet, and Dr. Ameya Kasture for all the guidance, help, and support throughout my journey of Masters in Vienna.

Special thanks to Gaurav for all the good times. I want to thank him for always being a patient listener to all my complaints and for dealing with my whining. His presence has always filled me with hope, strength, and love.

My acknowledgment would not be complete without thanking the biggest source of my strength, my family. I thank my wonderful parents for having faith in me, supporting me, and giving me the freedom to pursue all that I wanted. It is their belief that drove me towards my goal. I am fortunate to have grown under their wings and learn moral values and that it is worth my eternal gratitude. I would also like to thank my grandparents and other family members for their never-ending love and care.

References

[1] Bibb MJ, Janssen GR, Ward JM. Cloning and analysis of the promoter region of the erythromycin resistance gene (ermE) of *Streptomyces erythraeus*. Gene 1985;38(1-3):215–26.

[2] Zaburannyi. Insights into naturally minimised *Streptomyces albus* J1074 genome 2014.

[3] Gabriella Molinari. Natural Products in Drug Discovery Present Status and Perspectives.

[4] Avato P. Editorial to the Special Issue-"Natural Products and Drug Discovery". Molecules (Basel, Switzerland) 2020;25(5).

[5] Patwardhan B, Vaidya ADB, Chorghade M. Ayurveda and natural products drug discovery. Current Science 2004;86(6):789–99.

[6] Brahmachari G. Bioactive Natural Products: Opportunities and Challenges in Medicinal Chemistry (694 Pages).

[7] Amit Koparde A, Chandrashekar Doijad R, Shripal Magdum C. Natural Products in Drug Discovery. In: Perveen S, Al-Taweel A, editors. Pharmacognosy - Medicinal Plants: IntechOpen; 2019.

[8] Newman DJ, Cragg GM. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. Journal of natural products 2020;83(3):770–803.

[9] Amirkia V, Heinrich M. Natural products and drug discovery: a survey of stakeholders in industry and academia. Frontiers in pharmacology 2015;6:237.

[10] Calixto JB. The role of natural products in modern drug discovery. Anais da Academia Brasileira de Ciencias 2019;91 Suppl 3:e20190105.

[11] Hutchings MI, Truman AW, Wilkinson B. Antibiotics: past, present and future. Current opinion in microbiology 2019;51:72–80.

[12] 1,000-year-old onion and garlic eye remedy kills MRSA. BBC News 2015, 30 March 2015; Available from: https://www.bbc.com/news/uk-england-nottinghamshire-32117815. [August 12, 2021].

[13] Atanasov AG, Zotchev SB, Dirsch VM, Supuran CT. Natural products in drug discovery: advances and opportunities. Nature reviews. Drug discovery 2021;20(3):200–16.

[14] Genilloud O. Natural products discovery and potential for new antibiotics. Current opinion in microbiology 2019;51:81–7.

[15] 13.1B: Antibiotic Discovery. Libretexts 2018, 25 June 2018; Available from: https://bio.libretexts.org/Bookshelves/Microbiology/Book%3A_Microbiology_(Boundless)/13%3A_Antimicrobial_Drugs/13.1%3A_Overview_of_Antimicrobial_Therapy/13.1B%3A_Antibiotic_Discovery. [August 13, 2021].

[16] Hug JJ, Bader CD, Remškar M, Cirnski K, Müller R. Concepts and Methods to Access Novel Antibiotics from *Actinomycetes*. Antibiotics (Basel, Switzerland) 2018;7(2).

[17] Davies J, Davies D. Origins and evolution of antibiotic resistance. Microbiology and molecular biology reviews MMBR 2010;74(3):417–33.

[18] Organisation for Economic Co-operation and Development. Stemming the superbug tide: Just a few dollars more. Paris: OECD Publishing; 2018.

[19] Katz L, Baltz RH. Natural product discovery: past, present, and future. Journal of industrial microbiology & biotechnology 2016;43(2-3):155–76.

[20] Shen B. A New Golden Age of Natural Products Drug Discovery. Cell 2015;163(6):1297–300.

