Development of methods for the culture-independent discovery of natural products from soil metagenomes

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Timo Negri

aus Mannheim

Tübingen

2022

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	05.12.2022
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter/-in:	Prof. Dr. Nadine Ziemert
2. Berichterstatter/-in:	apl. Prof. Dr. Evi Stegmann

Table of contents

1. Abstract/Zusammenfassung

The diverse bioactivities of microbial secondary metabolite natural products enabled their use as medical drugs for decades. Especially antibiotics produced by soil bacteria proved to be indispensable for the successful treatment of bacterial infectious diseases. However, the rise of multi-drug resistant bacteria resulting from the ongoing misuse and overuse of antibiotics is currently causing an increasing number of deadly infections, yet the clinical approval rate of new antimicrobial drugs has decreased in the recent past, which is why there is a high demand for new antibiotics. While culture-based natural product discovery from predominantly soil bacteria has so far provided the majority of antibiotics, it nowadays struggles with high rediscovery rates of already known compounds. Since this is partly due to the uncultivable nature of the majority of soil bacteria under conventional laboratory conditions, culture-independent metagenomic natural product discovery approaches provide a promising solution to access this previously missed biosynthetic capacity. However, for that purpose, the natural product encoding biosynthetic gene clusters (BGCs) need to be recovered from soil metagenomes to enable subsequent heterologous expression, which often consists of a laborious and time-consuming process.

In this thesis, an efficient strategy for the rapid identification and recovery of complete natural product BGCs from soil metagenomes was developed. BGCs were recovered from a soil metagenome of the Schönbuch forest, which enabled subsequent heterologous expression experiments aiming at the production of the encoded molecules.

As a first step of the overall process, three different soil types sampled from the Schönbuch forest were investigated for their biosynthetic potential via amplicon and Illumina shotgun sequencing, which revealed a huge biosynthetic capacity of all three soils. High-quality, high molecular weight (HMW) metagenomic DNA was isolated from one of the soil types to construct a metagenomic fosmid library of 83700 clones that was subsequently sequenced using Illumina and Nanopore technologies. Short- and long-reads were used in a hybrid assembly approach to generate contigs that were analyzed to detect complete BGCs. The method single Nanopore read cluster mining (SNRCM) was developed to identify complete BGCs on single fosmids by aligning the detected BGCs directly to Nanopore long-reads. Using SNRCM, two lasso peptide BGCs were recovered and tested in various heterologous expression experiments.

Additionally, the corresponding HMW metagenomic DNA was sequenced, which enabled the direct amplification and recovery of a complete lasso peptide BGC via PCR and subsequent cloning into an expression vector, followed by heterologous expression experiments.

Furthermore, using both the sequencing data of the metagenomic DNA as well as the corresponding fosmid library enabled the assembly and recovery of a nonribosomal peptide synthetase (NRPS) BGC from three distinct fosmids of the library that was subsequently transferred to various heterologous hosts aiming at the production of the encoded compound.

Overall, this thesis contributes to facilitate and accelerate natural product discovery from soil metagenomes, which can potentially lead to the isolation of new medically relevant compounds such as antibiotics in the near future.

Die vielfältigen Bioaktivitäten von mikrobiellen Sekundärmetabolit-Naturstoffen ermöglichen seit Jahrzehnten deren Nutzung als Medikamente. Besonders von Bodenbakterien produzierte Antibiotika haben sich als unverzichtbar für die erfolgreiche Behandlung von bakteriellen Infektionskrankheiten erwiesen. Der Anstieg an multiresistenten Bakterien, der aus der andauernden unsachgemäßen und übermäßigen Verwendung von Antibiotika resultiert, verursacht aktuell jedoch eine steigende Zahl an tödlichen Infektionen und dennoch ist die klinische Zulassungsrate von neuen antimikrobiellen Wirkstoffen in der jüngsten Vergangenheit gesunken, weshalb es einen großen Bedarf an neuen Antibiotika gibt. Während kulturbasierte Naturstoffentdeckung aus vorwiegend Bodenbakterien bisher die Mehrheit an Antibiotika geliefert hat, kämpft sie heutzutage mit hohen Wiederentdeckungsraten von bereits bekannten Stoffen. Da dies teilweise daran liegt, dass die Mehrheit von Bodenbakterien unter konventionellen Laborbedingungen nicht kultivierbar ist, stellen kulturunabhängige metagenomische Naturstoffentdeckungsmethoden eine vielversprechende Lösung dar, um Zugriff auf die zuvor entgangene biosynthetische Kapazität zu erhalten. Dazu müssen Naturstoff codierende biosynthetische Gencluster (BGC) jedoch von Boden-Metagenomen isoliert werden, um eine anschließende heterologe Expression zu ermöglichen, was oft aus einem arbeitsaufwändigen und zeitintensiven Prozess besteht.

In dieser Dissertation wurde eine effiziente Strategie für die schnelle Identifizierung und Isolierung von vollständigen biosynthetischen Naturstoff-Genclustern aus Boden-Metagenomen entwickelt. BGC wurden aus einem Boden-Metagenom des Schönbuchwalds isoliert, was anschließende heterologe Expressionsexperimente ermöglichte, deren Ziel die Produktion der codierten Moleküle war.

Als erster Schritt des gesamten Prozesses wurden drei verschiedene Bodenprobentypen aus dem Schönbuchwald mittels Amplicon und Illumina Shotgun Sequenzierung auf ihr biosynthetisches Potential untersucht, was eine sehr große biosynthetische Kapazität aller drei Böden enthüllte. Hochqualitative, hochmolekulare metagenomische DNA wurde aus einer der

Bodenprobentypen isoliert, um eine metagenomische Fosmid Bibliothek aus 83700 Klonen zu konstruieren, die anschließend unter Verwendung von Illumina und Nanopore Technologien sequenziert wurde. Kurze und lange reads wurden in einer Hybridassemblierungsmethode verwendet, um contigs zu generieren, die analysiert wurden, um vollständige BGC zu detektieren. Die Methode single Nanopore read cluster mining (SNRCM) wurde entwickelt, um vollständige BGC auf einzelnen Fosmiden zu identifizieren, indem die detektierten BGC direkt mit langen Nanopore reads abgeglichen wurden. Unter Verwendung von SNRCM wurden zwei Lasso-Peptid BGC isoliert und in verschiedenen heterologen Expressionsexperimenten getestet.

Zusätzlich wurde die entsprechende hochmolekulare metagenomische DNA sequenziert, was die direkte Amplifizierung und Isolierung eines vollständigen Lasso-Peptid BGC per PCR und die anschließende Klonierung in einen Expressionsvektor ermöglichte, worauf heterologe Expressionsexperimente folgten. Des Weiteren ermöglichte die Verwendung der Sequenzierdaten von sowohl der metagenomischen DNA als auch der entsprechenden Fosmid Bibliothek die Assemblierung und Isolierung eines nichtribosomalen Peptidsynthetase (NRPS) BGCs von drei verschiedenen Fosmiden der Bibliothek, das anschließend in verschiedene heterologe Stämme transferiert wurde, um die Produktion des codierten Stoffes zu erzielen.

Insgesamt trägt diese Dissertation dazu bei die Naturstoffentdeckung von Boden-Metagenomen zu vereinfachen und zu beschleunigen, was potentiell zur Isolierung von neuen medizinisch relevanten Stoffen, wie zum Beispiel Antibiotika, in der nahen Zukunft führen kann.

2. Introduction

2.1 Medical relevance of natural products in the past and the present day

Natural products comprise a large family of diverse molecules with a wide range of biological activities that are biosynthesized by living organisms such as plants, fungi, bacteria and animals. Of particular interest are secondary metabolite natural products that, unlike primary metabolites, such as amino acids, are not essential for the survival and growth of an organism but often confer an advantage to the producer (Katz and Baltz 2016). The various biological activities found within this group of metabolites make them a valuable source of compounds with a broad range of therapeutic applications including their use as antiparasitic, antiviral, anticancer, antifungal or antibacterial drugs (Newman and Cragg 2020). While the medical effects of natural products, e.g. the ones derived from plants, have been applied for thousands of years, microorganisms became a major focus for drug discovery after the discovery of penicillin in the 20th century. The natural products produced by bacteria served as lead structures for the development of a large part of widely used drugs (Wohlleben et al. 2016). The antibiotic fosfomycin, the anti-cancer drug doxorubicin and the immunosuppressant rapamycin, all produced by species of *Streptomyces*, represent only a few examples of compounds that are used in medicine to this day (Hendlin et al. 1969; Seto 2012; Newman and Cragg 2020). However, the most common medical application of natural products has been located in the field of anti-invectives, in particular with the use of antibiotics to treat bacterial infections (Katz and Baltz 2016). These drugs specifically inhibit vital processes in bacteria such as DNA replication, cell wall synthesis, transcription or protein biosynthesis. Depending on the effect of an antibiotic to the respective target bacteria, it is distinguished between a bactericidal effect that kills them and a bacteriostatic effect that inhibits their growth (Clatworthy et al. 2007). While the discovery of antibiotics has led to a tremendous success in the fight against bacterial infectious diseases, it was accompanied by the problem of antibiotic resistance that could be observed for any antibiotic shortly after its clinical application (Palumbi 2001). The ongoing excessive misuse and overuse of antibiotics have led to a tremendous increase in multi-drug resistant bacteria that are currently responsible for a high number of deadly infections. The Centers for Disease Control and Prevention (CDC) estimates that globally this number will increase to 10 million per year by 2050 if no novel antibiotics for clinical use are discovered. At the same time we experienced an antimicrobial drug discovery decline in the past decades due to higher rediscovery rates of known compounds and, as a consequence, a shifting research focus of pharmaceutical companies from infectious diseases

towards the discovery and development of more profitable drugs. As a result, no new antibiotic classes have been introduced to clinical use in more than 30 years, which is why new antibiotics are desperately needed (Martens and Demain 2017; Hutchings et al. 2019; Church and McKillip 2021).

The secondary metabolite natural products of microbes are encoded by specific sets of genes called biosynthetic gene clusters (BGCs), in which the genes responsible for building the respective compound are located in close vicinity to each other within the bacterial genome (Wohlleben et al. 2016). The different classes of secondary metabolites are generally determined by the mode of their biosynthesis. The major classes contain molecules synthesized by polyketide synthases (PKSs), nonribosomal peptide synthetases (NRPSs), terpenoids and ribosomally synthesized and post-translationally modified peptides (RiPPs) (Ziemert et al. 2016).

2.2 Secondary metabolite classes

2.2.1 Polyketides

Polyketides are a structurally diverse group of secondary metabolites including polyethers, polyphenols, polyenes, enediynes and macrolides. They exhibit a wide range of bioactivities, which enables their use as immunosuppressants and antibiotics as well as anticancer, antiparasitic or antifungal drugs (Hertweck 2009; Ogasawara et al. 2015).

Polyketides are synthesized by PKSs that generally perform multiple decarboxylative condensation reactions using acetyl-coenzyme A (CoA) and derivatives as starter units, while malonyl-CoA and derivatives are used as chain extender units (Miyanaga 2019). PKSs can be divided into three different types based on their structural organization. Non-iterative type I PKSs are large multimodular enzymes mainly found in prokaryotes, where each module consists of multiple functional domains and is responsible for the incorporation of one building block (Rawlings 2001; Keatinge-Clay 2012; Zhang et al. 2017). The acyltransferase (AT) loads the extender unit onto the acyl carrier protein (ACP), which carries a phosphopantetheine residue resulting from a post-translational modification. The latter functions as an arm anchoring the growing chain and enables its delivery to the adjacent enzymatic domains for chain extension (Staunton and Weissman 2001). The decarboxylative condensation reaction is performed by the Ketosynthase (KS). Additional optional domains such as a Ketoreductase (KR), a Dehydratase (DH) and an Enoyl reductase (ER) can be found within each module that determine the level of reduction of the respective extender units. Upon release of the product

by the thioesterase (TE) domain, the polyketide can experience further tailoring modifications such as hydroxylation, glycosylation or acylation (Rix et al. 2002; Walsh 2004; Fischbach and Walsh 2006). Iterative type II PKSs produce numerous aromatic polyketides such as the chemotherapeutic agent doxorubicin and are exclusively found in bacteria, predominantly in *Actinomycetes* (Brachmann et al. 2007; Hertweck 2009). The polyketide is synthesized by a set of enzymes that are used iteratively. The minimal PKS is composed of the two KS units KS_α and KS_β as well as an ACP domain. While the KS_α unit performs the condensation reaction for chain elongation, the KS_β unit determines the carbon chain length and is therefore also known as chain length factor. Additional domains such as aromatases, cyclases or ketoreductases determine the folding characteristics of the synthesized polyketide (Rawlings et al. 1999; Shen 2000; Hertweck et al. 2007). Type III PKS are predominantly known from plants, where they can be found as chalcone/stilbene synthases but have also already been found in bacteria and fungi (Moore and Hopke 2001; Pfeifer et al. 2001; Seshime et al. 2005). They are multifunctional homodimeric enzymes that work iteratively and provide all functions necessary for polyketide assembly (Shen 2003). Additionally, mixtures of different PKS types such as type I/type II, type III/type I or fatty acid synthase/PKS hybrids have also been observed (Hertweck 2009).

2.2.2 Nonribosomal peptides

Nonribosomal peptides (NRPs) are mainly produced by bacteria and fungi (Süssmuth and Mainz 2017) and comprise compounds with a wide range of therapeutic application such as immunosuppressants (e.g. Cyclosporin A), antibiotics (e.g. Daptomycin) or antitumor drugs (e.g. Bleomycin A2) (Walsh 2008).

NRPs are synthesized by NRPSs, which are large multimodular enzymes, where each module consists of a minimal set of three enzymatic domains that catalyze the incorporation of an amino acid into a growing peptide chain. The adenylation (A) domain selects an amino acid and activates it by adenylation using adenosine triphosphate (ATP), followed by its transfer to a peptidyl carrier protein (PCP) domain (Schwarzer and Marahiel 2001) (Fig. 1). In addition to the proteinogenic amino acids, various other building blocks such as hydroxy acids, fatty acids or nonproteinogenic amino acids can be incorporated, which greatly contributes to the structural diversity of NRPs (Caboche et al. 2007). The PCP domain, also called thiolation (T) domain, carries a phosphopantetheine arm resulting from a post-translational modification that anchors the growing peptide chain and enables its access to adjacent enzymatic domains (Stachelhaus

et al. 1996; Kittilä et al. 2016). The condensation (C) domain performs the condensation reaction between PCP-bound substrates resulting in an amide or ester bond formation (Stachelhaus et al. 1998). This process is repeated by transferring the growing peptide chain from one module to the adjacent one until the peptide reaches the final module. The peptide is released upon hydrolysis or intramolecular cyclization usually catalyzed by a thioesterase (TE) domain contained within the final module (Fig. 1) (Kopp and Marahiel 2007; McErlean et al. 2019).

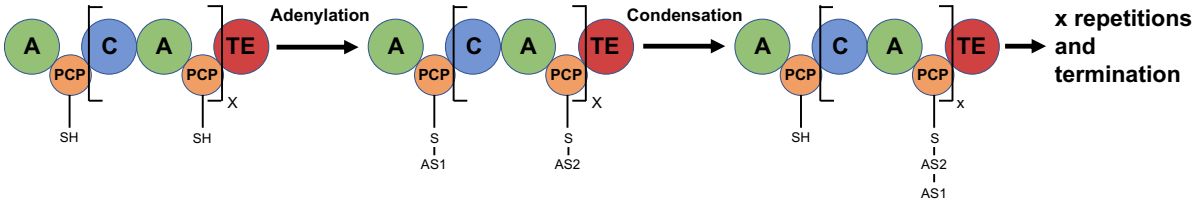


Fig. 1 Classical NRPS assembly line structure. The A domain activates and transfers the amino acid (AS) to the PCP domain. The C domain performs the condensation reaction between two PCP-bound substrates. This process is repeated x times depending on the number of adjacent modules until the peptide is released by the TE domain of the final module

Additionally, NRPs can undergo further tailoring during their synthesis but also after assembly of the complete peptide. Such modifications can include glycosylations, methylations, oxidations or acylations (Walsh et al. 2001; Condurso and Bruner 2012). There are also polyketide-peptide hybrid compounds that are synthesized by hybrids of a type I PKS and a NRPS (Fischbach and Walsh 2006).

2.2.3 Terpenoids

Terpenoids build a huge group of natural products that are especially widespread in plants and fulfill various functions in ecological interactions of organisms such as the defense against predators. Numerous molecules within this group of compounds show different biological activities such as antibacterial, anti-fungal or anticancer activity, which is why they have been used for various medical applications (Rastogi et al. 1998; Lunde and Kubo 2000; Wagner and Elmadfa 2003; Lange 2016). The carbon backbone of terpenoids is synthesized by the condensation of the isoprene derived C5 units isopentenyl diphosphate (IDP) and dimethylallyl diphosphate (DMADP). These building blocks are synthesized by two different pathways namely the mevalonate (MVA) pathway generally found in eukaryotes and archaea and the 2-C-methyl-D-erythritol-4-phosphate (MEP) pathway found in bacteria, protists, algae and higher plants (PHILLIPS et al. 2008; Lombard and Moreira 2011; Frank and Groll 2017).

2.2.4 Ribosomally synthesized and post-translationally modified peptides (RiPPs)

RiPPs comprise a large group of natural products that can be found in all three domains of life. These ribosomally synthesized peptides show a considerable amount of post-translational modifications, making them a structurally diverse group of compounds with a wide range of biological activities (Arnison et al. 2013). Generally, a precursor peptide consisting of a leader and a core peptide is synthesized by the ribosome, followed by post-translational modification of the core peptide resulting from the activity of RiPP tailoring enzymes. Subsequently, the leader peptide that often contains recognition sequences for the binding of the tailoring enzymes is removed from the modified core peptide by cleavage, which generates the final RiPP (Arnison et al. 2013; Ortega and van der Donk 2016). RiPPs are divided into different families based on their mode of biosynthesis and structural characteristics. Major families include lanthipeptides, thiopeptides, linear azoline-containing peptides, linaridins, cyanobactins, bottromycins, thioviridamide-like molecules, sactipeptides and lasso peptides (Russell and Truman 2020). The latter are described in more detail in the following section.

2.2.5 Lasso peptides

Most known lasso peptides were discovered from Proteobacteria or *Actinobacteria* and show various biological activities, although putative BGCs from other phyla of the bacterial domain have been discovered by genome mining approaches (Maksimov and Link 2014; Hegemann et al. 2015). Some lasso peptides act as receptor antagonists, others show antiviral or antimicrobial activity, the latter being the most common one (Maksimov et al. 2012). Two well-known examples of lasso peptides that only function as antibiotics targeting the gram-negative RNA polymerase are MccJ25 and capistruin (Delgado et al. 2001; Kuznedelov et al. 2011), the latter being the first lasso peptide isolated using a genome mining approach (Knappe et al. 2008).

The unique structure of lasso peptides generally shows the following features. Seven to nine amino acids build a macrolactam ring that is formed by an isopeptide bond between the N-terminal amino group of usually a glycine or cysteine and an aspartate or glutamate sidechain at position seven, eight or nine. The C-terminal linear peptide tail is threaded through the macrolactam ring, which generates a structure that resembles a lariat knot (Fig. 2) and gave lasso peptides their name. Apart from the conserved characteristics, the amino acid composition and size of the peptides can vary tremendously. The resulting structure is mainly supported by steric interactions that are sometimes further supported by disulfide bonds. Depending on the number and location of these bonds, the lasso peptides are divided into four classes. Class I

lasso peptides have two, class II have no, and class III and IV have one disulfide bond (Knappe et al. 2010; Maksimov et al. 2012; Hegemann et al. 2015). In class III lasso peptides the disulfide bond is located between the ring and the tail, whereas in class IV it is located in the tail (Fig. 2) (Cheng and Hua 2020).