[21] Zotchev SB, Sekurova ON, Katz L. Genome-based bioprospecting of microbes for new therapeutics. Current opinion in biotechnology 2012;23(6):941–7.
[22] Patridge E, Gareiss P, Kinch MS, Hoyer D. An analysis of FDA-approved drugs: natural products and their derivatives. Drug discovery today 2016;21(2):204–7.
[23] Zhang L, An R, Wang J, Sun N, Zhang S, Hu J et al. Exploring novel bioactive compounds from marine microbes. Current opinion in microbiology 2005;8(3):276–81.
[24] Sekurova ON, Schneider O, Zotchev SB. Novel bioactive natural products from bacteria via bioprospecting, genome mining and metabolic engineering. Microb. Biotechnol. 2019;12(5):828–44.
[25] Liu Z, Zhao Y, Huang C, Luo Y. Recent Advances in Silent Gene Cluster Activation in Streptomyces. Front. Bioeng. Biotechnol. 2021;9:632230.
[26] Bernardini S, Tiezzi A, Laghezza Masci V, Ovidi E. Natural products for human health: an historical overview of the drug discovery approaches. Natural Product Research 2018;32(16):1926–50.
[27] Zotchev SB. Antibiotics: Biosynthesis.
[28] Olano C, Lombó F, Méndez C, Salas JA. Improving production of bioactive secondary metabolites in actinomycetes by metabolic engineering. Metabolic engineering 2008;10(5):281–92.
[29] Marnix H Medema. Computational approaches to natural product discovery.
[30] Weissman KJ. The structural biology of biosynthetic megaenzymes. Nature chemical biology 2015;11(9):660–70.
[31] Calcott MJ, Ackerley DF. Genetic manipulation of non-ribosomal peptide synthetases to generate novel bioactive peptide products. Biotechnology letters 2014;36(12):2407–16.
[32] Lim YP, Go MK, Yew WS. Exploiting the Biosynthetic Potential of Type III Polyketide Synthases. Molecules (Basel, Switzerland) 2016;21(6).
[33] Gomes ES, Schuch V, Macedo Lemos EG de. Biotechnology of polyketides: new breath of life for the novel antibiotic genetic pathways discovery through metagenomics. Braz. J. Microbiol. 2013;44(4):1007–34.
[34] Jenke-Kodama H, Sandmann A, Müller R, Dittmann E. Evolutionary implications of bacterial polyketide synthases. Molecular Biology and Evolution 2005;22(10):2027–39.
[35] Genome Mining | German Center for Infection Research. [August 15, 2021]; Available from: https://www.dzif.de/en/glossary/genome-mining.
[36] Shi J, Zeng YJ, Zhang B, Shao FL, Chen YC, Xu X et al. Comparative genome mining and heterologous expression of an orphan NRPS gene cluster direct the production of ashimides. Chemical science 2019;10(10):3042–8.
[37] Gomez-Escribano JP, Bibb MJ. Heterologous expression of natural product biosynthetic gene clusters in *Streptomyces coelicolor*: from genome mining to manipulation of biosynthetic pathways. Journal of industrial microbiology & biotechnology 2014;41(2):425–31.
[38] Medema MH, Blin K, Cimermancic P, Jager V de, Zakrzewski P, Fischbach MA et al. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic acids research 2011;39(Web Server issue):W339-46.
[39] Rutledge PJ, Challis GL. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. Nature reviews. Microbiology 2015;13(8):509–23.
[40] Aigle B, Corre C. Waking up *Streptomyces* secondary metabolism by constitutive expression of activators or genetic disruption of repressors. Methods in enzymology 2012;517:343–66.

[41] Baral B, Akhgari A, Metsä-Ketelä M. Activation of microbial secondary metabolic pathways: Avenues and challenges. Synthetic and systems biotechnology 2018;3(3):163–78.
[42] Xue T, Sun M, Chen D, Yuan X, Liu K, Chen J et al. Molecular Cloning and Overexpression of the Transglutaminase Gene from *Streptomyces mobaraensis*. FSTR 2019;25(5):687–94.
[43] Krause J, Handayani I, Blin K, Kulik A, Mast Y. Disclosing the Potential of the SARP-Type Regulator PapR2 for the Activation of Antibiotic Gene Clusters in *Streptomycetes*. Frontiers in microbiology 2020;11:225.
[44] Bilyk O, Sekurova ON, Zotchev SB, Luzhetskyy A. Cloning and Heterologous Expression of the Grecocycline Biosynthetic Gene Cluster. PloS one 2016;11(7):e0158682.
[45] Sharma V, Kaur R, Salwan R. *Streptomyces*: host for refactoring of diverse bioactive secondary metabolites. 3 Biotech 2021;11(7):340.
[46] Ongley SE, Bian X, Neilan BA, Müller R. Recent advances in the heterologous expression of microbial natural product biosynthetic pathways. Nat. Prod. Rep. 2013;30(8):1121–38.
[47] Tocchetti A, Donadio S, Sosio M. Large inserts for big data: artificial chromosomes in the genomic era. FEMS microbiology letters 2018;365(9).
[48] Jiang W, Zhao X, Gabrieli T, Lou C, Ebenstein Y, Zhu TF. Cas9-Assisted Targeting of CHromosome segments CATCH enables one-step targeted cloning of large gene clusters. Nat Commun 2015;6(1):8101.
[49] Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 2009;6(5):343–5.
[50] Nah H-J, Pyeon H-R, Kang S-H, Choi S-S, Kim E-S. Cloning and Heterologous Expression of a Large-sized Natural Product Biosynthetic Gene Cluster in Streptomyces Species. Frontiers in microbiology 2017;8:394.
[51] Yamanaka K, Reynolds KA, Kersten RD, Ryan KS, Gonzalez DJ, Nizet V et al. Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. Proc Natl Acad Sci USA 2014;111(5):1957–62.
[52] Zhang JJ, Yamanaka K, Tang X, Moore BS. Direct cloning and heterologous expression of natural product biosynthetic gene clusters by transformation-associated recombination. In: Shukla AK, editor. Chemical and synthetic biology approaches to understand cellular functions. Part A. Cambridge, MA: Academic Press; 2019, p. 87–110.
[53] Wu N, Huang H, Min T, Hu H. TAR cloning and integrated overexpression of 6-demethylchlortetracycline biosynthetic gene cluster in *Streptomyces aureofaciens*. Acta Biochimica et Biophysica Sinica 2017;49(12):1129–34.
[54] Court DL, Sawitzke JA, Thomason LC. Genetic engineering using homologous recombination. Annual review of genetics 2002;36:361–88.
[55] Wang W, Zheng G, Lu Y. Recent Advances in Strategies for the Cloning of Natural Product Biosynthetic Gene Clusters. Front. Bioeng. Biotechnol. 2021;9:692797.
[56] Maghembe R, Damian D, Makaranga A, Nyandoro SS, Lyantagaye SL, Kusari S et al. Omics for Bioprospecting and Drug Discovery from Bacteria and Microalgae. Antibiotics (Basel, Switzerland) 2020;9(5).
[57] Niu G, Tan H. Biosynthesis and regulation of secondary metabolites in microorganisms. Science China. Life sciences 2013;56(7):581–3.
[58] Chavali AK, Rhee SY. Bioinformatics tools for the identification of gene clusters that biosynthesize specialized metabolites. Briefings in bioinformatics 2018;19(5):1022–34.