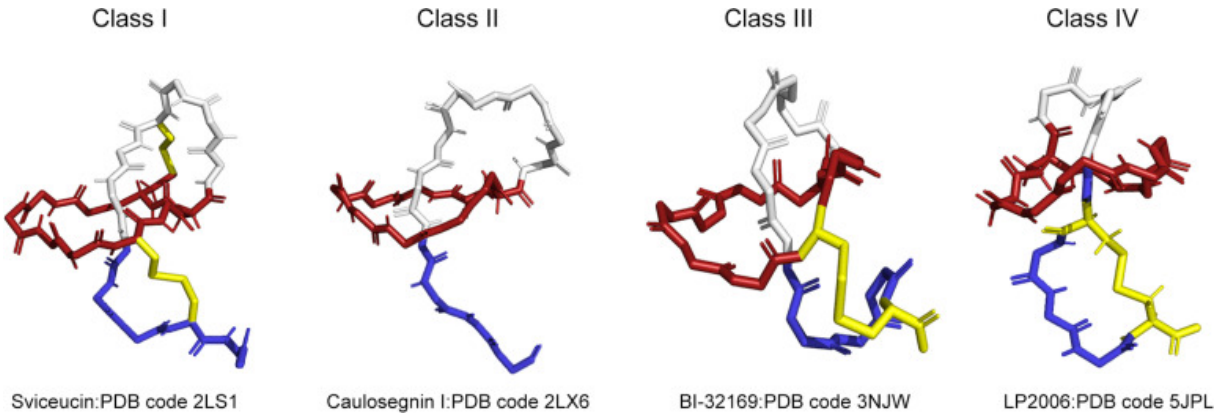


Fig. 2 Examples for lasso peptide classes I to IV. Each class with a representative structure, corresponding lasso peptide name and Protein Data Bank (PDB) code. Red, macrolactam ring; blue, tail; yellow, disulfide bonds. Adopted from (Cheng and Hua 2020)

The minimal set of genes necessary for lasso peptide biosynthesis consists of three genes (Fig. 3). The A gene codes for the precursor peptide consisting of the leader and core peptide. The leader peptide is recognized and cleaved from the core peptide by an ATP-dependent cysteine protease encoded by the B gene. The C gene codes for an ATP-dependent macrolactam synthetase that catalyzes the isopeptide bond formation leading to the macrolactam ring (Martin-Gómez and Tulla-Puche 2018). The B protein can also be present in a split form consisting of B1 representing the N-terminal domain of a complete B protein, while B2 represents the C-terminal domain. In these cases B1 binds to the leader peptide, which enables the cleavage by B2. In addition to this minimal set, further genes such as a self-resistance conferring D gene encoding an ABC-transporter can be found. BGCs with genes coding for a lasso peptide specific isopeptidase instead of a D gene have also been found. Moreover, BGCs containing a kinase for lasso peptide tailoring are known (Hegemann et al. 2013b; Zhu et al. 2016b; Zhu et al. 2016a) (Fig. 3).

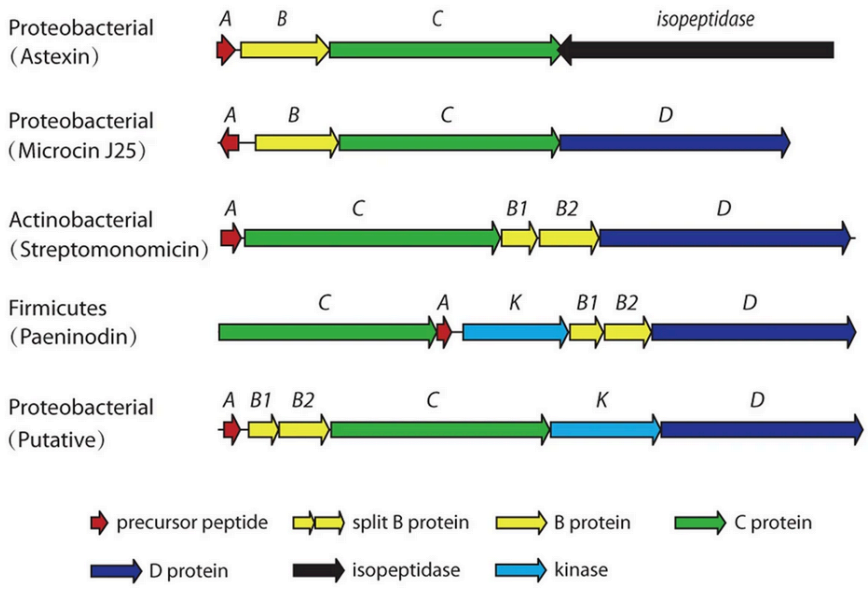


Fig. 3 Structural organization of different lasso peptide BGCs. All BGCs contain the minimal set of genes (A, B and C) for lasso peptide biosynthesis. Examples with additional genes and variations are shown; modified from (Zhu et al. 2016a)

2.3 Natural product discovery

2.3.1 Culture-based natural product discovery

Starting from the discovery of penicillin in the 20th century, more than 23 thousand natural products of predominately bacterial origin, mainly from the *Actinomycetaceae* family, have been described (Bérdy 2012). Early natural product screening efforts using actinomycetes isolated from soil led to the discovery of well-known compounds such as actinomycin or streptomycin, which marked the starting point for the extensive screening of actinomycete extracts for various bioactivities. These approaches made use of bioassays to perform a phenotypic screening. For that purpose, microorganisms were isolated by plating soil suspensions on agar plates. Grown colonies were cultured in different media with varying conditions such as temperature and resulting culture broths and extracts were tested for growth inhibition against other organisms such as bacteria or yeast. This was done by applying filter paper discs carrying the extracts to agar plates containing the respective test organism. After incubation overnight, the plates could be investigated for growth inhibition of the test organisms (Fig. 4).

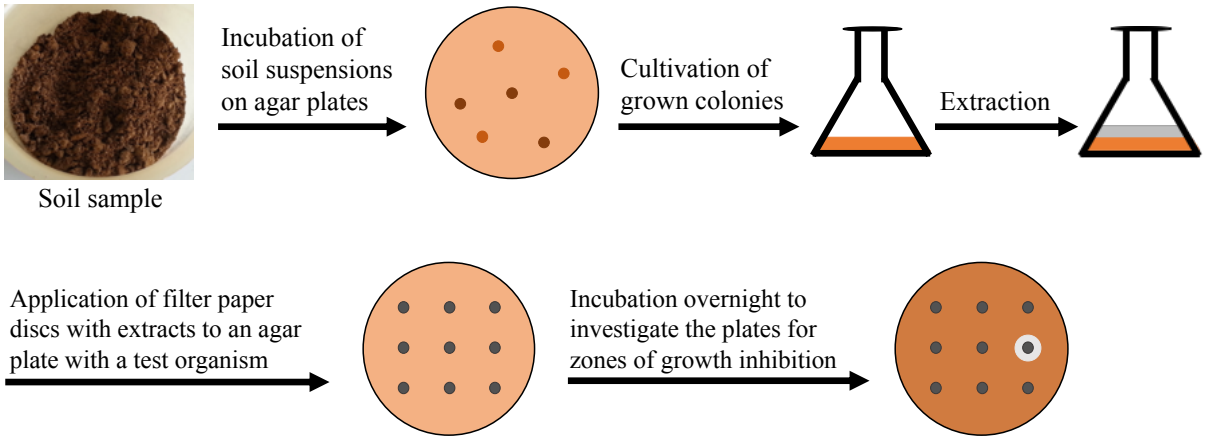


Fig. 4 Culture-based natural product discovery. Plating of soil suspensions on agar plates. Cultivation of grown colonies in different media and subsequent extraction of the cultures. Application of filter paper discs carrying the extracts to agar plates containing test organisms. Incubation overnight and investigation of plates for zones of growth inhibition

For compounds that could not be extracted using a solvent, an agar plug assay was performed. For that purpose, small agar plugs were removed from a plate with a test organism and the resulting holes were filled with culture broth from a soil isolate. Zones of growth inhibition that appeared after incubation overnight identified producers of antibiotics. Compounds responsible for the observed activity were subsequently separated using chromatographic methods, followed by their isolation and structural characterization using mass spectrometry (MS) and nuclear magnetic resonance (NMR) spectroscopy (Breton and Reynolds 2013; Havlicek et al. 2013; Katz and Baltz 2016). For some compounds the mode of action was additionally investigated. Applying this approach, and later also to alternative sources such as marine environments, numerous natural products with antibacterial, antifungal or anticancer activity were discovered, of which many were later approved as drugs for clinical use (Katz and Baltz 2016). However, the ongoing use of these approaches finally led to the more and more frequent rediscovery of already known compounds (Baltz 2006), which was partially due to the fact that the majority of soil bacteria such as the previously exploited actinomycetes could not be cultured using conventional cultivation methods (Kellenberger 2001; Schloss and Handelsman 2003). To address this problem, various new cultivation methods for soil-derived *Actinobacteria* such as the cultivation in microfluidic droplets have been developed. The latter allows the generation of thousands of pure cultures per hour and enables access to potential natural products of slow growers that could not have been isolated using standard cultivation methods (Zang et al. 2013; Wohlleben et al. 2016). Another method for accessing previously uncultivable environmental microorganisms was provided by the development of an isolation chip (ichip), which consists of hundreds of tiny diffusion chambers that are each inoculated with a single cell. The ichip is incubated in the natural environment of the cells, which allows

the diffusion of nutrients necessary for the growth of the respective microorganisms (Nichols et al. 2010). Applying this cultivation method, the new antibiotic teixobactin was discovered upon isolation of a previously uncultured β-proteobacterium called *Eleftheria terrae* (Ling et al. 2015).

2.3.2 Genome mining

The genome mining approach differs from the classical screening approach in that it first aims at the detection and analysis of secondary metabolite encoding BGCs, followed by their connection to the respective compounds. Genome-based approaches started in the beginning of the 21st century with the first complete genome sequences of *Streptomyces coelicolor* and *Streptomyces avermitilis*. The high number of secondary metabolite BGCs found within their genomes revealed that the capacity for natural product production is much higher than previously observed based on actually discovered compounds (Bentley et al. 2002; Ikeda et al. 2003; Challis 2014; Ikeda et al. 2014). With more genomes available, these findings could also be observed for other actinomycetes, which led to the conclusion that only a small part of natural product BGCs is expressed in detectable amounts under fermentation conditions. The remaining ones, often called silent gene clusters, can only be accessed by adjusting culture conditions or applying genetic manipulations. The latter can include deletions of negative regulators, overexpression of positive regulators, duplication of entire BGCs (Bachmann et al. 2014; Baltz 2016) or the replacement of a BGC's native promoter with a constitutive one (Franke et al. 2012; Rutledge and Challis 2015). Technological advances in genome sequencing and bioinformatics as well as the growing knowledge on secondary metabolite biosynthesis enabled the detection of new secondary metabolite gene clusters and in some cases the prediction of encoded compounds based on the analysis of genomic sequences (Medema and Fischbach 2015; I. Tietz and A. Mitchell 2016). Taken together, these advances resulted in the availability of huge amounts of genomic sequencing data as well as efficient and readily accessible genome mining tools such as antiSMASH (Medema et al. 2011) or PRISM (Skinnider et al. 2015) and enabled access to a large variety of potential new compounds encoded by respective BGCs. The possibility of structure predictions based on sequence information of BGCs also contributed to circumvent the increasing problem of rediscovering known compounds (Baltz 2006). Additionally, the comparison of BGCs of interest to databases such as the MIBiG (Minimum Information about a Biosynthetic Gene cluster) database (Kautsar et al. 2019) also enabled the selection of novel BGCs potentially encoding new compounds. Furthermore, the more targeted nature of genome-based approaches in

combination with new cloning techniques for large BGCs, e.g. transformation-associated recombination (TAR) cloning, opened up new possibilities such as direct cloning of sequenced BGCs for subsequent heterologous expression (Yamanaka et al. 2014; Zhang et al. 2019). For these purposes, various strains have been genetically modified to develop expression hosts that are optimized for the heterologous production of secondary metabolites. Such genetic modifications can include the deletion of native active secondary metabolite BGCs for better precursor availability and easier detectability of heterologously produced compounds or the introduction of point mutations to genes encoding RNA polymerase subunits as it has been applied to develop a *Streptomyces coelicolor* host for example (Gomez-Escribano and Bibb 2011).

With increasing numbers of available secondary metabolite BGCs it became more and more important to prioritize the ones encoding bioactive compounds. However, even if a structure prediction is feasible, the prediction of bioactivity with potential therapeutic application remains challenging. Therefore, target-directed genome mining (Tang et al. 2015) has been developed, which analyzes detected BGCs for the presence of potential resistance genes as microbial producers of antibiotics need to be resistant to their own products. This resistance can be conferred by the presence of a second modified copy of the target gene within an antibiotic encoding BGC, which enables the target-directed mining by exploiting this knowledge. An automation of this process has been realized with the development of the Antibiotic Resistant Target Seeker (ARTS) tool (Alanjary et al. 2017; Mungan et al. 2020). Although these and other new methods for natural product discovery have been developed in the recent past, overall, we nevertheless experienced an antimicrobial drug discovery decline with no new antibiotic classes introduced to clinical use in more than 30 years as pointed out earlier (Martens and Demain 2017; Hutchings et al. 2019; Church and McKillip 2021).

2.3.3 Culture-independent metagenomic natural product discovery

Nowadays we know from numerous metagenomic studies that even with the advances made in culture-based approaches we still only have access to a tiny proportion of the thousands of individual bacterial species and thus the biosynthetic diversity that can be present in as little as one gram of soil. Furthermore, these studies have shown that the so far uncultivated microbes also contain bacteria that are very distantly related to the frequently isolated and closely related species of actinobacteria (Roesch et al. 2007; Hug et al. 2016). Moreover, studies focusing on the generation and subsequent analysis of metagenomic biosynthetic domain sequences

similarly found that the majority of these are only distantly related to characterized BGCs (Charlop-Powers et al. 2015). Therefore, it can be assumed that secondary metabolites potentially produced by so far uncultivable bacteria also differ significantly from the known ones, which might be a solution to the discovery of new antibiotic classes for example. To access this hidden bacterial and biosynthetic diversity, culture-independent natural product discovery approaches, i.e. metagenomic approaches, have been developed. For that purpose, the DNA of the microorganisms contained within an environmental sample such as soil is directly isolated without any culturing step. The DNA is then cloned into a cultivable host organism that serves a genetic library (Fig. 5A), which is the starting point for different approaches that finally aim at the heterologous expression of new secondary metabolite BGCs encoded within the genomes of environmental bacteria (Handelsman et al. 1998). For that purpose, the clones of a generated library need to be screened for the presence of secondary metabolite BGCs, which is mainly performed using either a functional or a sequence-based screening approach.

The functional screening is generally based on an untargeted phenotypic or chromatographic screening that aims at identifying clones producing bioactive secondary metabolites (Fig. 5B). In the past, this kind of screening was mainly performed by generating metagenomic *E. coli* cosmid or bacterial artificial chromosome (BAC) libraries and clones were investigated for phenotypes indicating secondary metabolite production such as antimicrobial effects, color changes or HPLC peaks (Rondon et al. 2000; MacNeil et al. 2001; Brady et al. 2001; Brady et al. 2002; Brady et al. 2004). The advantage of functional screening is that it is not dependent on knowledge of sequence information for the identification of bioactive clones and thus positive clones can potentially harbor novel secondary metabolite genes without any similarity to known ones. Although functional screening approaches led to the discovery of some new natural products (MacNeil et al. 2001; Gillespie et al. 2002), they revealed to be comparatively inefficient with low hit rates because of the following reasons. First, the commonly used host for library generation and screening *E. coli* is a comparatively unsuitable host for the heterologous expression of metagenomic secondary metabolite BGCs as compared to the previously more successful host *Streptomyces* for example (Wexler and Johnston 2010). Therefore BGCs might remain undiscovered because of failing heterologous expression in *E. coli*, which can result from issues with promoter recognition, translation or proper folding of resulting proteins (Madhavan et al. 2017). Second, the untargeted nature of the functional screening contributes to the inefficiency because generally only a small part of the bacterial

genome codes for secondary metabolite BGCs (Bentley et al. 2002; Ikeda et al. 2003; Wexler and Johnston 2010; Craig et al. 2010).

The sequence-based screening uses similarity-based DNA sequence analysis as a first step to identify clones carrying BGCs of interest (Fig. 5C), followed by heterologous expression of the biosynthetic pathways to produce the respective compounds. Thus, this screening method circumvents the aforementioned drawbacks of functional screening as the detection of clones of interest does not necessitate heterologous expression as a first step and moreover the sequence analysis provides a more targeted approach (Katz et al. 2016).

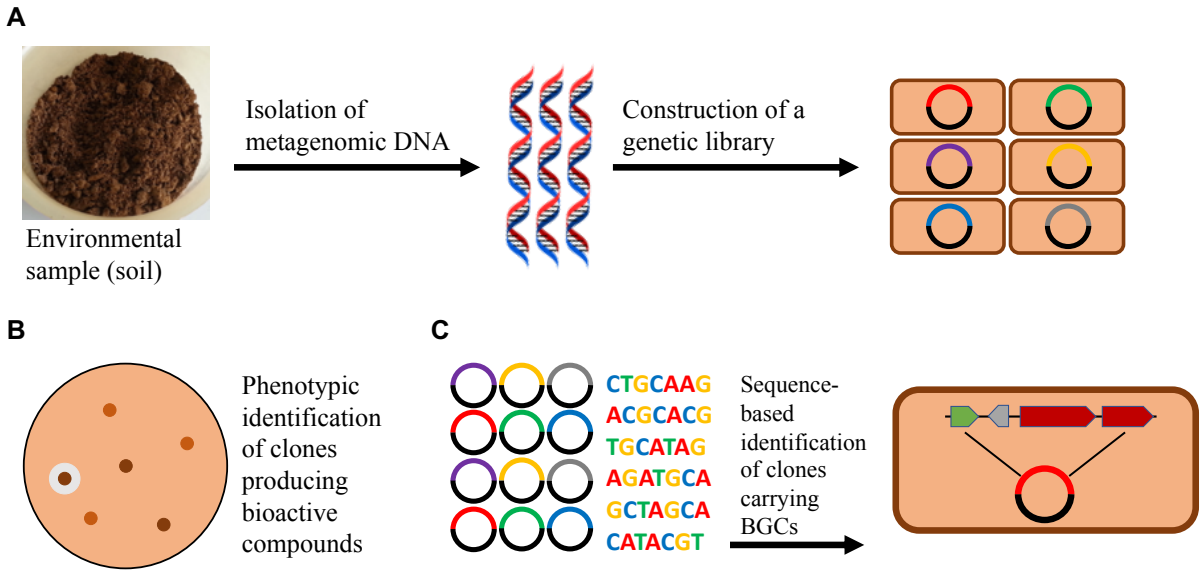


Fig. 5 Culture-independent natural product discovery. A) Isolation of metagenomic DNA from an environmental sample and subsequent construction of a genetic library. **B)** Functional screening based on phenotypic identification of clones producing bioactive compounds. **C)** Sequence-based screening for the identification of clones carrying secondary metabolite BGCs; partly adopted and modified from (Negri et al. 2022)

The sequence analysis is also often directly applied to the metagenome of respective samples prior to metagenomic library generation in order to prioritize sites for high biosynthetic potential or specific BGCs of interest. Technological advances in sequencing and bioinformatics combined with a significant decrease in sequencing cost nowadays enable the direct sequencing of soil metagenomes. Furthermore, efficient tools enable the subsequent assembly of metagenomic sequencing data (Nurk et al. 2017), which can yield the recovery of complete BGC sequences in silico (Crits-Christoph et al. 2018; Waschulin et al. 2021). However, in the past it was not possible to access natural product BGCs in this manner. Therefore, early attempts used degenerate primers for conserved regions of biosynthetic domains, mainly A and KS domains, to generate amplicons via PCR. Sequencing and bioinformatic analysis of generated amplicons further confirmed the assumption of a huge

undiscovered biosynthetic diversity of soil metagenomes in general (Charlop-Powers et al. 2014; Charlop-Powers et al. 2015; Charlop-Powers et al. 2016). As standard bioinformatics tools for secondary metabolite genome mining such as antiSMASH usually depend on large input sequences or whole genomes, specific tools like NaPDoS (The Natural Product Domain Seeker) or eSNaPD (Environmental Surveyor of Natural Product Diversity) have been developed that use amplicon sequences of biosynthetic domains as an input. These tools perform phylogenetic analyses of biosynthetic domain sequence datasets by comparing them to characterized biosynthetic genes to assess biosynthetic diversity as well as predict potential chemical novelty and medicinal relevance of compounds encoded by the corresponding BGCs (Ziemert et al. 2012; Reddy et al. 2014). Applying the PCR-based screening for biosynthetic domains to metagenomic libraries enabled the identification of clones carrying respective BGCs. An alternative method for identifying clones of interest relied on the generation of DNA probes that were similarly designed based on conserved regions of biosynthetic domains and subsequently used in hybridization experiments (Parsley et al. 2011). After the recovery of positive clones, the BGCs could be investigated for subsequent heterologous expression attempts (Bauer et al. 2010; Amos et al. 2015). The analysis of recovered clones often revealed the presence of only partially captured BGCs, predominantly in cases of BGCs that were larger than the maximum insert size of the used vector. Nevertheless, in general, also smaller BGCs can reveal to be incomplete due to the random cloning process. To obtain full BGCs, the libraries were screened for clones harboring overlapping corresponding parts of the respective BGC using specific primers. The BGC parts were then assembled to a complete cluster using transformation-associated recombination (TAR) cloning (Hover et al. 2018; Wu et al. 2019). This cloning technique is performed by the co-transformation of yeast with a linearized pathway-specific capture vector and the linearized plasmids carrying the overlapping cluster parts, which results in the assembly of the BGC via homologous recombination. Subsequently, the assembled BGC can be transferred to suitable heterologous hosts (Zhang et al. 2019). This approach is not only time-consuming but also dependent on a high coverage of the metagenome to guarantee that a BGC of interest is completely covered by the library. A more recent strategy applied a short-read shotgun sequencing approach to a metagenomic library, followed by the generation of contigs that were subsequently analyzed for encoded BGCs (Santana-Pereira et al. 2020). In contrast to the PCR screening approach, this strategy directly reveals BGCs that are completely covered within the library and thus suitable for heterologous expression. Nevertheless, it does not provide information on completeness of BGCs within single clones either.

So far metagenomics has already proven to be a promising alternative for natural product discovery that enables access to the previously hidden biosynthetic potential of microorganisms, especially from soil. Nevertheless, the identification and recovery of BGCs remains a laborious and time-consuming process that leaves potential for optimization to accelerate and facilitate the overall process. Similarly, heterologous expression of metagenomic BGCs is currently a bottleneck often hindering metagenomic natural product discovery.

3. Materials and methods

3.1 Materials

3.1.1 Equipment

Table 1 Devices

Device	**Manufacturer**
T100 Thermal Cycler	Bio-Rad
Gel imaging system	Nippon Genetics Europe
Nanodrop 2000c spectrophotometer	Thermo Fisher Scientific
Electrophoresis power supply MP-300V	Major Science
Centrifuge 5424 (R)	Eppendorf
Heraeus multifuge 3SR+ centrifuge	Thermo Fisher Scientific
Heraeus multifuge X1R centrifuge	Thermo Fisher Scientific
Sorvall RC 6+ centrifuge	Thermo Fisher Scientific
Ultrospec 10 cell density meter	Amersham Biosciences
MicroPulser Electroporator	Bio-Rad
Precellys 24	Bertin Technologies
pH meter	METTLER TOLEDO
Special accuracy weighing machine Analytic AC 210P	Sartorius
Blockthermostat BT 1303	HLC
Incubator Shaker Innova 44	New Brunswick
Ultrapure water system GenPure Pro	TKA
Rotary evaporator	Heidolph Instruments
Magnetic stirrer RSM-01HS	Phoenix Instrument GmbH
High-performance liquid chromatography (HPLC) device 1260 Infinity	Agilent Technologies
InfinityLab LC/MSD	Agilent Technologies

3.1.2 Media

Table 2 Media

Medium	**Ingredients**	**Amount**
Lysogeny broth (LB) medium (Sigma Aldrich)	Yeast extract	5 g/l
	Tryptone	10 g/l
	Sodium chloride	5 g/l
Tryptic Soy Broth (TSB) medium (BD Bacto)	Pancreatic digest of casein	17 g/L
	Papaic digest of soybean	3 g/L
	Dextrose	2.5 g/L
	Sodium chloride	5 g/L
	Dipotassium phosphate	2.5 g/L

2S4G medium (Eustáquio et al. 2016)	Glycerol (Sigma-Aldrich)	40 g/l
	Soya peptone (Oxoid)	20 g/l
	Ammonium sulfate (Roth)	2 g/l
	Magnesium sulfate (Roth)	0.06 g/l
	Calcium carbonate (Merck)	2 g/l
M9 medium (Zhu et al. 2016b)	disodium hydrogen phosphate dihydrate (CHEMSOLUTE)	8.5 g/l
	potassium dihydrogenphosphate (fisher chemical)	3 g/l
	Sodium chloride (Merck)	0.5 g/l
	Ammonium chloride (Merck)	1 g/l
	1 M Magnesium sulfate (AppliChem) solution	2 ml/l
	1 M calcium chloride (Merck) solution	0.1 ml/l
	Adjusted to pH 7, followed by the addition of the remaining ingredients	
	40% glucose (Roth) solution for cultures with no arabinose induction	10 ml/l
	Glycerol for cultures with arabinose induction	4 g/l
	Vitamin mix	2 ml/l
Vitamin mix for M9 medium (Zhu et al. 2016b)	Choline chloride (Sigma-Aldrich)	1 g/0.3l
	Folic acid (Merck)	1 g/0.3l
	Pantothenic acid (Aldrich-Chemie)	1 g/0.3l
	Nicotinamide (Sigma-Aldrich)	1 g/0.3l
	Myo-inositol (Merck)	2 g/0.3l
	Pyridoxal hydrochloride (Alfa Aesar)	1 g/0.3l
	Thiamine (Merck)	1 g/0.3l
	Riboflavin (Alfa Aesar)	0.1 g/0.3l
	Disodium adenosine 5'-triphosphate (Sigma-Aldrich)	0.3 g/0.3l
	Biotin (Sigma-Aldrich)	0.2 g/0.3l
	Adjusted to pH 12 with 10 M sodium hydroxide (VWR)	
M20 medium (Knappe et al. 2008)	Glutamic acid (AppliChem)	20 g/l
	L-alanine (Serva)	0.2 g/l
	Sodium citrate (VWR)	1 g/l
	Disodium hydrogen phosphate (CHEMSOLUTE)	20 g/l
	Potassium chloride (Sigma-Aldrich)	0.5 g/l
	Sodium sulfate (Roth)	0.5 g/l
	Magnesium chloride (Roth)	0.2 g/l
	Calcium chloride (Roth)	0.0076 g/l
	Iron (II) sulfate (Merck)	0.01 g/l
	Manganese sulfate (Merck)	0.0076 g/l
	Adjusted to pH 7	
	Thiamine added for *E. coli* cultures	2 mg/l
	Biotin added for *E. coli* cultures	2 mg/l
M8 medium	Glycerol (Sigma-Aldrich)	20 g/l

	Sodium chloride (Merck)	0.5 g/l
	Ammonium sulfate (Roth)	2 g/l
	Magnesium sulfate heptahydrate (AppliChem)	0.2 g/l
	Calcium chloride dihydrate (Roth)	0.1 g/l
	potassium dihydrogenphosphate (fisher chemical)	1 g/l
	TES (Roth)	22.9 g/l
	Adjusted to pH 7, followed by the addition of the remaining ingredient	
	Separately autoclaved 6 mM Zinc sulfate heptahydrate (Merck)	1 ml/l
R5 medium	Sucrose (Roth)	103 g/l
	Glucose (Roth)	10 g/l
	Yeast extract (Roth)	5 g/l
	Magnesium chloride (Roth)	10.12 g/l
	TES (Roth)	5.73 g/l
	Potassium sulfate (Roth)	0.25 g/l
	Casamino acids (Bacto)	0.1 g/l
	R5 Trace element solution	2ml/l
	Adjusted to pH 7.2, followed by the addition of the remaining and separately autoclaved ingredients	
	1 M Calcium chloride solution	20 ml/l
	0.54% potassium dihydrogenphosphate solution	10 ml/l
	20% L-Proline (Roth) solution	20 ml/l
R5 Trace element solution	Iron(III) chloride hexahydrate (Merck)	200 mg/l
	Zinc chloride (Fluka Chemika)	40 mg/l
	Copper(II) chloride dihydrate (Fluka Chemika)	10 mg/l
	Manganese(II) chloride tetrahydrate (Merck)	10 mg/l
	Ammonium heptamolybdate tetrahydrate (Merck)	10 mg/l
	sodium tetraborate decahydrate (Merck)	10 mg/l
Cullum medium	Mannitol (Merck)	20 g/l
	Soy flour (Hensel)	20 g/l
For solid media 16 g/l agar-agar (Roth) was added		

3.1.3 Buffers and solutions

Table 3 Buffers and solutions

Buffer/Solution	**Ingredients**	**Concentration**
P1 buffer	Trizma-Base (Sigma-Aldrich)	50 mM
	ethylenediaminetetraacetic acid (Merck)	10 mM
	RNase A (Sigma-Aldrich)	100 μg/ml
	Adjusted to pH 8 with hydrochloric acid (fisher chemical)	
P2 buffer	sodium hydroxide (Sigma-Aldrich)	200 mM

	sodium dodecyl sulfate (Serva)	1 %
P3 buffer	potassium acetate (AppliChem)	3 M
	Adjusted to pH 5.5 with acetic acid (Sigma-Aldrich)	
50x TAE buffer	Trizma-Base (Sigma-Aldrich)	2 M
	sodium acetate (Roth)	500 mM
	ethylenediaminetetraacetic acid (Merck)	50 mM
	Adjusted to pH 7.8 with acetic acid	
Tris-EDTA (TE) buffer	Trizma-Base (Sigma-Aldrich)	10 mM
	ethylenediaminetetraacetic acid (Merck)	1 mM
	Adjusted to pH 8	
Lysis buffer HMW metagenomic DNA isolation 1 (Brady 2007)	Trizma-Base (Sigma-Aldrich)	100 mM
	ethylenediaminetetraacetic acid (Merck)	100 mM
	Sodium chloride (Merck)	1.5 M
	cetyl trimethyl ammonium bromide (Serva)	1 %
	sodium dodecyl sulfate (Serva)	2 %
	Adjusted to pH 8	
Extraction buffer HMW metagenomic DNA isolation 3 (Verma et al. 2017)	Trizma-Base (Sigma-Aldrich)	100 mM
	ethylenediaminetetraacetic acid (Merck)	100 mM
	Sodium phosphate buffer	100 mM
	Sodium chloride (Merck)	1.5 M
	cetyl trimethyl ammonium bromide (Serva)	1 %
	Calcium chloride (Roth)	100 mM
	proteinase K (Roth)	10 mg/ml
	Lysozyme (Sigma-Aldrich)	10 mg/ml
	Adjusted to pH 8	
Solution 1	Sucrose	10 %
	Trizma-Base (pH 8)	50 mM
	Ethylenediaminetetraacetic acid	10 mM

3.1.4 Additional chemicals

Table 4 Additional chemicals

Chemical	**Manufacturer**
L-arabinose	Roth
Agarose	Genaxxon bioscience
6x DNA loading Dye	Thermo Fisher Scientific
Phenol:chloroform:isoamyl alcohol	Roth
Ethidium bromide	Roth
Maltose	Merck
Ethanol	Honeywell

Isopropanol	Honeywell
Yeast nitrogen base without amino acids and ammonium sulfate	Sigma-Aldrich
Yeast synthetic drop-out medium supplements without tryptophan	Sigma Aldrich
Adenine	Merck
5-fluoroorotic acid	Thermo Scientific
2-Mercaptoethanol	Merck
Sorbitol	Sigma-Aldrich
Peptone	Merck
Dimethyl sulfoxide	Merck
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)	Serva
PEG8000	Roth
Chloroform	Sigma-Aldrich
Isoamyl alcohol	Merck
Acetonitrile	Honeywell
Formic acid	VWR

3.1.5 Antibiotics

Table 5 Antibiotics

Antibiotic (Manufacturer)	**Concentration in media**	**Abbreviation**
Chloramphenicol (Serva)	12.5 μg/ml	CHL
Kanamycin sulfate (Roth)	50 μg/ml (*E. coli*)	KAN
	500 μg/ml (*Burkholderia*)	KAN500
Nalidixic acid (AppliChem)	25 μg/ml	NAL
Ampicillin (AppliChem)	100 μg/ml	AMP
Gentamycin sulfate (Roth)	50 μg/ml	GEN

3.1.6 Restriction enzymes

Table 6 Restriction enzymes

Restriction enzyme	**Manufacturer**
BamHI	Thermo Scientific
SalI	Thermo Scientific
SacI	Thermo Scientific
DraI	Thermo Scientific
Ecl136II	Thermo Scientific
Eco105I	Thermo Scientific
PsiI	Thermo Scientific
XhoI	Thermo Scientific

NdeI	Thermo Scientific
PmeI	Thermo Scientific
SmiI	Thermo Scientific
PvuI	Thermo Scientific
DpnI	Thermo Scientific

3.1.7 Enzymes

Table 7 Enzymes

Enzyme	**Manufacturer**
T4 DNA Ligase	Thermo Fisher Scientific
Exonuclease V	NEB
Taq DNA-Polymerase	Genaxxon bioscience
Zymolyase-20T	MP Biomedicals
RNase I	Boehringer Mannheim
RNase A	Sigma-Aldrich
Lysozyme	Sigma-Aldrich
Proteinase K	Roth

3.1.8 DNA ladders

Table 8 DNA ladders

DNA ladder	**Manufacturer**
Easy Ladder I	BioLine
Lambda DNA/HindIII Marker	Thermo Scientific
GeneRuler 1kb DNA Ladder	Thermo Scientific

3.1.9 Kits

Table 9 Kits

Kit	**Manufacturer**
CopyControl Fosmid Library Production Kit with pCC1FOS Vector	Lucigen
PowerLyzer PowerSoil DNA Isolation Kit (#12855-100)	MO BIO Laboratories
Genomic DNA Clean & Concentrator-10 kit	Zymo Research
Q5 High-Fidelity DNA Polymerase kit (M0491L)	NEB
Zymoprep Yeast Plasmid Miniprep I kit	Zymo Research
RevertAid RT Reverse Transcription Kit	Thermo Scientific
QIAquick PCR purification Kit (50)	QIAGEN
NEBuilder HiFi DNA Assembly Master Mix	NEB

QIAquick Gel Extraction Kit	QIAGEN
Monarch Total RNA Miniprep Kit	NEB
DNA-free DNA Removal Kit	Invitrogen

3.1.10 Sequencing primers

All sequencing primers were designed using Geneious version 9.1.8.

Table 10 Sequencing primers

Sequencing primer	Sequence	Target
pCC1FOS_Cl40_seq_1	GAAATCAGTGATTGCCGGCG	Sequencing primers targeting pCC1FOS_Cl40 for the generation of sequences with overlap that cover lasso peptide BGC 40.1
pCC1FOS_Cl40_seq_2	GGATGAGAAAGTTTGGGCGC	
pCC1FOS_Cl40_seq_3	CCCGCAAGAATTCGCGATTT	
pCC1FOS_Cl40_seq_4	CCAAGCCGAGATGCTGAGAA	
pCC1FOS_Cl40_seq_5	TTGACGTCAAGAACACGACT	
pCC1FOS_Cl40_seq_6	CACGATGTGCATGGCGCTAG	
pCC1FOS_Cl40_seq_7	CGGAAGAGGGAACGGAGCGG	
pCC1FOS_Cl482_seq_1	ATTGTCTGGCTTCCTCTGCG	Sequencing primers targeting pCC1FOS_Cl482 for the generation of sequences with overlap that cover lasso peptide BGC 482.1
pCC1FOS_Cl482_seq_2	CGTAACTACGCTGTGATGGA	
pCC1FOS_Cl482_seq_3	ATTTCGCCTTTCGTTGCTCG	
pCC1FOS_Cl482_seq_4	ACCATGTCACTGCTTGTCCC	
pCC1FOS_Cl482_seq_5	TCTAGTTGTCGCATGGGACG	
pCC1FOS_Cl482_seq_6	TGACCTGGAAATGTCCCGTG	
pCC1FOS_Cl482_seq_7	TACGACGACAACTTGCGTCC	
pCC1FOS_Cl482_seq_8	TCGAGGAGTGAAGGACAACG	
pSK019_Cl44_D_seq_1	GCGTCACACTTTGCTATGCC	Sequencing primers targeting pSK019_Cl44_D for the generation of sequences with overlap that cover the cloned lasso peptide BGC 44.1
pSK019_Cl44_D_seq_2	GTGACGAGCGCAACTTTGTT	
pSK019_Cl44_D_seq_3	GAGCACCTCCACTTCACTCC	
pSK019_Cl44_D_seq_4	GCTGGGCGAGCTTTTTCTTT	
pSK019_Cl44_D_seq_5	CCTAAAGATCCCGCAGCAGT	
pSK019_Cl44_D_seq_6	GAGCAGGAAAATCCGGGAGT	
pSK019_Cl44_D_seq_7	GGTGAGGCTGTAGCAATCGT	
pSK019_Cl44_D_seq_8	ACCTGTCTCGCTTCGTTGAG	

3.1.11 Primers

Table 11 Primers

Primer	Sequence	Target	Annealing temperature/ Amplicon size
Adom_fw	GCSTACSYSATSTACACSTCSGG	conserved regions in NRPS A domains (Pimentel-Elardo et al. 2012)	58.5 °C/ approx. 750 bp
Adom_rv	SASGTCVCCSGTSCGGTAS		
KSI_fw	CCSCAGSAGCGCSTSYTSCTSGA	conserved regions in PKSI KSI domains (Ginolhac et al. 2004)	68 °C/ approx. 670 bp
KSI_rv	GTSCCSGTSCCGTGSGYSTCSA		
Cl_44_SalI_fw	GTCGACCCTCCGTCGCAGAGCTGTAT (SalI recognition site underlined)	Lasso peptide BGC 44.1 genes A, B1, B2 and C	67 °C/ 3034 bp
Cl_44_SacI_rv	GAGCTCAAGATGTTCCTGACCTGCGG (SacI recognition site underlined)		
Cl44_D_gene_OV_fw	GGTCAGGAACATCTTGAGCCCGAACAGCAATGACAGAAC (overhang underlined)	Lasso peptide BGC 44.1 genes D1 and D2	66 °C/ 1814 bp
Cl44_D_gene_OV_rv	CATGATTACGAATTCGAGCGCCGCCTTCTTGCAATTAA (overhang underlined)		
NRPS_BGC_screen_left_fw	GGTGACCCGACAATTCCCAT	Left part of NRPS BGC 76.1	67 °C (Q5 Polymerase), 55 °C (Taq Polymerase)/ 615 bp
NRPS_BGC_screen_left_rv	TCACCGTGAGCTTCAGTGAC		
NRPS_BGC_screen_middle_fw	CGGATTCCTGTGCTCTGGTT	Middle part of NRPS BGC 76.1	67 °C (Q5 Polymerase), 55 °C (Taq Polymerase)/ 594 bp
NRPS_BGC_screen_middle_rv	TTGCCAATTAGACCGGACCC		
NRPS_BGC_screen_right_fw	CAAAGACACGCAAGCAGCTT	Right part of NRPS BGC 76.1	66 °C (Q5 Polymerase), 54 °C (Taq Polymerase)/ 476 bp
NRPS_BGC_screen_right_rv	TCTTTGAGCAGGGTTCCGTC		
Caparm_fw	CATGGTATAAATAGTGGC	144 bp fragment for BGC 76.1 TAR cloning	53 °C/ 144 bp
Caparm_rv	TATGTAGCTTTCGACATA		
Lambda_red_pCC1FOS_fw	GAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGCAAGGGCTGCTAAAGGAAG	KAN resistance, oriT and ori1600 gene of pSK019	66 °C/ 3434 bp

Lambda_red_pCC1FOS_rv	AGGCGTTTAAGGGCACCAATAACTGCCTTAAAAAAATTATCTAGGCCAGATCCAGCG		
Lambda_red_control_primer_fw	GTTGATACCGGGAAGCCCTG	Flanking regions of pCC1FOS CHL resistance gene	66 °C/ 933 bp for non-replaced CHL resistance gene; 3635 bp for lambda red recombination event
Lambda_red_control_primer_rv	GCCGTCGACCAATTCTCATG		
Cl482_A_tr_fw	ACAAAGCCAGAAGTCGTCGT	A gene parts of BGC 482.1	54 °C/ 366 bp and 65 bp
Cl482_A_tr_rv	CCCACCGACTTCGACAGTTG		
Cl482_B1_tr_fw	GTTCGCAGCACACAAACCAT	B1 gene part of BGC 482.1	54 °C/ 243 bp
Cl482_B1_tr_rv	GACTGCATTCGTGGACTGGA		
Cl482_C_tr_fw	GAGTCGTATTCCAGCGAGGG	C gene part of BGC 482.1	54 °C/ 628 bp
Cl482_C_tr_rv	CCGCCTCGGCAAATACATTG		
Cl44_B1_tr_fw	TGGATGCACCTCTCTCCTCA	B1 gene part of BGC 44.1	55 °C/ 276 bp
Cl44_B1_tr_rv	CCTCGATCAGTCCTTCGCTC		
Cl44_B2_C_tr_fw	GAACTGGCGGAATTGATCGC	B2 to C gene part of BGC 44.1	54 °C/ 874 bp
Cl44_B2_C_tr_rv	TCACGGAAAGTAGTGGTGGC		
Cl44_C_tr_fw	AAGGAATGGAGCCAGCTTCC	C gene part of BGC 44.1	53 °C/ 513bp
Cl44_C_tr_rv	TTAAGCATTGCGAACCTCCG		
Cl44_D_tr_fw	ATTGCTGCTCGATGAACCCA	D gene part of BGC 44.1	54 °C/ 809 bp
Cl44_D_tr_rv	AAGAGTACCGCATAGCTCGC		

3.2 Methods

3.2.1 General methods

All prepared agarose gels were composed of 1 % agarose in 1x TAE buffer. If not described differently, electrophoresis was conducted at 90 V for 45-55 minutes. Gels were stained with ethidium bromide and subsequently visualized using a gel imaging system.