[59] MIBiG: Minimum Information about a Biosynthetic Gene cluster. [August 14, 2021]; Available from: https://mibig.secondarymetabolites.org/.
[60] Eicosapentaenoic acid (EPA) biosynthetic gene cluster of *Shewanella oneidensis* MR-1: Cloning, heterologous expression, and effects of temperature and glucose on the production of EPA in *Escherichia coli*.
[61] Choi S-U, Lee C-K, Hwang Y-I, Kinoshita H, Nihira T. Intergeneric conjugal transfer of plasmid DNA from Escherichia coli to *Kitasatospora setae*, a bafilomycin B1 producer. Archives of microbiology 2004;181(4):294–8.
[62] Technical Document, Protocol Manual. CopyRight® v2.0 BAC Cloning Kits: Product: Lucigen.
[63] TransforMax EPI300 Chemically_Electrocompetent *E. coli* _ LG: Lucigen.
[64] Brachmann, Davies, Cost, Caputo, Li, Hieter and Boeke. Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: A useful set of strains and plasmids for PCR-mediated gene disruption and other applications.
[65] Sci-Ed Software - Clone Manager 10: Information. [August 25, 2021]; Available from: https://www.scied.com/in_order.htm.
[66] Soldatova LI, Sladkova IA, Orekhov AV. Funktsionirovanie integrativnykh vektorov, sozdannykh na osnove faga phiC31, v shtamme *Streptomyces bambergiensis* 712. Antibiotiki i khimioterapiia = Antibiotics and chemoterapy [sic] 1994;39(6):3–7.
[67] Mindat.org. [August 25, 2021]; Available from: https://www.mindat.org/taxon-3223697.html.
[68] Orekhov Av, Zotchev Sb, G. Florova, A. Foors, L. Gul'ko, S. Rosenfeld et al. Molecular Genetic Investigation of *Streptomyces Bambergiensis* and *Streptomyces Fradiae* Strains, the Producers of Flavomycin and Tylosin. undefined 1991.
[69] Fu C, Keller L, Bauer A, Brönstrup M, Froidbise A, Hammann P et al. Biosynthetic Studies of Telomycin Reveal New Lipopeptides with Enhanced Activity. J. Am. Chem. Soc. 2015;137(24):7692–705.
[70] Leskiw BK, Lawlor EJ, Fernandez-Abalos JM, Chater KF. TTA Codons in Some Genes Prevent Their Expression in a Class of Developmental, Antibiotic-Negative, Streptomyces Mutants. Proceedings of the National Academy of Sciences of the United States of America 1991;88(6):2461–5.
[71] Silov S, Zaburannyi N, Anisimova M, Ostash B. The Use of the Rare TTA Codon in Streptomyces Genes: Significance of the Codon Context; Indian J Microbiol 2021;61(1):24–30.
[72] Tran PN, Yen M-R, Chiang C-Y, Lin H-C, Chen P-Y. Detecting and prioritizing biosynthetic gene clusters for bioactive compounds in bacteria and fungi. Appl Microbiol Biotechnol 2019;103(8):3277–87.
[73] Jones PM, George AM. The ABC transporter structure and mechanism: perspectives on recent research. Cellular and Molecular Life Sciences (CMLS) 2004;61(6):682–99.

Appendix I

1. BLAST analysis of BGC 18

The table below gives a list of individual genes in BGC 18 with the details of the protein it encodes, along with homologous protein found in other species and the extent of homology. The table highlights core biosynthetic genes in red, transposase encoding genes in yellow and transporter genes in blue.

Table 18: BLAST analysis of BGC18

Orf	**Gene number Ga0065125_**	**Protein**	**Homologous Protein**	**% Identity**
1	115814	Fungal chitosanase of glycosyl hydrolase group 75	glycoside hydrolase family 75 protein from *Streptomyces prasinus*	100
2	115815	dihydroxyacetone kinase DhaM subunit (EC 2.7.1.121)	PTS fructose transporter subunit IIA from *Streptomyces prasinus*	100
3	115816	dihydroxyacetone kinase DhaL subunit (EC 2.7.1.121)	dihydroxyacetone kinase subunit L from *Streptomyces prasinus*	100
4	115817	dihydroxyacetone kinase DhaK subunit (EC 2.7.1.121)	dihydroxyacetone kinase subunit DhaK from *Streptomyces prasinus*	100
5	115818	hypothetical protein	Pirin from *Streptomyces prasinus*	98
6	**115819**	**Transposase**	**ISL3 family transposase from *Streptomyces antimycoticus***	**71**

7	**115820**	**zinc-finger of transposase IS204/IS1001/IS1096/IS1165/Transposase**	**transposase from *Streptacidiphilus jeojiense***	**89**
8	115821	hypothetical protein	ISL3 family transposase from *Streptomyces* sp. NEAU-C40	88
9	115822	hypothetical protein	hypothetical protein M271_50710 from *Streptomyces rapamycinicus* NRRL 5491	76
10	115823	hypothetical protein	hypothetical protein ACZ90_43440 from *Streptomyces albus* subsp. *albus*	86
11	**115824**	**Transposase and inactivated derivatives**	**transposase from *Streptomyces prasinus***	**99**
12	**115825**	**Transposase and inactivated derivatives, IS30 family**	**IS30 family transposase from *Streptomyces prasinus***	**99**
13	**115826**	**Transposase InsO and inactivated derivatives**	**IS3 family transposase from *Streptomyces prasinus***	**99**
14	**115827**	**Transposase**	**IS1182 family transposase from *Streptomyces prasinus***	**100**
15	**115828**	**Transposase DDE domain**	**Transposase from *Streptomyces atratus***	**89**

16	115829	branched-chain amino acid aminotransferase, group II	branched-chain amino acid aminotransferase from *Streptomyces prasinus*	100
17	115830	Formyl transferase	methionyl-tRNA formyltransferase from *Streptomyces prasinus*	100
18	115831	hypothetical protein	LLM class flavin-dependent oxidoreductase from *Streptomyces prasinus*	100
19	115832	hypothetical protein	L-tyrosine 3-hydroxylase from *Streptomyces prasinus*	100
20	115833	Catechol 2,3-dioxygenase or other lactoylglutathione lyase family enzyme	VOC family protein from *Streptomyces prasinus*	100
21	115834	Gamma-glutamyltranspeptidase	gamma-glutamyltransferase from *Streptomyces prasinus*	100
22	115835	Uncharacterized conserved protein YbdZ, MbtH family	MbtH family protein from *Streptomyces prasinus*	100
23	**115836**	**methyltransferase, FkbM family/non-ribosomal peptide synthase domain TIGR01720/amino acid adenylation domain**	**amino acid adenylation domain-containing protein from *Streptomyces prasinus***	**100**
24	115837	Methyltransferase domain	tRNA (adenine (22)-N (1))-methyltransferase	99

			TrmK from *Streptomyces prasinus*	
25	**115838**	**non-ribosomal peptide synthase domain TIGR01720/amino acid adenylation domain**	**amino acid adenylation domain-containing protein from *Streptomyces prasinus***	**100**
26	**115839**	**methyltransferase, FkbM family/non-ribosomal peptide synthase domain TIGR01720/amino acid adenylation domain**	**non-ribosomal peptide synthetase from *Streptomyces prasinus***	**99**
27	**115840**	**amino acid adenylation domain**	**amino acid adenylation domain-containing protein from *Streptomyces prasinus***	**99**
28	115841	Histidinol-phosphate/aromatic aminotransferase or cobyric acid decarboxylase	histidinol-phosphate aminotransferase family protein from *Streptomyces prasinus*	100
29	115842	Uncharacterized conserved protein YbdZ, MbtH family	MbtH family protein from *Streptomyces prasinus*	99
30	115843	Ketosteroid isomerase-related protein	nuclear transport factor 2 family protein from *Streptomyces prasinus*	100
31	115844	Tryptophan 2,3-dioxygenase (vermilion)	hypothetical protein from *Streptomyces prasinus*	99

32	115845	O-methyltransferase	methyltransferase from *Streptomyces prasinus*	100
33	115846	hypothetical protein	transposase from *Streptomyces* sp. DSM 15324	66
34	115847	Cyclopropane fatty-acyl-phospholipid synthase and related methyltransferases	methyltransferase domain-containing protein from *Streptomyces prasinus*	100
35	115848	hypothetical protein	FAD-binding oxidoreductase from *Streptomyces violens*	62
36	115849	HemK family putative methylases	peptide chain release factor N(5)-glutamine methyltransferase from *Streptomyces prasinus*	100
37	115850	chorismate mutase related enzymes	chorismate mutase family protein from *Streptomyces prasinus*	100
38	115851	aminodeoxychorismate synthase, component I, bacterial clade	aminodeoxychorismate synthase component I from *Streptomyces prasinus*	100
39	115852	Prephenate dehydrogenase	prephenate dehydrogenase/arogenate dehydrogenase family protein from *Streptomyces prasinus*	99

40	115853	hypothetical protein	hypothetical protein *Streptomyces* sp. me109	43
41	115854	hypothetical protein	hypothetical protein *Streptomyces* sp. NRRL F-2747	48
42	**115855**	**ATP-binding cassette protein, ChvD family**	**energy-dependent translational throttle protein EttA from *Streptomyces prasinus***	**100**
43	**115856**	**Transposase**	**IS5 family transposase from *Streptomyces prasinus***	**99**
44	115857	hypothetical protein	hypothetical protein from *Streptomyces olivochromogenes*	86
45	115858	hypothetical protein	hypothetical protein *Streptomyces* sp. NL15-2K	91
46	115859	Phage integrase family	tyrosine-type recombinase/integrase from *Streptomyces prasinus*	99

2. BLAST analysis of BGC 20

The table below gives a list of individual genes in BGC 20 with the details of the protein it encodes, along with homologous protein found in other species and the extent of homology. The table highlights core biosynthetic genes in red and regulatory genes in green.