If not described differently, all plasmids and fosmids were isolated by alkaline lysis as described in section 3.2.4.

If not described differently, all PCR reactions using the Q5 High-Fidelity DNA Polymerase kit were performed according to the conditions listed in Tables 12 and 13, while conditions for PCR reactions performed with the Taq DNA-Polymerase are listed in Tables 14 and 15. Respective annealing temperatures used for each primer pair are listed in Table 11. For all negative controls water was used instead of template DNA.

Table 12 Q5 PCR reaction mixture

Component	**Volume**
5x Q5 reaction buffer	5 µl
5x Q5 High GC Enhancer	5 µl
10 µM forward/reverse primer	0.5 µl
10 mM deoxynucleoside triphosphates (dNTPs)	0.5 µl
template DNA	3 µl
Q5 high-fidelity DNA polymerase	0.25 µl
nuclease-free water	10.25 µl
Total volume	**25 µl**

Table 13 Q5 Thermocycling conditions

PCR step	**Duration**	**Temperature**	**Cycles**
Initialization	30 s	98 °C	1x
Denaturation	10 s	98 °C	30x
Annealing	30 s	Primer specific (see Table 11)	
Elongation	20 s/kb	72 °C	
Final extension	2 min	72 °C	1x

Table 14 Taq DNA-Polymerase PCR reaction mixture

Component	**Volume**
10x CoralLoad PCR Buffer (QIAGEN)	2.5 µl
10 µM forward/reverse primer	0.5 µl
10 mM deoxynucleoside triphosphates (dNTPs)	0.5 µl
template DNA	3 µl
Taq DNA-Polymerase 5 U/µl (Genaxxon bioscience)	0.2 µl
nuclease-free water	17.8 µl
Total volume	**25 µl**

Table 15 Taq DNA-Polymerase Thermocycling conditions

PCR step	Duration	Temperature	Cycles
Initialization	3 min	95 °C	1x
Denaturation	30 s	94 °C	30x
Annealing	30 s	Primer specific (see Table 11)	
Elongation	1 min/kb	72 °C	
Final extension	5 min	72 °C	1x

If not described differently, all precultures were grown in glass culture tubes, while main cultures were grown in Erlenmeyer flasks.

3.2.2 Generation of electrocompetent *E. coli* DH10ß and *E. coli* GB05-MtaA cells

For *E. coli* GB05-MtaA, LB media was supplemented with 50 µg/ml gentamycin (GEN) to maintain the selection pressure for the phosphopantetheinyl transferase (PPtase) gene integrated into the chromosome.

5 ml LB medium was inoculated with 5 µl of an *E. coli* DH10ß glycerol stock or 5 µl of an *E. coli* GB05-MtaA glycerol stock and the culture was grown at 37 °C with shaking overnight. 1 ml of the overnight culture was used to inoculate 100 ml LB medium and the culture was grown to an OD_{600} of 0.5-0.7 at 37 °C with shaking. Following, the culture was split to 2x 50 ml and pelleted by centrifugation (3000 rpm, 4 °C, 10 min). Each pellet was resuspended in 25 ml of cooled 10 % glycerol and subsequently pelleted again. Resuspension in 25 ml of cooled 10 % glycerol followed by centrifugation was repeated. Subsequently, the supernatant was discarded and the pellet resuspended in the backflow. The cells were stored in 100 µl aliquots at -80 °C until use.

3.2.3 Transformation of electrocompetent *E. coli* DH10ß and *E. coli* GB05-MtaA cells via electroporation

100 µl of electrocompetent *E. coli* DH10ß or *E. coli* GB05-MtaA cells were mixed with 5 µl of the respective plasmid and the mixture was transferred to an electroporation cuvette (2 mm gap, VWR). Cells were electroporated using a MicroPulser electroporator (program EC2) and the cuvette was subsequently filled with 1 ml of LB medium. The mixture was transferred to a reaction tube and incubated for 1 hour at 37 °C with shaking. The cells were pelleted,

resuspended in the backflow and plated on LB agar supplemented with the respective antibiotic, followed by incubation at 37 °C overnight.

3.2.4 Plasmid and fosmid isolation by alkaline lysis

All centrifugation steps were carried out at 13000 rpm.

1 ml of the respective culture was pelleted by centrifugation (3 min) and the pellet was resuspended in 200 µl of P1 buffer. 200 µl of P2 buffer was added and the tube was inverted multiple times for mixing. After the addition of 350 µl of P3 buffer, the tube was inverted multiple times again and subsequently incubated at 4 °C for 5 min. Following, the reaction tube was centrifuged for 5 min and the supernatant was transferred to a new tube containing 750 µl of cooled isopropanol. After mixing by inversion, the tube was centrifuged for 20 min at 4 °C. The resulting pellet was washed with 1 ml of 70 % ethanol, followed by centrifugation for 5 min at 4 °C. The supernatant was discarded and the pellet dried at room temperature for 10-15 min, followed by resuspension in 30-50 µl of water.

3.2.5 Sampling of seven soil samples for Illumina and amplicon sequencing

In May 2019, soils were sampled from the Schönbuch Forest nature reserve that is located close to Tübingen (Germany) and harbors the three different soil types podzol, cambisol and stagnosol all in close proximity to each other. The soil types each consisted of multiple horizons with different characteristics. Using a soil probe the organic layer as well as the A horizon were sampled from each of the three soil types and the B horizon was additionally sampled for cambisol soil. All samples were stored at -20 °C and sieved through a fine mesh screen prior to metagenomic DNA isolation.

3.2.6 Soil sampling for the isolation of high molecular weight (HMW) metagenomic DNA

In November 2016 the A horizon of cambisol soil was sampled from the Schönbuch forest and subsequently stored at -20 °C. The soil sample was sieved through a coarse mesh screen prior to HMW metagenomic DNA isolation.

The methods of sections 3.2.7 and 3.2.8 were published in (Mantri et al. 2021) and are briefly summarized with the main steps in the following.

3.2.7 Isolation of metagenomic DNA from the seven soil samples and subsequent Illumina and amplicon sequencing

Metagenomic DNA from the O and A horizons of 250 mg of podzol, cambisol, and stagnosol soil was isolated using the PowerLyzer PowerSoil DNA isolation kit and following an alternative protocol provided by Mo Bio. Metagenomic DNA from the B horizon of cambisol soil was isolated using the method described for HMW DNA 3 in section 3.2.8. Illumina sequencing was performed by CeGaT GmbH on a NovaSeq 6000 PE150 system.

Amplicons were generated via PCR using the metagenomic DNA of the 7 soils and degenerate primers for conserved regions of A domains in NRPSs (Adom_fw/rv) and KSI domains in PKSs (KSI_fw/rv) (Table 11). PCR was performed using the Q5 High-Fidelity DNA Polymerase kit. Generated amplicons of four PCR reactions for each sample and primer pair were pooled, purified (QIAquick PCR purification kit for A domain amplicons; QIAquick gel extraction kit for KSI domain amplicons) and subsequently verified via agarose gel electrophoresis. Sequencing was performed by the NGS Competence Center Tübingen (NCCT) on a MiSeq system.

Clustering of generated amplicon sequences into operational biosynthetic units (OBUs) and generation of rarefaction curves was performed by Shrikant Mantri and is described in detail in (Mantri et al. 2021).

3.2.8 Isolation and sequencing of HMW metagenomic DNAs from the A horizon of cambisol soil, hybrid assembly of short- and long-reads, and identification of BGCs on contigs

The methods for the isolation of HMW metagenomic DNA 1 and 3 were additionally published in (Negri et al. 2022).

HMW metagenomic DNA 1 was isolated from a 250 g soil sample using a published protocol (Brady 2007) and adding a modification to increase the DNA purity. In the following, the main steps including the applied modification are listed. The soil sample was treated with lysis buffer for two hours at 70 °C, followed by centrifugation and subsequent addition of isopropanol to the resulting supernatant. The precipitated DNA was pelleted, resuspended in TE and subsequently gel purified using a large agarose gel. The gel was size selected for HMW DNA

and the DNA was electroeluted into a dialysis tube (ZelluTrans dialysis tube, MWCO 8000-10000, width 25 mm, Roth) containing 1x TAE. As an addition to the protocol, the dialysis tube was incubated in 0.5x TE buffer overnight. The DNA was concentrated using a centrifugal concentrator (MWCO 30000) (Amicon) and subsequently analyzed on an agarose gel. Library preparation and Nanopore sequencing were performed by genXone on a GridIon device.

Metagenomic DNA 2 was generated by further purifying HMW metagenomic DNA 1 using the spin columns of the PowerLyzer PowerSoil DNA isolation kit. Briefly, the DNA sample was filled up to 650 µl with H_2O and subsequently mixed with 650 µl of solution C4 and 650 µl of 100% ethanol. The resulting mixture was transferred to a Mo Bio spin column and the DNA was bound by centrifugation. After washing the membrane with 650 µl of 100 % ethanol and 500 µl of solution C5, the column was dried by centrifugation. The DNA was eluted with H_2O and subsequently analyzed on an agarose gel. Illumina sequencing was performed by the CeGaT GmbH on a NovaSeq 6000 PE150 system.

HMW metagenomic DNA 3 was isolated from 6x 5 g of soil using a published protocol (Verma et al. 2017) with several modifications to increase DNA yield and purity. The published protocol was based on enzymatic and chemical lysis of cells within the soil, followed by multiple DNA purification and precipitation steps. In the following, the applied modifications are listed. After the DNA was pelleted for the first time, followed by resuspension in TE buffer, 1 µl of RNase I was added and the mixture was incubated for 30 min at 37 °C. For the second DNA precipitation step with isopropanol, a 0.1 volume of 5 M sodium acetate was included. Finally, the isolated DNA was further gel purified using a large agarose gel as described for isolation of HMW metagenomic DNA 1 and subsequently analyzed on an agarose gel. Sequencing was performed by the NGS Competence Center Tübingen (NCCT) on a PromethION device.

Metagenomic DNA concentrations and absorbance ratios were measured using the Nanodrop 2000c spectrophotometer.

Contigs were generated from resulting short- and long-reads using a metaSPAdes (Nurk et al. 2017) based hybrid assembly approach. Hybrid assembled contigs were filtered for sizes greater than 25 kb and BGCs were identified using antiSMASH version 5 (Blin et al. 2019) (hybrid assembly, size filtering and antiSMASH analysis performed by Shrikant Mantri).

The methods of sections 3.2.9 to 3.2.13 were published in (Negri et al. 2022) and are briefly summarized with the main steps in the following.

3.2.9 Generation and sequencing of a metagenomic fosmid library

A metagenomic fosmid library was generated using the CopyControl Fosmid Library Production Kit with pCC1FOS Vector and HMW metagenomic DNA from the A horizon of cambisol soil. Approximately 8000 clones were generated using HMW metagenomic DNA 1, while approx. 75700 clones were generated from HMW metagenomic DNA 3. The following modifications have been applied to the manufacturer's instructions of the CopyControl Fosmid Library Production Kit: The size selection step after end-repair of the DNA was excluded as this step was already contained in the DNA isolation process and therefore the DNA already had the appropriate size. For each packaging reaction the ligation reaction was scaled up to 20 µl and performed at 16 °C overnight.

After plating, clones were stored in pools of 2000 clones (storage method C of the manual). The fosmids of ten randomly selected clones were isolated and digested with BamHI overnight to release the inserts and subsequently assess their sizes via agarose gel electrophoresis. 100 ml LB medium supplemented with 12.5 µg/mL chloramphenicol (CHL) and 1x CopyControl Fosmid Autoinduction Solution (Autoinduction solution) was inoculated with 20 µl of each pool and the culture was grown at 37 °C with shaking overnight. After harvesting the cells by centrifugation, the fosmids were isolated, further processed and sequenced by the NGS Competence Center Tübingen (NCCT). For Illumina sequencing, a library was prepared using the TruSeq DNA PCR-Free kit (Illumina) and paired-end sequencing (2 x 150 bp) was performed on a NextSeq 550 system using a NextSeq Mid Output flow cell. For Nanopore sequencing a library was prepared using the Rapid Barcoding Kit (SQK-RBK004) and sequencing was conducted on a MinION device using a MinION flow cell.

3.2.10 Hybrid assembly approach for the fosmid library sequencing data and identification of complete natural product BGCs captured on single fosmids via single Nanopore read cluster mining (SNRCM)

Hybrid assembly of short- and long-reads of the fosmid library sequencing data was performed using metaSPAdes (Nurk et al. 2017). Resulting hybrid contigs were filtered for different lengths (greater than 1 kb, 25 kb and 40 kb) and BGCs were identified using antiSMASH version 5 (Blin et al. 2019). Identified BGCs on contigs greater than 25 kb and 40 kb were

investigated for BGCs that were annotated as being complete. For the identification of clones harboring complete BGCs on a single fosmid via SNRCM, complete BGCs on contigs greater than 40 kb were aligned with Nanopore reads using BLAST (Zhang et al. 2000) (hybrid assembly, antiSMASH analysis, size filtering and alignment with Nanopore reads performed by Shrikant Mantri).

3.2.11 Clone recovery from pools of 2000 clones via serial dilution PCR

Using specific primers for each clone of interest, a serial dilution PCR strategy (Owen et al. 2015) was applied to recover clones from pools of 2000 clones. Briefly, the dilution of a positive *E. coli* pool overnight LB culture (CHL) to an OD_{600} of 0.25 x 10^{-5} (≈ 2000 cells/ml) allowed the inoculation of multiple LB cultures (CHL, Autoinduction solution) with approx. 200 cells (100 µl) and their subsequent screening for a positive signal via PCR. The positive pool was then further diluted to an OD_{600} of 0.25 x 10^{-6} (≈ 200 cells/ml), which allowed the inoculation of multiple cultures with approx. 40 cells (200 µl) and their subsequent screening for a positive signal. The positive pool was then diluted and plated on LB agar (CHL) to yield single colonies that were finally screened for the target gene via PCR.

In case of disappearing positive PCR signals after further dilution and screening of the positive 200 clones pool, the following alternative dilution method was used. The positive *E. coli* pool was grown in LB medium (CHL) at 37 °C with shaking overnight and subsequently diluted to an OD_{600} of 0.25 x 10^{-5} (≈ 2000 cells/ml). 200 µl (≈ 400 cells) were used to inoculate 96 wells of four 24-well plates each containing 0.8 ml LB (CHL, Autoinduction solution) and the cultures were grown at 37 °C with shaking overnight. Subsequently, 12 wells each were pooled and fosmids isolated, followed by screening for positive pools via PCR. The 12 wells of the resulting positive pool were then screened individually. The resulting positive well was diluted and plated to yield single colonies. 144 individual colonies were picked and grown in three 48 well plates containing 1 ml LB (CHL, Autoinduction solution) at 37 °C with shaking overnight. 16 wells each containing individual clones were pooled and the resulting nine pools were screened again. The 16 wells of the resulting positive pool were subsequently screened individually, which led to the identification of a single clone positive for the target gene.

3.2.12 Bioinformatics methods for the identification of lasso peptide specific genes and sequencing confirmation of lasso peptide BGCs on fosmids

Blastx and antiSMASH analyses were conducted to identify the minimal set of the lasso peptide biosynthesis gene homologues A, B and C for the metagenomic BGCs 40.1 and 482.1. For BGC 44.1 the genes A, B, C and D were identified using the same approach. For the latter BGC, the A gene could directly be identified by antiSMASH annotation. For BGC 482.1 the two A genes were identified by manual investigation of the translated nucleotide sequence of candidate genes for the features common to a lasso peptide. The B1 genes were identified by comparing the Blastx analysis results, gene sizes and antiSMASH annotations of known lasso peptide B1 genes from the MIBiG database (Kautsar et al. 2019) to those of the metagenomic lasso peptide BGCs. The B2 genes could directly be identified by Blastx analysis and the C genes could directly be identified by antiSMASH, which annotates identified C genes as "Asn synthase". The D gene in BGC 44.1 was identified by Blastx and antiSMASH analysis and often codes for a transporter of the ABC type.

The two clones carrying the BGCs 40.1 and 482.1 on fosmids (pCC1FOS_Cl40 and pCC1FOS_Cl482) were isolated from the library via serial dilution PCR. After fosmid isolation, the lasso peptide specific genes were confirmed by Sanger sequencing (Eurofins Genomics) using the sequencing primers pCC1FOS_Cl40_seq_1 - pCC1FOS_Cl40_seq_7 and pCC1FOS_Cl482_seq_1 - pCC1FOS_Cl482_seq_8 respectively (Table 10). The resulting overlapping sequences were aligned with the reference sequence using Geneious.

3.2.13 Amplification of a lasso peptide BGC from HMW metagenomic DNA and subsequent cloning into an expression vector

For the amplification of the metagenomic lasso peptide BGC 44.1, isolated HMW metagenomic DNA 3 was used. The specific primers Cl_44_SalI_fw and Cl_44_SacI_rv (Table 11) were used to amplify the lasso peptide specific genes A, B1, B2 and C of BGC 44.1 via PCR using the Q5 high-fidelity DNA polymerase kit. For that purpose, five reactions with varying metagenomic DNA template amounts (114 ng, 54 ng, 30 ng, 15 ng and 3 ng) were performed. After analysis of the amplicons via agarose gel electrophoresis, the five reactions were pooled, purified (Genomic DNA Clean & Concentrator-10 kit) and digested with SacI overnight. After purification of the digested amplicons, they were digested with SalI overnight and purified again. Using the T4 DNA Ligase the digested amplicons were ligated into the equally digested and purified expression vector pSK019 (Kunakom and Eustáquio 2020) to generate

pSK019_Cl44. The construct was transferred to *E. coli* DH10ß via electroporation and plasmids of resulting transformants were isolated and screened for the presence of BGC 44.1 by PCR using the primers Cl_44_SalI_fw and Cl_44_SacI_rv. 10 µl of PCR-positive plasmids were digested with PvuI in a 20 µl reaction overnight that was subsequently analyzed on an agarose gel to verify the expected fragments. In a next step, the D gene of BGC 44.1 was amplified from metagenomic DNA by PCR using the Q5 high-fidelity DNA polymerase kit and the primer pairs Cl44_D_gene_OV_fw/rv (Table 11). The primer pairs contained overhangs allowing for the subsequent assembly of the PCR product with pSK019_Cl44 thereby inverting the D genes to adjust the gene orientation. After analysis of the PCR products on an agarose gel, the five reactions were pooled, purified (Genomic DNA Clean & Concentrator-10 kit) and subsequently assembled with the Ecl136II digested and purified pSK019_Cl44 plasmid using the NEBuilder HiFi DNA Assembly Master Mix (NEB). The resulting construct pSK019_44_D was transferred to *E. coli* DH10ß and plasmids of transformants were isolated and verified via PCR using the primer pairs Cl44_D_gene_OV_fw/rv. The plasmids of three PCR-positive clones were digested with PvuI and 15 µl and 5 µl of each digested plasmid were analyzed on an agarose gel for verification of the expected fragments. A positive pSK019_Cl44_D plasmid containing the complete lasso peptide BGC 44.1 was confirmed by Sanger sequencing (Eurofins Genomics) using the sequencing primers PSK019_Cl44_D_seq_1 - PSK019_Cl44_D_seq_8 (Table 10) that generate sequences with overlap. The resulting sequences were aligned with the reference sequence using Geneious.

3.2.14 Integration of the *Burkholderia* replicon ori1600 into pCC1FOS_Cl40 and pCC1FOS_Cl482

A modified version of the protocol published in (Gust et al. 2003) was applied and is described in the following.