Table 19: BLAST analysis of BGC 20

Orf	**Gene number Ga0065125_**	**Protein**	**Homologous Protein**	**% Identity**
1	116161	hypothetical protein	hypothetical protein from *Streptomyces prasinus*	99
2	116162	L, D-peptidoglycan transpeptidase YkuD, ErfK/YbiS/YcfS/YnhG family	hypothetical protein from *Streptomyces prasinus*	100
3	116163	Integral membrane sensor domain MASE1	MASE1 domain-containing protein from *Streptomyces prasinus*	100
4	116164	Stage II sporulation protein E (SpoIIE)	serine/threonine-protein phosphatase from *Streptomyces prasinus*	100
5	116165	hypothetical protein	MULTISPECIES: hypothetical protein from *Streptomyces*	100
6	116166	Membrane dipeptidase (Peptidase family M19)/F5/8 type C domain	discoidin domain-containing protein from *Streptomyces prasinus*	100
7	116167	Glycerol-3-phosphate dehydrogenase	glycerol-3-phosphate dehydrogenase/oxidase from *Streptomyces prasinus*	100
8	**116168**	**transcriptional regulator, TetR family**	**TetR/AcrR family transcriptional regulator from *Streptomyces prasinus***	**100**
9	116169	FAD/FMN-containing dehydrogenase	FAD-binding oxidoreductase from *Streptomyces prasinus*	100

10	116170	lipid kinase, YegS/Rv2252/BmrU family	YegS/Rv2252/BmrU family lipid kinase from *Streptomyces prasinus*	100
11	116171	Pimeloyl-ACP methyl ester carboxylesterase	alpha/beta hydrolase from *Streptomyces prasinus*	100
12	116172	Cytochrome P450	cytochrome P450 from *Streptomyces prasinus*	99
13	116173	hypothetical protein	hypothetical protein from *Streptomyces prasinus*	100
14	116174	hypothetical protein	hypothetical protein CP972_02570 from *Streptomyces prasinus*	100
15	116175	Uncharacterized membrane protein YjfL, UPF0719 family	DUF350 domain-containing protein from *Streptomyces prasinus*	100
16	116176	hypothetical protein	MULTISPECIES: hypothetical protein from unclassified *Streptomyces*	66
17	116177	Beta-glucanase, GH16 family	family 16 glycosylhydrolase from *Streptomyces prasinus*	100
18	116178	hypothetical protein	hypothetical protein from *Streptomyces prasinus*	100
19	116179	Tellurite resistance protein	TerB family tellurite resistance protein from *Streptomyces prasinopilosus*	98
20	**116180**	**amino acid adenylation domain**	**hybrid non-ribosomal peptide synthetase/type I polyketide synthase from *Streptomyces* sp. H-KF8**	**95**

21	116181	WGR domain predicted DNA-binding domain in MolR	WGR domain-containing protein from *Streptomyces prasinus*	100
22	116182	CO or xanthine dehydrogenase, FAD-binding subunit	FAD binding domain-containing protein from *Streptomyces prasinus*	100
23	116183	Gamma-glutamyl:cysteine ligase YbdK, ATP-grasp superfamily	hypothetical protein from *Streptomyces prasinus*	100
24	116184	Pyridine nucleotide-disulphide oxidoreductase/Reductase C-terminal	FAD-dependent oxidoreductase from *Streptomyces prasinus*	100
25	116185	Ferredoxin	ferredoxin from *Streptomyces* sp. H-KF8	99
26	116186	Cytochrome P450	cytochrome P450 from *Streptomyces prasinus*	100
27	116187	hypothetical protein	hypothetical protein A4U61_15220 in *Streptomyces* sp. H-KF8	93
28	**116188**	**DNA-binding response regulator, OmpR family, contains REC and winged-helix (wHTH) domain**	**MULTISPECIES: response regulator transcription factor in *Streptomyces* spp.**	**100**
29	**116189**	**pyruvate carboxylase**	**pyruvate carboxylase from *Streptomyces prasinus***	**100**
30	116190	hypothetical protein	hypothetical protein from *Streptomyces prasinus*	99
31	**116191**	**PucR C-terminal helix-turn-helix domain**	**helix-turn-helix domain-containing protein from *Streptomyces prasinus***	**100**
32	116192	Glycine/D-amino acid oxidase (deaminating)	FAD-binding oxidoreductase from *Streptomyces prasinus*	100