The primer pairs Lambda_red_pCC1FOS_fw/rv (Table 11) containing 39 nucleotide overhangs flanking the CHL resistance gene of pCC1FOS and the Q5 high-fidelity DNA polymerase kit were used to amplify a DNA fragment consisting of the KAN resistance, oriT and ori1600 gene (*Burkholderia* replicon) from the pSK019 vector via PCR. After analysis of the PCR products via agarose gel electrophoresis, the amplicons of four reactions were pooled, purified (QIAquick PCR purification Kit) and subsequently digested with DpnI overnight. The digestion reaction was purified and analyzed via agarose gel electrophoresis followed by gel purification. Electrocompetent cells of *E. coli* BW25113 carrying the lambda red plasmid pKD20

(BW25113/pKD20) (Datsenko and Wanner 2000) were generated as described in section 3.2.2 with ampicillin for selection and an adjusted culturing temperature of 30 °C as the plasmid had a temperature-sensitive origin of replication. Electrocompetent cells were transformed with pCC1FOS_Cl40 and pCC1FOS_Cl482 respectively and plated on LB agar (CHL/AMP) followed by incubation at 30 °C overnight, which resulted in the generation of the strains BW25113/pKD20/pCC1FOS_Cl40 and BW25113/pKD20/pCC1FOS_Cl482. The next steps aimed at exchanging the CHL resistance genes in both pCC1FOS constructs with the generated DNA fragment consisting of the KAN resistance, oriT and ori1600 gene to generate the plasmids pCC1FOS_Cl40_ori1600 and pCC1FOS_Cl482_ori1600 respectively. For that purpose, 500 µl of a BW25113/pKD20/pCC1FOS_Cl40 and 500 µl of a BW25113/pKD20/pCC1FOS_Cl482 overnight culture (30 °C) were used to inoculate 50 ml of LB medium (CHL/AMP) respectively, each supplemented with 500 µl of a 1 M L-arabinose solution for induction of the lambda red genes. The cultures were grown to an OD_{600} of approx. 0.6 at 30 °C with shaking followed by the addition of another 500 µl of a 1 M L-arabinose solution. After incubation for 30 more minutes, the cells were pelleted and made competent as described before. Subsequently, the cells were each transformed with 5 µl of the generated DNA fragment via electroporation, plated on LB agar (KAN) and incubated at 37 °C overnight for loss of pKD20. The plasmids of transformants were isolated by alkaline lysis with adding a phenol/chloroform extraction step in between. Subsequently, the plasmids were checked for a lambda red recombination event via PCR using the primer pairs Lambda_red_control_primer_fw/rv (Table 11) and amplicons were analyzed by agarose gel electrophoresis. The isolated plasmids consisted of mixtures with and without the lambda red modification. To select for plasmids with the lambda red modification only, *E. coli* DH10ß was transformed with each of the plasmid mixtures isolated from the BW25113 strains and subsequently plated on LB agar (KAN) for selection. Transformants only carrying the lambda red modified fosmids pCC1FOS_Cl40_ori1600 and pCC1FOS_Cl482_ori1600 were verified via PCR using the primer pairs Lambda_red_control_primer_fw/rv, followed by analysis of the PCR products via agarose gel electrophoresis.

3.2.15 Generation and transformation of electrocompetent *Burkholderia* sp. FERM BP-3421 cells

The generation and transformation of electrocompetent *Burkholderia* sp. FERM BP-3421 cells was conducted following a published protocol (Kunakom and Eustáquio 2020). Briefly, 50 ml

LB medium was inoculated with 1 ml of a *Burkholderia* sp. FERM BP-3421 overnight culture (LB medium, 30 °C) and the culture was grown to an OD_{600} of approx. 0.6 at 30 °C with shaking. The cells were pelleted by centrifugation (4000 rpm, 5 min at 4 °C) and the cell pellet was subsequently resuspended in 50 ml cooled sterile electroporation buffer (1 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), pH 7.0). The last step was repeated with resuspending the cell pellet in 25 ml of electroporation buffer. After pelleting the cells again, the supernatant was discarded and the cells were resuspended in the backflow. 50 µl of the cell suspension was mixed with 5 µl of the respective plasmid and the mixture was transferred to an electroporation cuvette (2 mm gap, VWR). Cells were electroporated using a MicroPulser electroporator (program EC2) and the cuvette was subsequently filled with 1 ml of LB medium. The mixture was transferred to a reaction tube and incubated for 1 hour at 30 °C with shaking. The cells were pelleted, resuspended in the backflow and plated on LB agar supplemented with 500 µg/ml kanamycin (KAN500) followed by incubation at 30 °C for 2-3 days.

3.2.16 *Burkholderia* colony PCR

Burkholderia transformants were streaked out on LB agar (KAN500) and grown for two days at 30 °C. Cell material of grown colonies was suspended in 20 µl of a solution consisting of 0.1% Tween 20 and 10 mM Trizma-Base (pH8) and incubated at 95 °C for 10 min. The suspension was subsequently centrifuged and 1 µl was used as template for a PCR screening using the primer pairs Lambda_red_control_primer_fw/rv and the Q5 High-Fidelity DNA Polymerase kit.

3.2.17 Culture conditions for lasso peptide BGC transcription analysis and heterologous expression experiments

The methods for heterologous expression experiments of BGCs 40.1 and 482.1 in *E. coli* as well as heterologous expression experiments of BGC 44.1 in *E. coli* and *Burkholderia* using M9 media were published in (Negri et al. 2022).

For heterologous expression experiments of BGCs 40.1 and 482.1 in *E. coli*, the fosmids carrying lasso peptide BGC 40.1 (pCC1FOS_Cl40) and lasso peptide BGC 482.1 (pCC1FOS_Cl482) were transferred to *E. coli* DH10ß via electroporation generating the strains DH10ß/pCC1FOS_Cl40 and DH10ß/pCC1FOS_Cl482. Three clones of each strain as well as

the control strain carrying the empty pCC1FOS vector (DH10ß/pCC1FOS) were inoculated from a glycerol stock in 5 ml LB medium (CHL) and the cultures were grown at 37 °C with shaking overnight. 1 ml of each preculture was used to inoculate 100 ml of M9 minimal media (CHL) respectively, followed by culturing at 37 °C with shaking. After three days, the culture supernatants were extracted with XAD-16 resin followed by butanol extraction, while the cell pellets were extracted with methanol (MeOH).

For heterologous expression experiments of BGCs 40.1 and 482.1 in *Burkholderia*, the lambda red modified fosmids carrying BGC 40.1 (pCC1FOS_Cl40_ori1600) and BGC 482.1 (pCC1FOS_Cl482_ori1600) were transferred to *Burkholderia* sp. FERM BP-3421 via electroporation generating the strains *Burkholderia*/pCC1FOS_Cl40_ori1600 and *Burkholderia*/pCC1FOS_Cl482_ori1600. Three clones of each strain (Clones 2, 10, 26 for *Burkholderia*/pCC1FOS_Cl40_ori1600 and clones 3, 6, 9 for *Burkholderia*/pCC1FOS_Cl482_ori1600) were inoculated from a glycerol stock in 100 ml LB medium (Kan500) respectively, while the *Burkholderia* control strain carrying no plasmid was inoculated in 100 ml LB medium without antibiotics. The cultures were grown at 30 °C with shaking and after two days a 1 ml aliquot for RNA isolation was taken and stored at -80 °C, followed by the extraction of entire cultures with ethyl acetate. Additional cultures were equally prepared and culture supernatants were extracted with XAD-16 resin, while cell pellets were extracted with MeOH. For heterologous expression experiments in 2S4G media, the BGC carrying strains were inoculated in 5 ml LB medium (Kan500) respectively and the empty *Burkholderia* strain was inoculated in 5 ml LB medium without antibiotics. The cultures were grown for two days at 30 °C with shaking and subsequently 1 ml of each preculture was used to inoculate 100 ml of 2S4G medium respectively (Kan500 for BGC carrying strains, no antibiotics for empty strain). After culturing for 5 days at 30 °C with shaking, the cultures were first extracted with ethyl acetate and subsequently with butanol. For heterologous expression experiments in M9 minimal medium, 1 ml of equally prepared precultures of the same strains were used to inoculate 100 ml of M9 minimal media respectively (Kan500 for BGC carrying strains, no antibiotics for empty strain). The 100 ml cultures were grown for three and five days at 30 °C with shaking, followed by extraction of the culture supernatants with XAD-16 resin and butanol, while the cell pellets were extracted with MeOH. Aliquots for RNA isolation were taken after three days and stored at -80 °C prior to culture extraction.

For heterologous expression experiments of BGC 44.1 in *E. coli*, the confirmed *E. coli* DH10ß clones 1, 3 and 5 carrying pSK019_Cl44_D (DH10ß/pSK019_Cl44_D) as well as the *E. coli* DH10ß control strain carrying the empty pSK019 vector (DH10ß/pSK019) were each

inoculated from a glycerol stock in 5 ml LB media (KAN) and the cultures were grown at 37 °C with shaking overnight. 1 ml of each preculture was used to inoculate 100 ml of M9 and M20 minimal media respectively each supplemented with KAN and 100 mM arabinose for induction of the PBAD promoter and the cultures were grown at 37 °C with shaking. After three days, the supernatants were first extracted with XAD-16 resin and subsequently with butanol. The pellets were frozen, thawed again and subsequently extracted with MeOH. For heterologous expression experiments with short induction time, the same strains were inoculated in 100 ml LB media (KAN) from 1 ml of equally prepared precultures and the cultures were grown to an OD_{600} of 0.2 at 37 °C with shaking. Subsequently, they were induced with 100 mM L-arabinose for three hours followed by taking an aliquot for RNA isolation and subsequent extraction of half of the cultures (XAD-16 resin/pellet MeOH extraction). The other halves were grown for an additional three days and subsequently equally extracted.

For heterologous expression experiments of BGC 44.1 in *Burkholderia*, the plasmid pSK019_Cl44_D was transferred to *Burkholderia* sp. FERM BP-3421 via electroporation generating the strain *Burkholderia*/pSK019_Cl44_D. Three confirmed clones as well as the control strain carrying the empty pSK019 vector (*Burkholderia*/pSK019) were each inoculated in 5 ml LB medium (KAN500) and the cultures were grown for two days at 30 °C with shaking. 1 ml of each preculture was used to inoculate 100 ml of M9 and M20 minimal media respectively each supplemented with KAN500 and 100 mM arabinose for induction of the PBAD promoter and the cultures were grown at 30 °C with shaking. 1 ml aliquots were taken after two and three days and stored at -80 °C for RNA isolation. After three days, the cultures were extracted equivalent to the procedure for the *E. coli* cultures. For transcription analysis with short induction time, the same strains were inoculated in 100 ml 2S4G medium (KAN500) from 1ml of equally prepared precultures and the cultures were grown for 6 hours at 30 °C with shaking. Subsequently, they were induced with 100 mM L-arabinose and grown overnight followed by taking an aliquot for RNA isolation.

3.2.18 Transcription analysis for *E. coli* and *Burkholderia* sp. FERM BP-3421 heterologous expression experiments

RNA was isolated from *E. coli* and *Burkholderia* sp. FERM BP-3421 cell pellets of 1 ml aliquots that were taken from various heterologous expression experiments using the Monarch Total RNA Miniprep Kit and following the manufacturer's instructions. Instead of the optional on-column DNase I treatment recommended by the manual, DNA was digested using the DNA-

free DNA Removal Kit following the manufacturer's instructions for rigorous DNase treatment with increasing the amount of rDNase I used to 3 µl. Subsequently, RNA was reverse transcribed to cDNA using the RevertAid RT Reverse Transcription Kit following the manufacturer's instructions. The generated cDNAs were used as templates in PCR reactions using the Taq DNA-Polymerase (Table 14/15) to check for the transcription of lasso peptide specific genes. For all PCR transcription analysis experiments RNAs of respective cDNAs were used as controls.

cDNA derived from *Burkholderia*/pCC1FOS_Cl482_ori1600 heterologous expression experiments and the primer pairs Cl482_A_tr_fw/rv, Cl482_B1_tr_fw/rv, Cl482_C_tr_fw/rv (Table 11) targeting the lasso peptide specific genes A, B1 and C were used to check for transcription of the respective genes.

For BGC 44.1 heterologous expression experiments in *E. coli* and *Burkholderia* the primer pairs Cl44_B1_tr_fw/rv, Cl44_B2_C_tr_fw/rv, Cl44_C_tr_fw/rv and Cl44_D_tr_fw/rv (Table 11) were used to check for transcription of the gene regions B1, B2 to C, C and D.

3.2.19 Assembly of a NRPS BGC from three different fosmids via TAR cloning

The methods of section 3.2.19 were published in (Negri et al. 2022) and are briefly summarized with the main steps in the following.

The complete NRPS BGC 76.1 was detected on a 138.907 bp contig upon antiSMASH analysis of the hybrid assembled metagenome sequencing data, while parts of it were detected on different contigs derived from the hybrid assembled fosmid library sequencing data. To TAR clone the BGC from the fosmid library, the contig derived from the metagenome sequencing that covered the complete cluster was used as a reference sequence. The latter enabled the design of the primer pairs NRPS_BGC_screen_left_fw/rv, NRPS_BGC_screen_middle_fw/rv and NRPS_BGC_screen_right_fw/rv (Table 11) targeting the left, middle and right part of the BGC. The primer pairs were used to screen the fosmid library for clones carrying respective cluster parts via PCR using the Taq DNA-Polymerase (Table 14/15) and positive clones were subsequently isolated from respective pools via serial dilution PCR. The isolated fosmids were end-sequenced from both sides (Eurofins Genomics) and the resulting sequences aligned (Geneious) with the reference sequence to confirm that the three fosmids covered the complete NRPS BGC with shared sequence overlap. The three fosmids were each digested with a unique restriction enzyme (DraI, Eco105I or PsiI) overnight to linearize them and release overlapping cluster parts, followed by phenol/chloroform extraction. TAR cloning was conducted using a

published protocol (Zhang et al. 2019) and applying a few modifications. Briefly, the BGC 76.1 specific 144 bp dsDNA fragment was synthesized (IDT), and subsequently amplified by PCR using the Q5 high-fidelity DNA polymerase kit and the primer pairs Caparm_fw/rv (Table 11). After Gibson Assembly of the dsDNA fragment with the XhoI/NdeI digested pCAP03 vector, the resulting capture vector was cloned in *E. coli* DH10ß and subsequently digested with PmeI. The digested capture vector and the three digested fosmids were used for TAR cloning in yeast to generate pCAP03_Cl76. Resulting yeast colonies were screened via colony PCR using the Q5 high-fidelity DNA polymerase kit and the same three primer pairs targeting the left, middle and right part of BGC 76.1 (Table 11) that were used for fosmid library screening. Plasmids of positive yeast colonies were isolated (Zymoprep Yeast Plasmid Miniprep I kit) and transferred to *E. coli* DH10ß by electroporation. After isolation of a PCR-positive plasmid from *E. coli*, residual host genomic DNA contamination was degraded (Exonuclease V) and the plasmid was purified using the Genomic DNA Clean & Concentrator-10 kit. For Nanopore sequencing the plasmid was digested (SmiI) overnight to generate two linear DNA fragments that were subsequently purified. The two fragments were sequenced by the NCCT on a PromethION device using a PromethION flow cell.

The resulting Nanopore reads were filtered for size ranges (15.5 kb – 16.5 kb and 56 kb – 58 kb respectively) matching the two fragments and subsequently aligned with the SmiI digested pCAP03_Cl76 reference sequence using Tablet (version 1.19.09.03) (alignment performed by Shrikant Mantri).

3.2.20 Generation of spore suspensions of *Streptomyces* species

The respective *Streptomyces* strains were grown in baffled Erlenmeyer flasks containing a steel spring and TSB medium at 27 °C with shaking. After two days, 1 ml of the respective culture was spread on Cullum agar and incubated for 7-9 days at 30 °C until sporulation occurred. 10 ml of sterile water was added to each plate and the spores were scraped off to generate a suspension. The suspension was transferred to a falcon tube and vortexed for 1 min. Subsequently, the spore suspension was passed through sterile cotton that was plugged into a syringe to separate the mycelium from the spores. The spores were pelleted by centrifugation (2100 x g, 10 min, 4 °C), resuspended in 1 ml of 15% glycerol and stored at -80 °C until use.

3.2.21 Triparental conjugation for the introduction of pCAP03_Cl76 and the empty pCAP03 vector into *Streptomyces* strains

Triparental conjugation was conducted following a modified version of the procedure described in the TAR cloning protocol (Zhang et al. 2019).

E. coli DH10ß carrying pCAP03_Cl76 (DH10ß/pCAP03_Cl76) and *E. coli* ET12567/pUB307 carrying the helper plasmid for conjugation (ET12567/pUB307) were each inoculated in 50 ml LB medium (KAN) and the cultures were grown at 37 °C with shaking overnight. 100 µl of the DH10ß/pCAP03_Cl76 and 50 µl of the ET12567/pUB307 overnight cultures were used to inoculate 12 ml of LB medium (KAN) and the cultures were grown to an OD_{600} of 0.4-0.6 at 37 °C with shaking. Following, the cells were each washed twice with 20 ml of LB medium and subsequently resuspended in 1 ml of LB medium. In parallel, for each conjugation 50 µl of *Streptomyces* spores were added to 200 µl of TSB medium and heat shocked for 10 min at 50 °C followed by cooling. 200 µl of each *E. coli* suspension was mixed with the heat shocked spores. 250 µl and 450 µl of the resulting mixture were plated on two Cullum agar plates supplemented with 10 mM calcium chloride and incubated for 16-20 hours at 30 °C. Subsequently, the plates were each overlaid with 1 ml of water containing 25 µl of nalidixic acid (NAL) and 25 µl of KAN and incubated for 5 days at 30 °C. Exconjugant colonies were transferred to Cullum agar containing NAL and KAN and were grown for three to five days at 30 °C. Spore suspensions of individual exconjugants were prepared. For the introduction of the empty pCAP03 vector into *Streptomyces*, *E. coli* DH10ß carrying pCAP03 (DH10ß/pCAP03) was used.

3.2.22 Isolation of genomic DNA from Actinomycetes

The respective strain was grown in a baffled Erlenmeyer flask containing a steel spring and 20 ml of TSB medium supplemented with the appropriate antibiotic at 27 °C with shaking overnight. 10 ml of the culture was pelleted by centrifugation and subsequently resuspended in 900 µl of Solution 1. 100 µl of 30 mg/ml lysozyme dissolved in solution 1 was added and the mixture was incubated for 1 hour at 37 °C with shaking. Subsequently, the following components were added: RNase A to a final concentration of 0.1 mg/ml, 20 µl of 20 mg/ml proteinase K and finally 60 µl of 15 % sodium dodecyl sulfate. The tube was incubated for 1 hour at 37 °C with shaking, followed by incubation at 50 °C for 30 min. 600 µl of phenol:chloroform:isoamyl alcohol was added and the tube was vortexed for 30 seconds. After

centrifugation for 5 min at maximum speed, the upper phase was transferred to a new tube and mixed with 700 µl of isopropanol followed by several inversions of the tube. The supernatant was removed from the precipitated DNA that was subsequently washed with 1 ml of 70 % ethanol. After removal of the ethanol, the washed DNA was solved in 300 µl of water.

3.2.23 Culture conditions for NRPS BGC 76.1 heterologous expression experiments

Spores of three *S. coelicolor* exconjugants carrying the NRPS BGC 76.1 (pCAP03_Cl76) as well spores of a *S. coelicolor* control strain carrying the empty pCAP03 vector were each grown in baffled Erlenmeyer flasks containing a steel spring and 100 ml R5 media (KAN) and 100 ml M8 minimal media (KAN) respectively. The R5 cultures were grown for seven days and the M8 cultures were grown for nine days at 27 °C with shaking. Aliquots of BGC carrying strains grown in R5 media were taken after two and seven days for genomic DNA isolation and subsequent PCR to check for the presence of the three cluster parts (left, middle, right). After seven days (R5 medium) and nine days (M8), half of the culture supernatants were extracted with ethyl acetate and subsequently with butanol and the other halves were extracted with XAD-16 resin. The pellets were frozen at -80 °C, thawed again and subsequently extracted with 50% methanol/50% acetone.

E. coli GB05-MtaA (Fu et al. 2012) was transformed with pCAP03_Cl76 and the empty pCAP03 vector respectively. Three clones of *E. coli* GB05-MtaA carrying NRPS BGC 76.1 (pCAP03_Cl76) as well as a control strain carrying the empty pCAP03 vector were inoculated in 100 ml LB media (KAN) and 100 ml M9 media (KAN) respectively each additionally supplemented with gentamycin (GEN) for PPtase gene selection and the cultures were grown at 30 °C with shaking. After three days, half of the supernatants were extracted with XAD-16 resin and the other halves were extracted with ethyl acetate. The pellets were frozen at -80 °C, thawed again and subsequently extracted with MeOH.

3.2.24 Extraction of cultures and subsequent analysis of extracts via high-performance liquid chromatography-mass spectrometry (HPLC-MS)

The methods for the extraction of cultures with XAD-16 resin and the analysis of the extracts were published in (Negri et al. 2022).

For extractions of culture supernatants and entire cultures with ethyl acetate and butanol, the following procedure was applied. An equal amount of the respective solvent was added to the

respective culture contained in an Erlenmeyer flask. The flask was sealed using parafilm and shaken for 30 min. Subsequently, the solvent was separated from the aqueous phase using a separation funnel. In cases where this method of separation was not sufficient, the separation was conducted by centrifugation (10 min, 4000 rpm) of the mixture in falcon tubes and subsequent collection of the resulting upper phases. The solvent was transferred to a round-bottom flask, followed by drying at 37 °C and reduced pressure using a rotary evaporator. The dried extract was solved in 2 ml of methanol.