33	116193	Polyisoprenoid-binding periplasmic protein YceI	YceI family protein from *Streptomyces prasinus*	100
34	**116194**	**transcriptional regulator, TetR family**	**TetR/AcrR family transcriptional regulator from *Streptomyces prasinus***	**100**
35	116195	Cytochrome P450	cytochrome in Streptomyces sp. H-KF8	99
36	116196	CubicO group peptidase, beta-lactamase class C family	serine hydrolase from *Streptomyces prasinus*	99

3. BGC fragments for homologous recombination

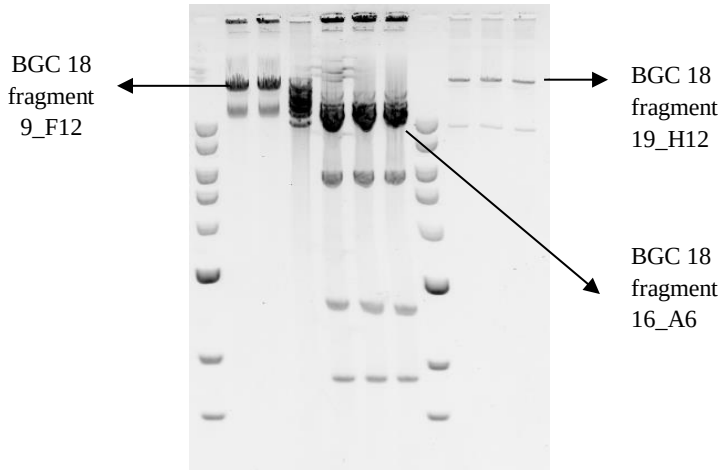
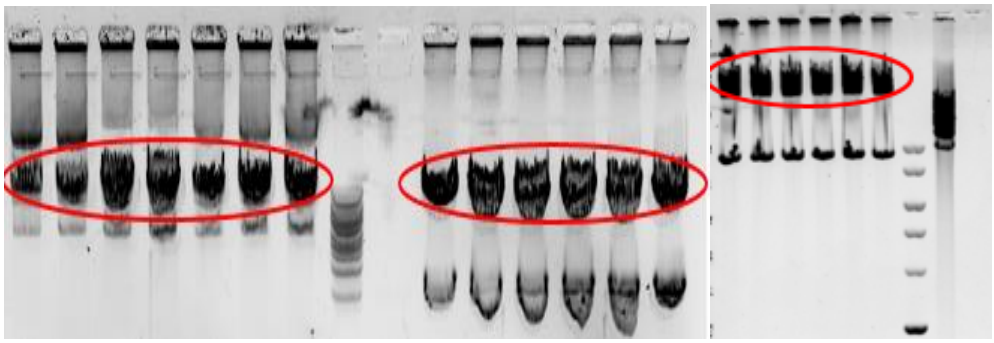


Figure 36: Fragments for BGC 18



8_F7(pFOS1_C20) **10_D7(pFOS2_C20)** **16_G5 (pFOS3_C20)**

Figure 37: Fragments for BGC 20

The fragments (sizes are mentioned in the Results section) marked in red were separated using agarose gel electrophoresis, after restriction digestion using appropriate restriction enzymes. These were then excised from the agarose gel, and the fragmented DNA was purified using a kit. These BGC fragments were used for homologous recombination in yeast.

4. Screening for BGCs after successful assembly in yeast

The screening was performed using colony PCR to confirm the presence of BGC in yeast clones. The clones were tested using primer sets respective for both the *S. bambergiensis* BGCs.

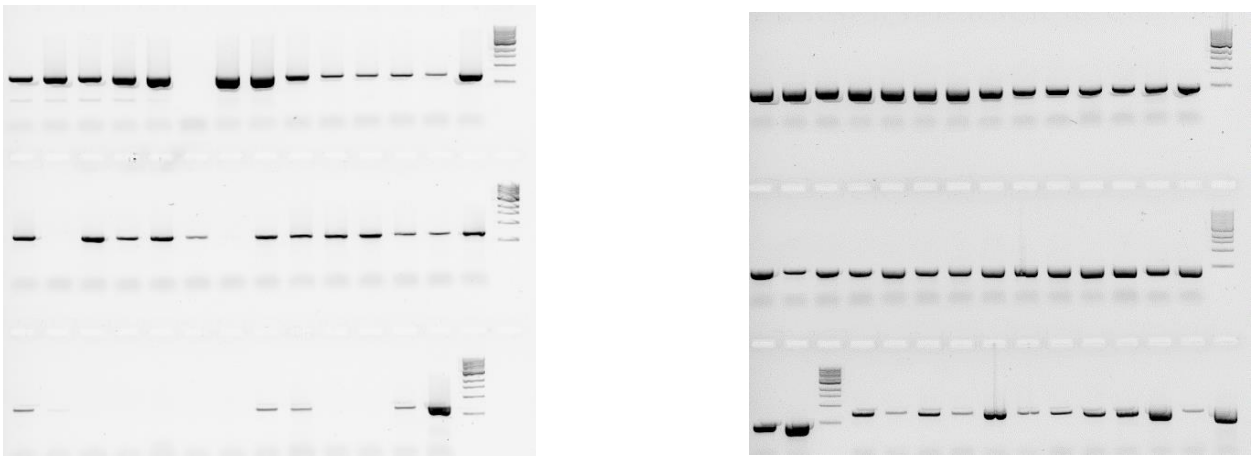


Figure 38: PCR based screening of yeast clones with BGC 18 with respective primers

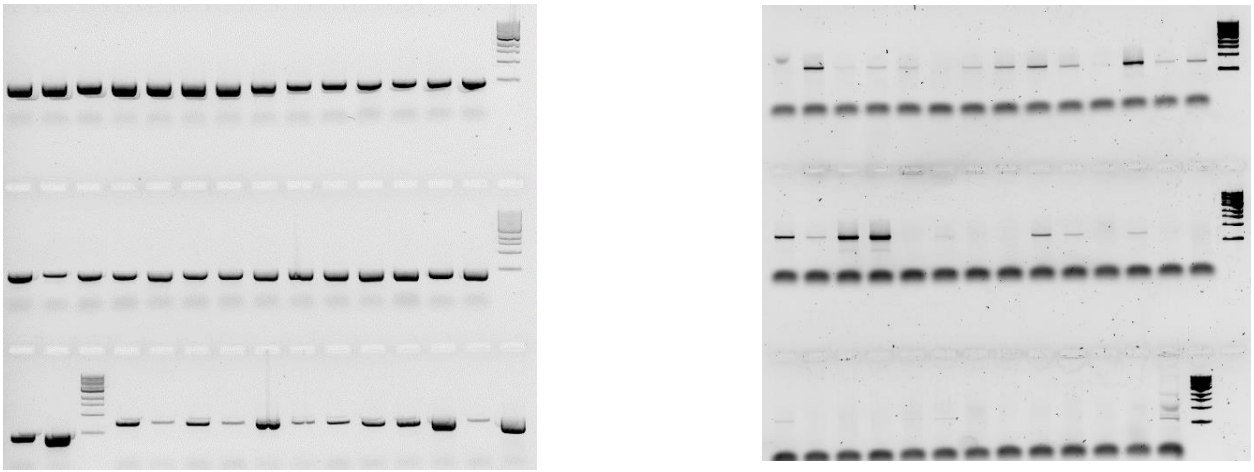


Figure 39: PCR based screening of yeast clones with BGC 20 with respective primers

Figure 37: Fragments for BGC 20