For extraction of cultures with XAD-16 resin (Amberlite), the following procedure was applied. XAD-16 resin was washed with acetone followed by washing twice with water prior to its use. The respective culture was centrifuged for 20 min at 4000 rpm to separate the cells from the culture medium. The cell pellet was frozen at -80 °C, thawed again and subsequently extracted with 50 ml MeOH and shaking overnight. After centrifugation (10 min, 4000 rpm), the MeOH was collected in a round-bottom flask, followed by drying at 37 °C and reduced pressure using a rotary evaporator. The supernatant of the respective culture was mixed with 15 ml of a XAD-16/water suspension and incubated for 1 hour with shaking. Subsequently, the supernatant was removed, followed by washing of the XAD-16 with water. After removal of the water, the XAD-16 was extracted with 100 ml of MeOH and shaking for 30 min. The MeOH was subsequently collected in a round-bottom flask, followed by drying at 37 °C and reduced pressure using a rotary evaporator. All dried lasso peptide extracts were solved in 50 % MeOH, while other extracts were solved in 100 % MeOH.

Actinomycete pellets were extracted with 50 % methanol/50 % acetone and following the procedure described for pellet extractions.

All extracts were analyzed on a HPLC 1260 Infinity device coupled to an InfinityLab LC/MSD mass spectrometry device. For HPLC a Kinetex 5 µm C18 100 Å LC column (100 x 4.6 mm) was used. A gradient (10-100 %) of acetonitrile (0.1 % formic acid) in water (0.1 % formic acid) was applied over 20 min with a flow rate of 1 ml/min. The mass spectrometer was configured to positive ion mode with a mass range of 100-2000 Da.

Generated data were analyzed using the data analysis tool of the LC/MSD ChemStation software (Agilent Technologies).

4. Aim of the study

The aim of this study consisted of developing an efficient approach for the identification and recovery of novel natural product BGCs from soil metagenomes. Furthermore, recovered metagenomic BGCs were tested in various heterologous expression experiments aiming at the heterologous production of novel compounds.

The overall aim of this study consisted of facilitating and accelerating future soil metagenomic natural product discovery efforts that can potentially lead to new therapeutically relevant compounds such as antibiotics.

5. Results

5.1 Assessing the biosynthetic potential of different soil types from the Schönbuch forest

The main part of section 5.1 was published in (Mantri et al. 2021).

The Schönbuch forest nature reserve (Germany) harbors the three different soil types podzol, cambisol and stagnosol in close proximity to each other, which ensures that they are subjected to the same environmental conditions. Therefore, the Schönbuch forest provides an optimal study area for comparing the biosynthetic potential of different soils. Each of the soil types consists of multiple horizons with different characteristics. The organic layer as well as the A horizon were sampled from each of the three types and the B horizon was additionally sampled for cambisol soil. In order to investigate the soil types and horizons for their biosynthetic potential, metagenomic DNA was isolated from each sample and verified via agarose gel electrophoresis (Fig. 6), followed by Illumina shotgun sequencing.

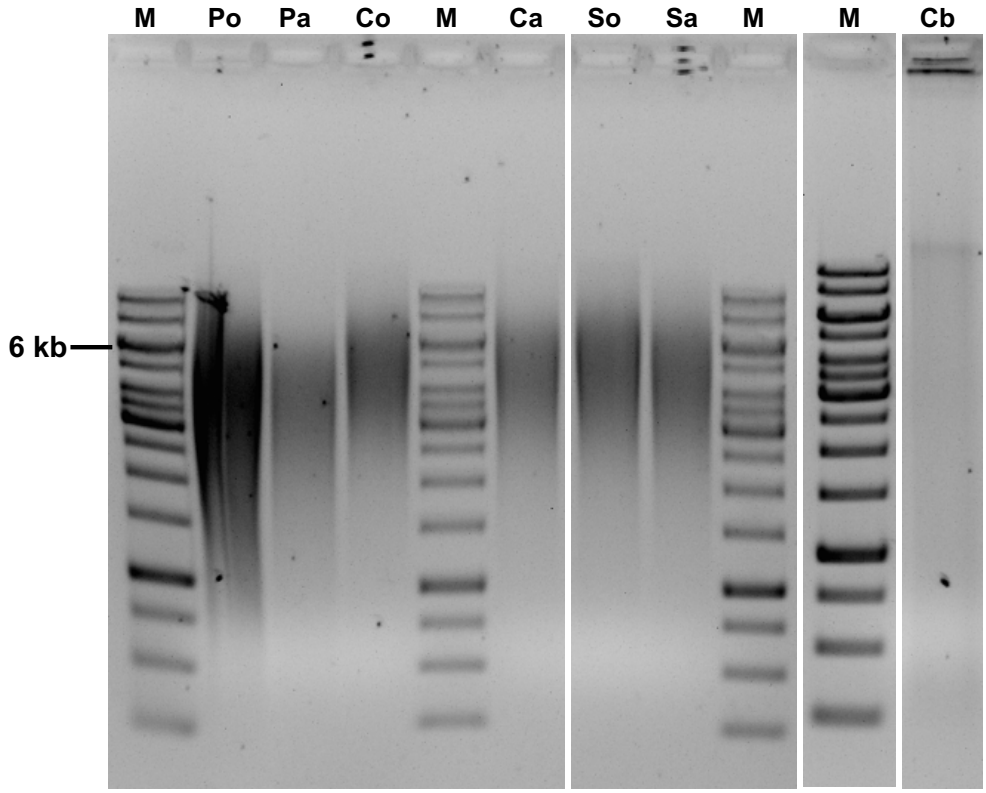


Fig. 6 Agarose gel electrophoresis of metagenomic DNA isolated from different soil types and horizons. Lane Po, podzol organic layer; lane Pa, podzol A horizon; lane Co, cambisol organic layer; lane Ca, cambisol A horizon; lane Cb, cambisol B horizon; lane So, stagnosol O layer; lane Sa, stagnosol A horizon; lane M, GeneRuler 1kb DNA Ladder

Additionally, PCR was performed using degenerate primers to generate A domain amplicons of NRPSs and KS domain amplicons of PKSs that served as indicators for biosynthetic diversity. Amplicons for each soil type and horizon were verified on an agarose gel (Fig. 7), followed by Illumina amplicon sequencing.

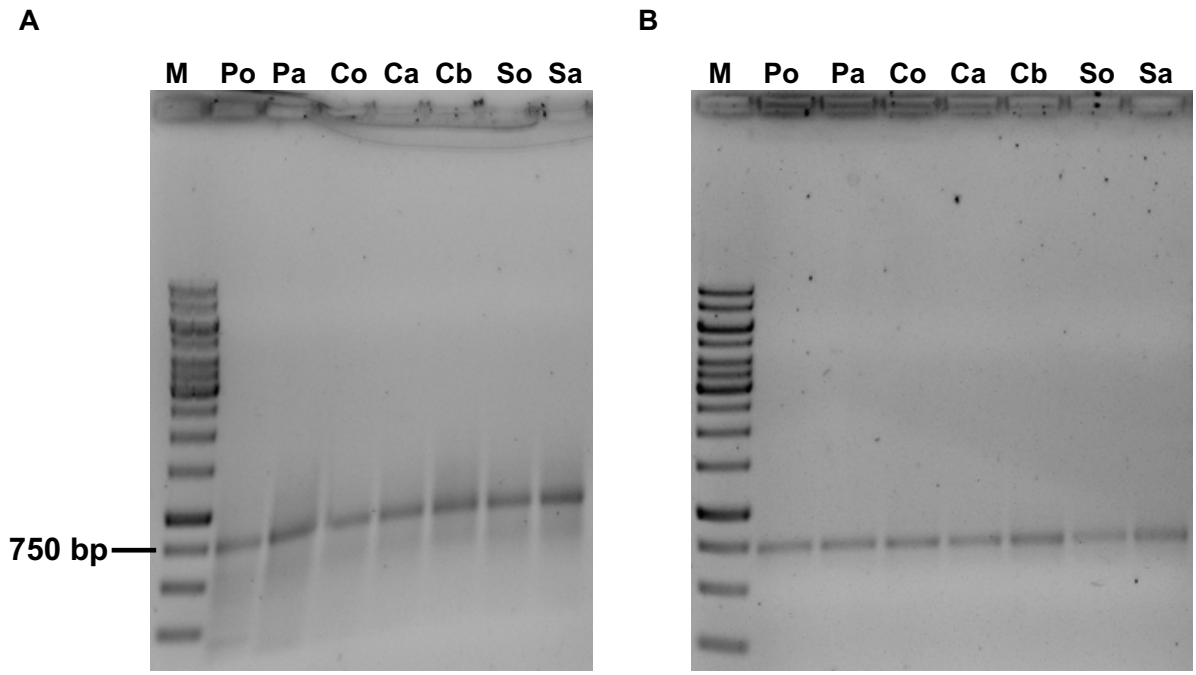


Fig. 7 Agarose gel electrophoresis of purified A domain (A) and KS domain (B) amplicons. Lanes Po, podzol organic layer; lanes Pa, podzol A horizon; lanes Co, cambisol organic layer; lanes Ca, cambisol A horizon; lanes Cb, cambisol B horizon; lanes So, stagnosol O layer; lanes Sa, stagnosol A horizon; lanes M, GeneRuler 1kb DNA Ladder

Generated amplicon sequences pertaining to each of the three soil types were clustered into operational biosynthetic units (OBUs) to generate groups of specific BGCs and rarefaction curves were generated revealing that all soils harbored a large biosynthetic potential that could not be captured completely by amplicon sequencing as the rarefaction curves did not reach saturation (Fig. 8) (analysis performed by Shrikant Mantri, data for comparison between horizons not shown). The soil types cambisol and stagnosol showed a higher number of OBUs with increasing sequencing depth for both A domains and KS domains as compared to podzol.

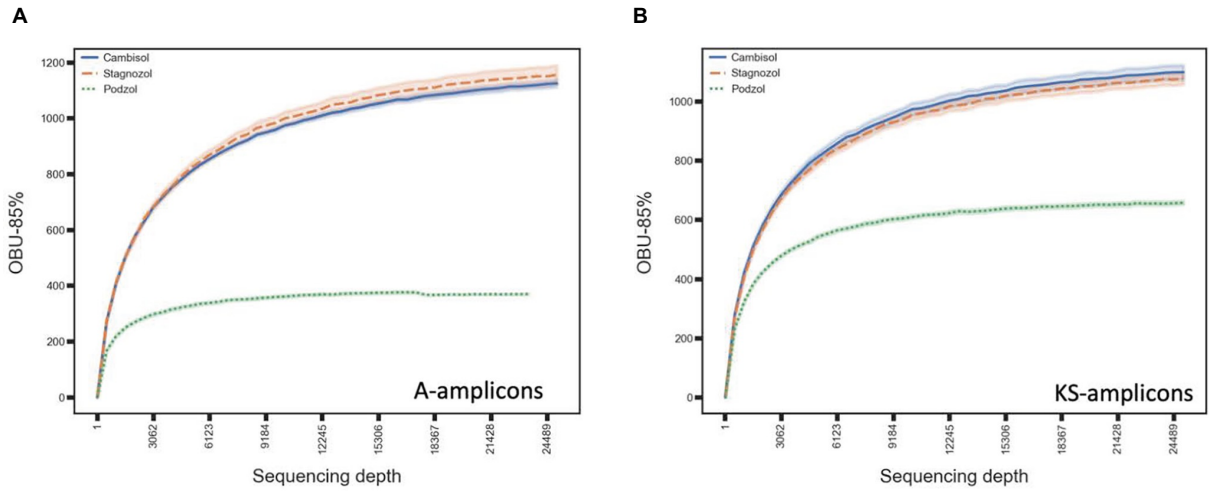


Fig. 8 Rarefaction curves for A domain amplicons (A) and KS domain amplicons (B). Mean value of operational biosynthetic units (OBUs) at a specific sequencing depth for all horizons of a soil type; modified from (Mantri et al. 2021).

Previous experience from DNA isolations of the different soil types has shown that the isolation of pure DNA from stagnosol is more challenging as compared to the other two types because of the higher clay content. Therefore, cambisol soil was chosen for the isolation of HMW DNA and the subsequent steps associated with BGC recovery.

Analysis of KS and A domains detected via Illumina shotgun sequencing revealed that more than 90% of these could not be detected by amplicon sequencing, which further confirmed the high biosynthetic potential of all sampled soils (data not shown).

5.2 Isolation and sequencing of high-quality, high molecular weight (HMW) metagenomic DNA from soil

The isolation of HMW metagenomic DNA as well as sequencing and hybrid assembly described in this section were published in (Mantri et al. 2021) and (Negri et al. 2022).

Accessing the biosynthetic potential of soils requires the isolation of high-quality, HMW metagenomic DNA from soil, which is crucial for the recovery and heterologous expression of novel metagenomic BGCs. For this study, the metagenomic DNA needed to be of sufficient quality for the downstream applications used, i.e. fosmid library generation, long- and short-read sequencing and direct amplification of complete BGCs via PCR. Within this thesis, different optimized protocols were developed to isolate high-quality HMW metagenomic DNA from the A horizon of cambisol soil that was sampled from the Schönbuch forest. The first isolated DNA was analyzed on an agarose gel, which revealed a HMW band migrating slightly below the lambda DNA/HindIII 27 kb band with only little smear below (Fig. 9A), confirming that the DNA was of a high molecular weight and showed only little shearing. 260/280 and 260/230 absorbance ratios (metagenomic DNA 1, Table 16) confirmed sufficient purity and the DNA was sequenced via Nanopore sequencing. For Illumina sequencing, where purity is of greater importance than DNA size, the same DNA was purified using a specific DNA column and subsequently analyzed on an agarose gel (Fig. 9B), which revealed an intense smear indicating sheared but highly concentrated DNA. 260/280 and 260/230 absorbance ratios (metagenomic DNA 2, Table 16) confirmed very high purity and the DNA was sequenced using Illumina. To further improve the HMW metagenomic DNA quality, a different optimized protocol was developed and used to isolate DNA from the same soil sample. Analysis of the DNA on an agarose gel revealed an intense band migrating above the lambda DNA/HindIII 23 kb band with almost no smear below (Fig. 9C) (Negri et al. 2022), which confirmed that the DNA was of a high molecular weight and showed minimal shearing. 260/280 and 260/230

absorbance ratios (metagenomic DNA 3, Table 16) confirmed high purity and the DNA was sequenced via Nanopore. Long-read Nanopore as well as short-read Illumina sequencing data were both used in a hybrid assembly approach to generate contigs (assembly performed by Shrikant Mantri).

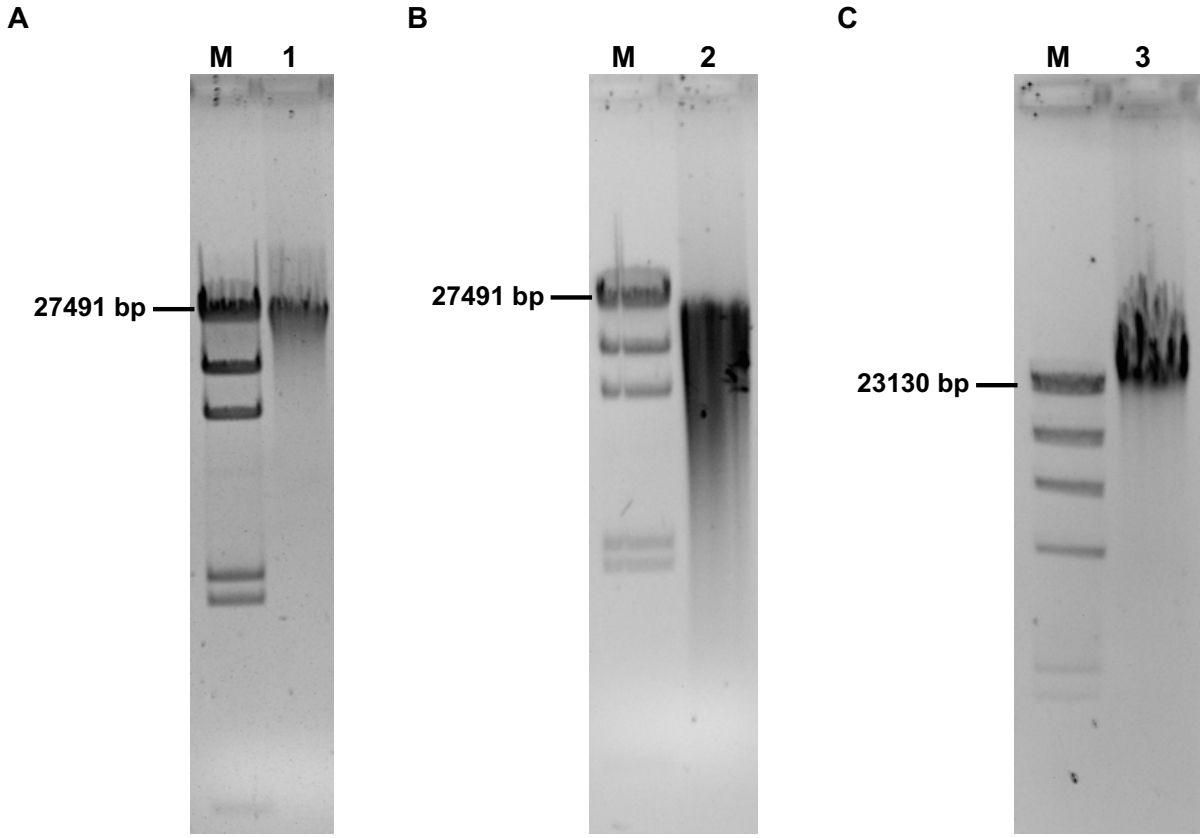


Fig. 9 Agarose gel electrophoresis of HMW metagenomic DNA from soil. A) Lane 1, HMW metagenomic DNA isolated from cambisol soil from the Schönbuch forest. **B)** Lane 2, same metagenomic DNA column purified. **C)** Lane 3, HMW metagenomic DNA isolated from the same soil sample of the Schönbuch forest using an alternative isolation method. Lanes M, Lambda DNA/HindIII Marker; partly adopted from (Negri et al. 2022)

Table 16 260/280 and 260/230 absorbance ratios of isolated metagenomic DNAs. Metagenomic DNA 1, first isolated HMW DNA; metagenomic DNA 2, same DNA column purified; metagenomic DNA 3, HMW DNA isolated from the same soil sample using an alternative isolation method

Sample	A260/280	A260/230
metagenomic DNA 1	1.72	1.14
metagenomic DNA 2	1.87	2.03
metagenomic DNA 3	1.86	1.67

The main part of sections 5.3 to 5.5.2 was published in (Negri et al. 2022).

5.3 Generation and sequencing of a metagenomic fosmid library

To capture and access the biosynthetic diversity contained within the metagenomic DNA, a metagenomic fosmid library consisting of approximately 83700 clones stored in pools of 2000 clones was generated (Fig. 11A) using mainly metagenomic DNA 3 as it showed the highest molecular weight. To assess the quality of the library, the fosmids of 10 randomly picked clones were isolated and their inserts released via restriction digestion. Subsequent analysis on an agarose gel revealed that the fosmids captured large inserts (Fig. 10).

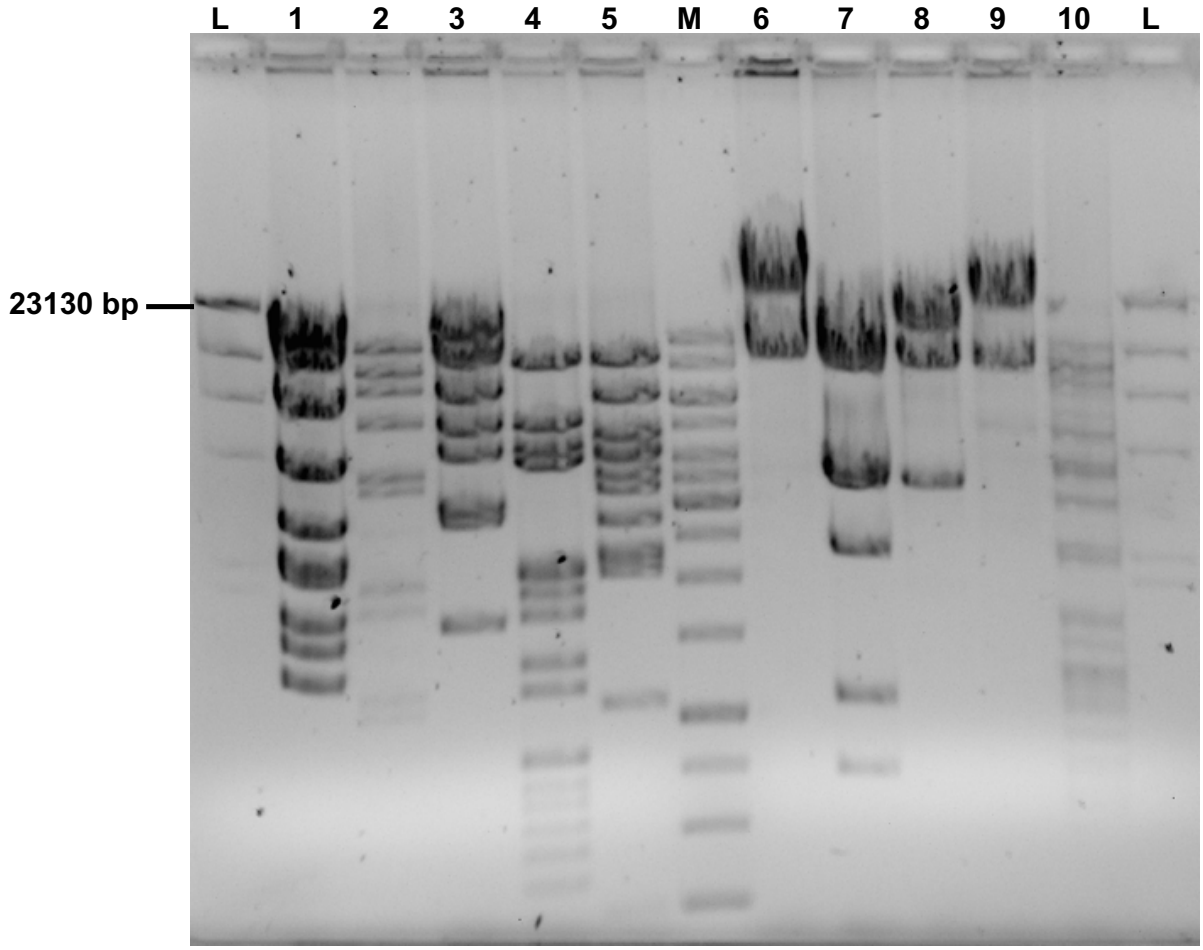


Fig. 10 Agarose gel electrophoresis of 10 randomly isolated and digested fosmids. Lanes 1-10, digested fosmids of 10 random clones; lane L, Lambda DNA/HindIII Marker; lane M, GeneRuler 1kb DNA Ladder

In a next step, all pools of the library were combined and total fosmid DNA was isolated and sequenced via Illumina and Nanopore technologies. Nanopore sequencing generated more than 367 thousand long-reads greater than 25 kb in size, which confirmed the high quality of the sequencing data. Both long- and short-reads were used in a hybrid assembly approach to enable the generation of large contigs (Fig. 11B), which was achieved with the assembly of almost 16

thousand contigs greater than 25 kb in size (assembly performed by Shrikant Mantri). Sequence data are summarized in Supplemental Table S1.

5.4 Identification of a large number of natural product BGCs contained within the fosmid library and its corresponding metagenome

To assess the abundance of BGCs present within the fosmid library, all hybrid assembled contigs greater than 1 kb were analyzed with antiSMASH (Blin et al. 2019), which revealed the presence of 2019 BGC regions.

A large part of the detected BGCs consisted of only partial clusters, which were annotated by antiSMASH as “region on contig edge”. Since the aim consisted of recovering complete BGCs for subsequent heterologous expression experiments, contigs greater than 40 kb in size were filtered to select for clusters with a higher probability of being completely present on a single contig (Fig. 11C). 100 BGCs remained after the filtering process and 98 of those revealed to be unknown as no significant similarity to any characterized BGC within the MIBiG database (Kautsar et al. 2019) was found. Out of these BGCs, 34 were annotated as being complete and were predicted to encode for compounds of various natural product classes: one heterocyst glycolipid synthase-like PKS (hglE-KS), one aryl polyene, one linear azol(in)e-containing peptide (LAP), nine terpenes, one lanthipeptide, five bacteriocins, five type 3 PKSs, one indole/terpene, one betalcatone, two NRPSs and seven lasso peptides. More than 100 additional complete BGCs were detected upon antiSMASH analysis of contigs greater than 25 kb, which were available as potential candidates for recovery and heterologous expression. However, these BGCs were not considered for the following steps as already a sufficient amount of complete BGCs to choose from were detected on contigs greater than 40 kb.

Additionally, hybrid assembled contigs greater than 25 kb derived from the direct sequencing of the corresponding metagenomic DNA were analyzed with antiSMASH leading to the detection of 113 BGCs.

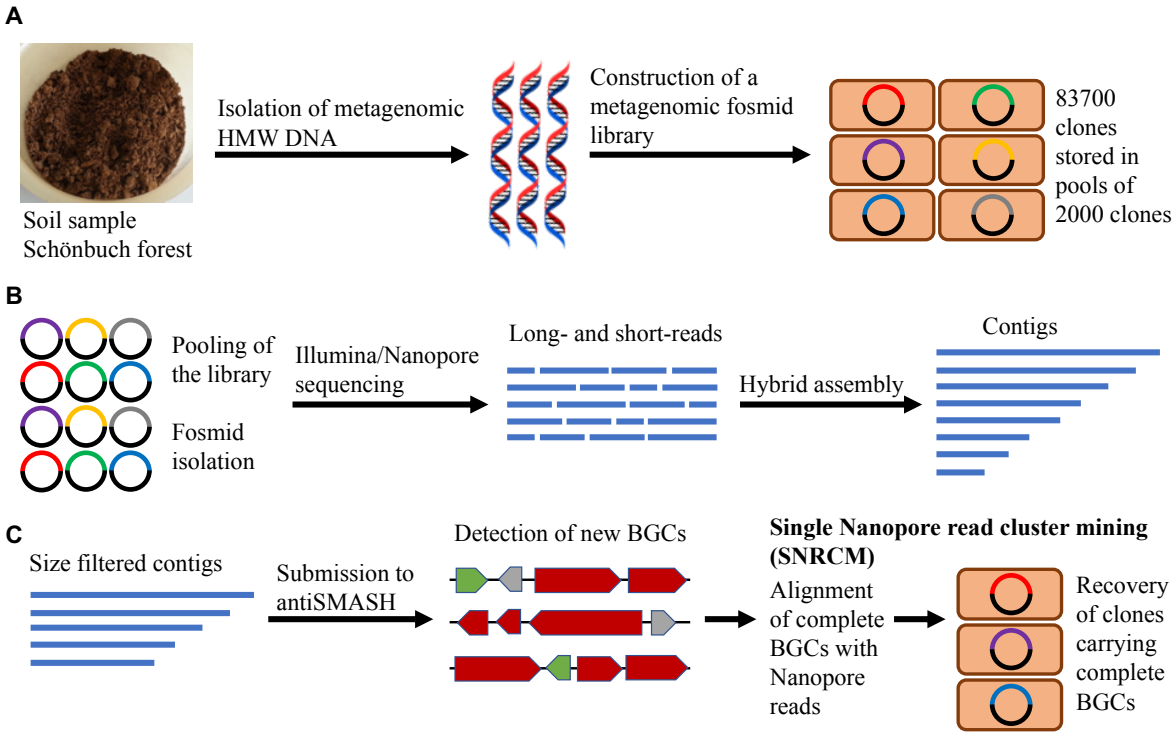


Fig. 11 Workflow for capturing metagenomic BGCs ready for heterologous expression. A) Isolation of high-quality HMW DNA from soil and subsequent construction of a metagenomic fosmid library consisting of approx. 83700 clones stored in pools of 2000 clones. **B)** Pooling of the library and isolation of fosmids for subsequent Illumina/Nanopore sequencing. Hybrid assembly of short- and long-reads. **C)** Size filtering of contigs greater than 40 kb and subsequent submission to antiSMASH for BGC detection. Alignment of complete BGCs with Nanopore reads for identification of fosmids harboring a complete BGC. Recovery of positive clones by serial dilution PCR; modified from (Negri et al. 2022)

5.5 Fast and efficient recovery of complete BGCs from a fosmid library and HMW metagenomic DNA and subsequent heterologous expression experiments

5.5.1 Identification and recovery of completely captured BGCs on single fosmids

The 34 complete BGCs that were detected on hybrid assembled fosmid library contigs greater than 40 kb could each potentially have been distributed over distinct fosmids, which would prevent the direct recovery of complete BGCs. Therefore, an efficient strategy called single Nanopore read cluster mining (SNRCM) (Negri et al. 2022) was developed that takes advantage of the generated high-quality long-read Nanopore sequencing data to enable the fast identification and recovery of complete BGCs captured on a single fosmid. Applying SNRCM, the 34 complete BGCs on contigs were directly aligned with the single Nanopore reads (Fig. 11C). 15 BGCs revealed to be present on a single fosmid as each of the BGCs aligned completely with a single Nanopore read that is generally derived from one fosmid molecule only. These BGCs were predicted to encode four terpenes, three lasso peptides, one betalactone, one LAP, one lanthipeptide and five bacteriocins. Since antimicrobial activity is the most common bioactivity of known lasso peptides, metagenomic BGCs encoding these molecules

on a single fosmid were prioritized for recovery. Therefore, the two fosmids carrying the complete lasso peptide BGCs 40.1 (Fig. 12A) and 482.1 (Fig. 12B) respectively were selected and isolated from the library via serial dilution PCR.

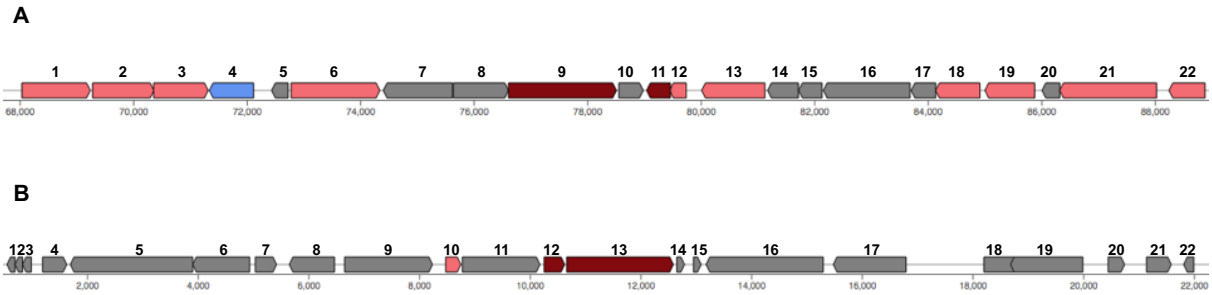


Fig. 12 antiSMASH output of lasso peptide BGCs 40.1 (A) and 482.1 (B) with numbered genes

Both lasso peptide BGCs were investigated for completeness by checking for the presence of the minimal set of the lasso peptide biosynthesis gene homologues A, B and C. The B2 and C genes could directly be identified using Blastx and antiSMASH analysis (Table 17, S2). To identify the putative B1 genes, the Blastx analysis results, gene sizes and antiSMASH annotations (Table 17, S2) were compared to known lasso peptide B1 genes (MIBiG) (Table 18, S3). This analysis revealed for both of the metagenomic BGCs a gene with high concurrence for the three compared parameters, which identified the putative B1 genes.

Table 17 Identification of putative lasso peptide genes of BGCs 40.1 and 482.1 using Blastx results, gene sizes and antiSMASH annotations. Blastx results, gene sizes and antiSMASH annotations with high similarities to known B1 genes highlighted in blue; modified from (Negri et al. 2022)

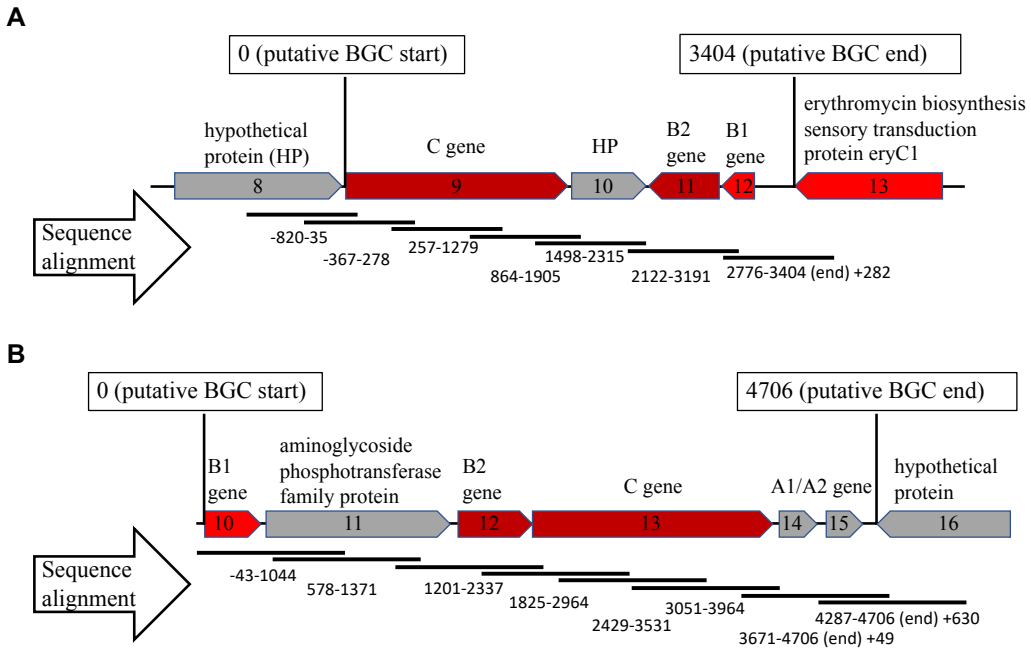
Metagenomic lasso peptide BGC 40.1				
Gene	**Blastx results**	**Gene size (bp)**	**antiSMASH annotation**	**Putative lasso peptide gene**
8	hypothetical protein [*Acidobacteriia bacterium*]	975	-	-
9	hypothetical protein [*Acidobacteriia bacterium*]	1905	Asn_synthase	C gene
10	hypothetical protein [*Betaproteobacteria bacterium*]	432	-	-
11	lasso peptide biosynthesis B2 protein [*Acidobacteriia bacterium*]	423	PF13471	B2 gene
12	PqqD family peptide modification chaperone [*Acidobacteriia bacterium*]	294	PF05402	B1 gene
13	erythromycin biosynthesis sensory transduction protein eryC1 [*Acidobacteria bacterium*]	1116	DegT_DnrJ_EryC1	-
Metagenomic lasso peptide BGC 482.1				
Gene	**Blastx results**	**Gene size (bp)**	**antiSMASH annotation**	**Putative lasso peptide gene**
10.1	hypothetical protein DMG37_22385 [*Acidobacteria bacterium*]	276	PF05402	B1 gene
10.2	PqqD family protein [*Acidobacteriia bacterium*]			
11	aminoglycoside phosphotransferase family protein [*Acidobacteriia bacterium*]	1404	-	Kinase
12	lasso peptide biosynthesis B2 protein [*Acidobacteria bacterium*]	375	PF13471	B2 gene
13	hypothetical protein [*Acidobacteriia bacterium*]	1929	Asn_synthase	C gene
14	hypothetical protein [*Acidobacteria bacterium*]	141	-	A1 gene
15	hypothetical protein [*Acidobacteria bacterium*]	141	-	A2 gene
16	hypothetical protein DMG78_32005 [*Acidobacteria bacterium*]	2118	-	-

Table 18 Exemplary Blastx results, gene sizes and antiSMASH annotations of known B1 genes. For each B1 gene the first two Blastx results are listed. Blastx results, gene sizes and antiSMASH annotations with high similarities to metagenomic lasso peptide BGC genes highlighted in blue; modified from (Negri et al. 2022)

BGC0001176: streptomonomicin biosynthetic gene cluster from *Streptomonospora alba*			
Gene	**Blastx results**	**Gene size (bp)**	**antiSMASH annotation**
B1	hypothetical protein LP52_05070 [*Streptomonospora alba*]	255	PF05402
	PqqD family peptide modification chaperone [*Streptomonospora alba*]		
BGC0001356: paeninodin biosynthetic gene cluster from *Paenibacillus dendritiformis C454*			
Gene	**Blastx results**	**Gene size (bp)**	**antiSMASH annotation**
B1	lasso peptide biosynthesis PqqD family chaperone [*Paenibacillus dendritiformis*]	300	PF05402
	lasso peptide biosynthesis PqqD family chaperone [*Paenibacillus dendritiformis*]		

No A gene coding for the precursor peptide could be detected by antiSMASH analysis in either of the BGCs, which is often the case for lasso peptide BGCs due to the small size of the gene. Manual inspection of BGC 482.1 led to the identification of two putative A genes (14 and 15) that were similar in size to known A genes. Additionally, their translated amino acid sequences (Fig. 13C) contained the common features of a lasso peptide (Maksimov et al. 2012): Both of the amino acid sequences contained a glycine with the necessary distance to an aspartate for ring formation. The remaining amino acids located right to the ring forming aspartate had the appropriate length to form the lasso peptide tail. The ones located left to the ring forming glycine would be cleaved from the precursor peptide. A threonine that is often found at position -2 relative to the core peptide can be replaced with amino acids similar in size (Pan et al. 2012). The leucine and serine predicted for the two precursor peptides of BGC 482.1 would be suitable candidates for such a replacement.

Resequencing of the detected putative lasso peptide specific genes on both isolated fosmids confirmed the recovery of two complete BGCs each captured on a single fosmid (Fig. 13A and B). Investigation of known lasso peptide BGCs revealed that the A gene is often located in close proximity to the other lasso peptide specific genes. Therefore, the surrounding part of the putative lasso peptide genes of BGC 40.1 was investigated, which led to the detection of various candidates. However, none of them could be determined as a promising precursor.



C

Gene 14, 141 bp

Amino acid sequence: MKYTKPEVVVLDNAISVVQMSKLRGVKDNADPTSELPSIAAYQSDE

Gene 15, 141 bp

Amino acid sequence: MKYTKPEVVVLDNAISVVQLSKSVGPHDSVDPGSMFHTPAAYQADE

Fig. 13 Sequencing confirmation for lasso peptide BGCs 40.1 and 482.1. Sequencing of lasso peptide BGC 40.1 (**A**) and 482.1 (**B**) using specific sequencing primers that generate sequences with overlap. Numbers of the alignment refer to the start and end point of each generated sequence that matches the reference sequence. Gene annotation via bioinformatics analysis as described. **C) Translated nucleotide sequences of genes 14 and 15 of lasso peptide BGC 482.1.** Most suitable amino acid candidates for ring formation in bold and red. Amino acids building the putative ring with grey shade. Amino acids similar in size to a commonly found threonine at position -2 relative to the core peptide underlined; modified from (Negri et al. 2022)

5.5.2 Heterologous expression experiments of lasso peptide BGCs 40.1 and 482.1 in *E. coli*

The next steps aimed at the heterologous expression of the metagenomic lasso peptide BGCs to obtain the encoded compounds. *E. coli* is known to be capable of lasso peptide production as exemplified by the strains producing the well-known lasso peptide microcin J25 (Bayro et al. 2003). Therefore, *E. coli* provided the most rapid and easy option for a first heterologous expression attempt as no modification of the fosmids was required for that purpose.

The fosmids carrying lasso peptide BGC 40.1 and lasso peptide BGC 482.1 were transferred to *E. coli* DH10ß respectively. Three clones of each BGC carrying strain as well as the control strain carrying the empty pCC1FOS vector were cultured in M9 minimal media to test for heterologous expression of the encoded lasso peptides. After three days, the culture supernatants and cell pellets were extracted separately and the extracts were analyzed via high-

performance liquid chromatography-mass spectrometry (HPLC-MS). The resulting mass chromatograms of BGC carrying strains and the empty vector control strain were compared for differences, which would indicate the presence of an additional compound. No differences could be observed upon analysis of the chromatograms (exemplary mass chromatograms are shown for later extractions, which are representative for lasso peptide extraction experiments conducted in this thesis).

5.5.3 Introduction of the *Burkholderia* replicon ori1600 into pCC1FOS_Cl40 and pCC1FOS_Cl482 and subsequent transfer to *Burkholderia* sp. FERM BP-3421

As heterologous expression experiments with *E. coli* did not lead to the detection of a lasso peptide, another heterologous host was tested for the ability to overcome possible limitations that prevented metagenomic lasso peptide production in *E. coli*.

Burkholderia sp. FERM BP-3421 was expected to be a promising host as another species of the genus *Burkholderia* is already known for native lasso peptide production. Furthermore, FERM BP-3421 has been shown to be able to heterologously express that same *Burkholderia* derived lasso peptide BGC in higher amounts than *E. coli* (Kunakom and Eustáquio 2020). The pCC1FOS vector used for fosmid library generation only allows for replication and maintenance of the fosmids in *E. coli*. In order to use the lasso peptide BGCs 40.1 and 482.1 on fosmids (pCC1FOS_Cl40 and pCC1FOS_Cl482) for heterologous expression in *Burkholderia* FERM BP-3421, the *Burkholderia* replicon (ori1600) needed to be integrated into the vector backbone. Therefore, lambda red recombination experiments were conducted that aimed at exchanging the chloramphenicol (CHL) resistance gene present on pCC1FOS with a DNA fragment amplified from the pSK019 vector that consisted of the *Burkholderia* replicon, a kanamycin (KAN) resistance gene for selection and an oriT gene lying between the genes of interest of the used template (Fig. 14).

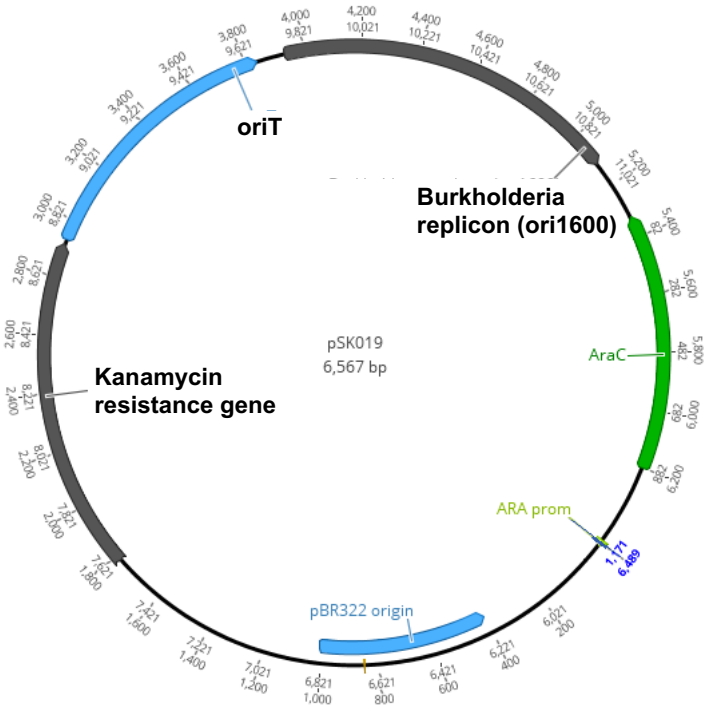


Fig. 14 Plasmid map of pSK019. Highlighted are kanamycin resistance, oriT and *Burkholderia* replicon (ori1600) genes that were amplified for the lambda red recombination experiment

Primers with overhangs flanking the CHL resistance gene of pCC1FOS from both sides were designed to amplify the *Burkholderia* replicon together with the KAN resistance and oriT gene from the pSK019 vector via PCR (3434 bp). In a next step, the template plasmid had to be removed from the generated amplicons as it would have generated false positive clones in the later step of transferring the linear DNA fragment. Therefore, the PCR reaction was digested with DpnI that only cuts methylated recognition sites and therefore only residual plasmid. The amplicons were subsequently purified and verified on an agarose gel (Fig. 15A), followed by a final gel purification.

In a next step, the *E. coli* lambda red strain BW25113/pKD20 was transformed with pCC1FOS_Cl40 and pCC1FOS_Cl482 respectively. The resulting strains were subsequently transformed with the purified linear DNA fragment containing the *Burkholderia* replicon for lambda red recombination to generate the plasmids pCC1FOS_Cl40_ori1600 and pCC1FOS_Cl482_ori1600 respectively. Individual clones were screened for each of the two plasmids using PCR with primers binding to the flanking regions of the CHL resistance gene. In case of no recombination event where the CHL resistance gene is still present, a PCR product of 933 bp was expected. In case of a recombination event where the CHL resistance gene was replaced with the *Burkholderia* replicon, oriT and KAN resistance gene, a PCR product of 3635 bp was expected. Analysis of the amplicons on an agarose gel revealed that both fragments were amplified from plasmids of all except one clone, which indicated that the majority of

clones carried a mixture of plasmids consisting of plasmids with and without a lambda red recombination modification (Fig. 15B and C).

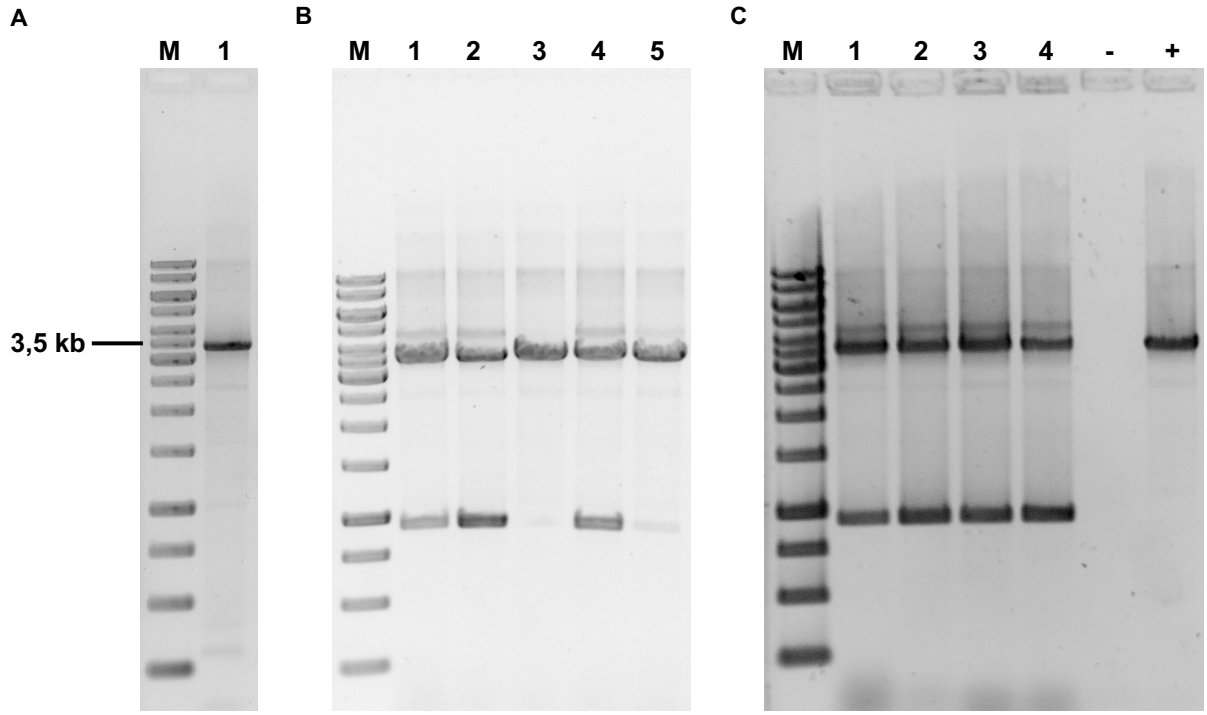


Fig. 15 A) Agarose gel electrophoresis of PCR products resulting from the *Burkholderia* replicon amplification. Lane 1, amplified *Burkholderia* replicon after DpnI digestion and purification. **B) Agarose gel electrophoresis of PCR products resulting from the screening of *E. coli* for a lambda red recombination event for pCC1FOS_Cl40.** Lanes 1-5, pCC1FOS_Cl40 fosmids of individual clones after lambda red recombination as templates. **C) Agarose gel electrophoresis of PCR products resulting from the screening of *E. coli* for a lambda red recombination event for pCC1FOS_Cl482.** Lanes 1-4, pCC1FOS_Cl482 fosmids of individual clones after lambda red recombination as templates; lane -, negative control with water (H_2O) instead of template; lane +, positive control with pSK019 plasmid as template. Lanes M, GeneRuler 1kb DNA Ladder

To select for plasmids with the lambda red modification, the mixtures were used to transform *E. coli* DH10ß with subsequent selection on LB agar (KAN). Plasmids of two transformants for each of the two modified fosmids were isolated and checked via PCR. Analysis of the PCR products on an agarose gel verified that the strains only harbored the lambda red modified fosmids pCC1FOS_Cl40_ori1600 (Fig. 16A) and pCC1FOS_Cl482_ori1600 (Fig. 16B) respectively as only the 3635 bp amplicon was detectable.

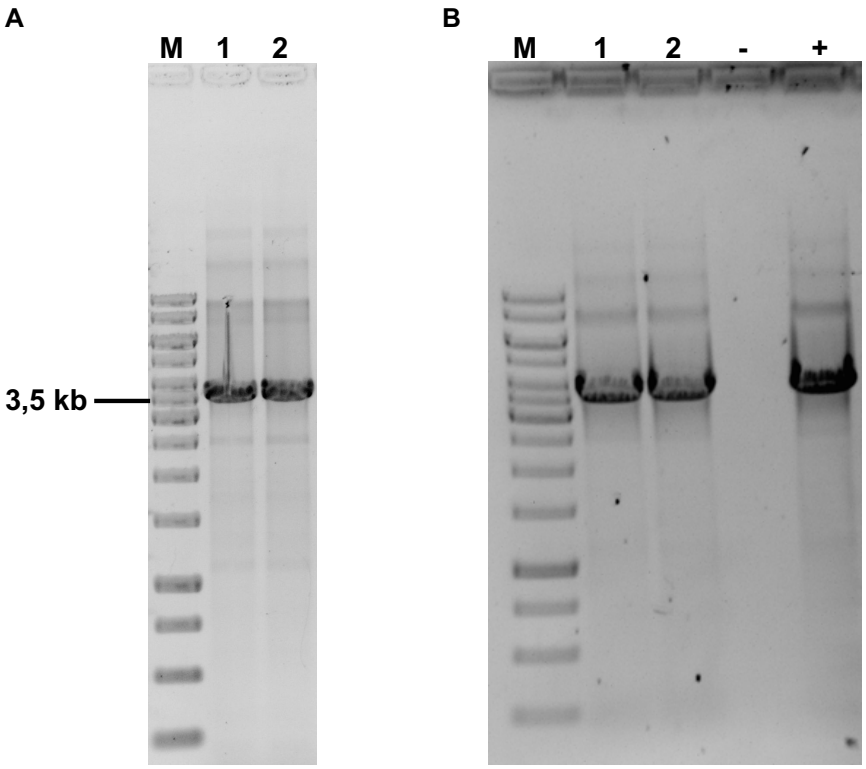


Fig. 16 Agarose gel electrophoresis of PCR products resulting from the screening of *E. coli* for clones only carrying the lambda red modified fosmids pCC1FOS_Cl40_ori1600 (A) and pCC1FOS_Cl482_ori1600 (B) respectively. A) Lanes 1-2, fosmids isolated from two individual clones as templates that were transformed with the pCC1FOS_Cl40 lambda red modified fosmid mixture. **B)** Lanes 1-2, fosmids isolated from two individual clones as templates that were transformed with the pCC1FOS_Cl482 lambda red modified fosmid mixture; lane -, negative control with H_2O instead of template; lane +, positive control with pSK019 plasmid as template. Lanes M, GeneRuler 1kb DNA Ladder

The fosmids pCC1FOS_Cl40_ori1600 and pCC1FOS_Cl482_ori1600 both carrying the *Burkholderia* replicon were transferred to *Burkholderia* sp. FERM BP-3421. Selected transformants that had grown on LB agar (KAN500) were screened for the presence of the modified fosmids via colony PCR. Analysis of the PCR products on an agarose gel revealed several positive transformants for both fosmids as the 3635 bp amplicon was detectable (Fig. 17A and B).

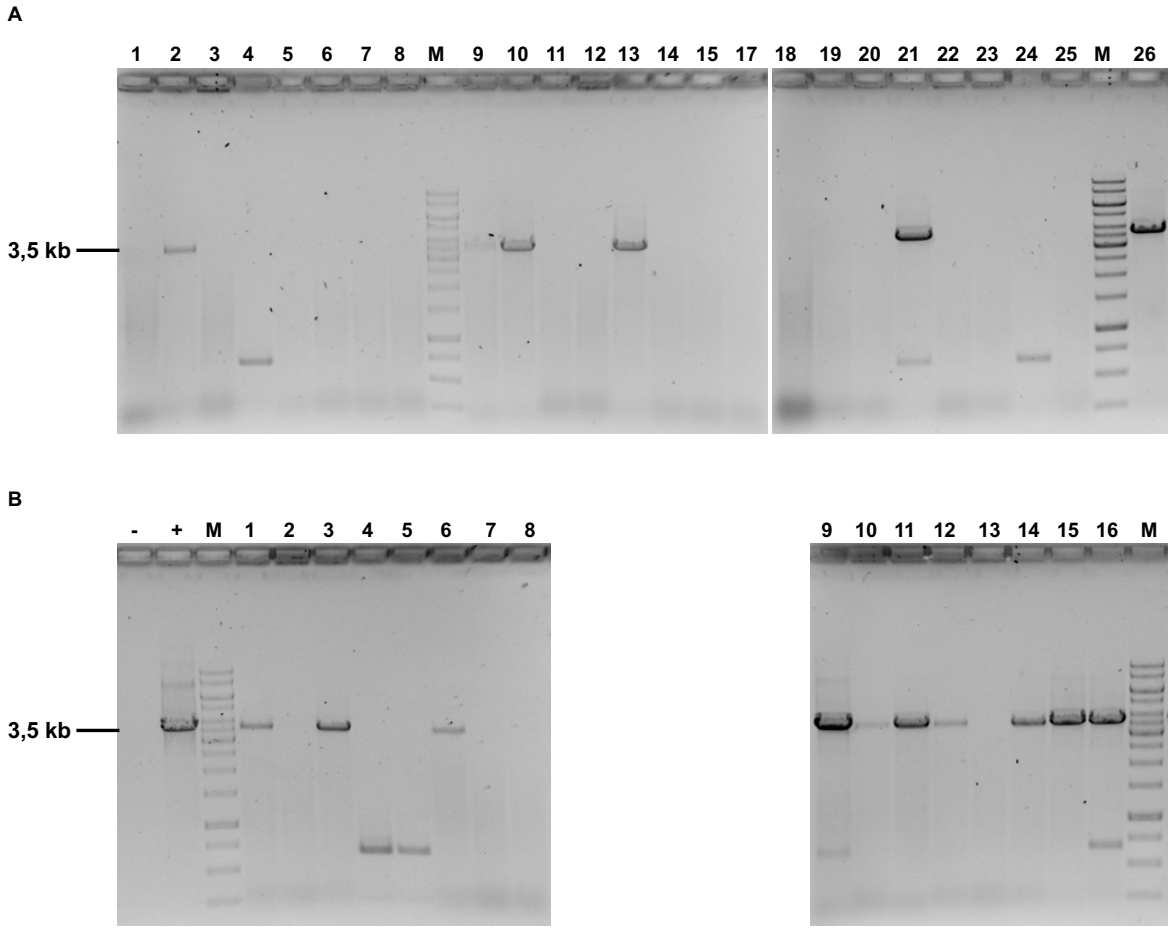


Fig. 17 Agarose gel electrophoresis of colony PCR products resulting from the screening of *Burkholderia* clones transformed with pCC1FOS_Cl40_ori1600 (A) and pCC1FOS_Cl482_ori1600 (B). A) Lanes 1-15 and 17-26, individual clones of *Burkholderia* transformed with pCC1FOS_Cl40_ori1600 as templates. **B)** Lanes 1-16, individual clones of *Burkholderia* transformed with pCC1FOS_Cl482_ori1600 as templates; lane -, negative control with H_2O instead of template; lane +, positive control with pSK019 plasmid as template. Lanes M, GeneRuler 1kb DNA Ladder

5.5.4 Heterologous expression experiments of lasso peptide BGCs 40.1 and 482.1 in *Burkholderia* sp. FERM BP-3421

Three confirmed clones of *Burkholderia* carrying pCC1FOS_Cl40_ori1600 as well as three clones of *Burkholderia* carrying pCC1FOS_Cl482_ori1600 were used for the following heterologous expression experiments, while the empty *Burkholderia* strain was used as a control.

In a first attempt, the strains were grown for two days in LB media. Different extraction methods were applied to extract both entire cultures as well as cultures that have been separated into supernatant and cell pellet before. Analysis of the extracts via HPLC-MS did not lead to the detection of any lasso peptide.

Therefore, in a second attempt the same strains were grown in 2S4G media, which is used as spliceostatin production medium and has previously been shown to be appropriate for the

induced heterologous production of a lasso peptide using the same strain (Kunakom and Eustáquio 2020). After 5 days, entire cultures were extracted using different solvents. Analysis of the extracts via HPLC-MS did not reveal the production of any lasso peptide either.

5.5.5 Transcription analysis and further heterologous expression experiments

Since no lasso peptide production could be detected, RNA was isolated from aliquots that were taken from LB cultures prior to extraction to check if the lasso peptide BGCs were transcribed. Isolated RNA was reverse transcribed to cDNA and in a first step PCR was performed using primers targeting the C genes of the lasso peptide BGCs 40.1 and 482.1 respectively. Since PCR products could only be detected for the C gene of BGC 482.1 but not for BGC 40.1 (data not shown), only cDNA resulting from the three strains carrying BGC 482.1 was used to check for transcription of further lasso peptide genes A and B1 via PCR. Analysis of the PCR products via agarose gel electrophoresis showed amplicons for each investigated gene confirming their transcription in each of the three BGC carrying clones (Fig. 18).

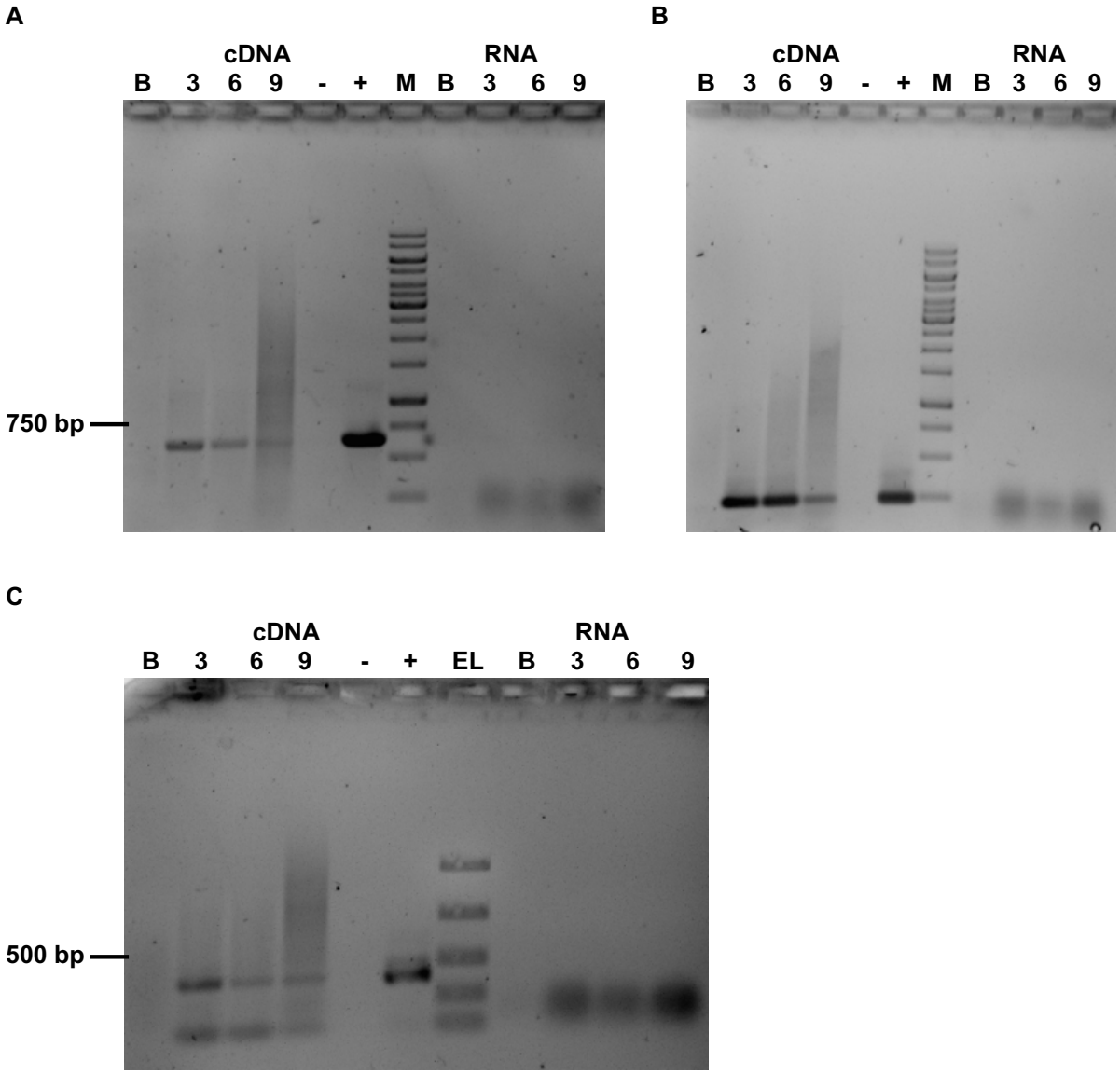


Fig. 18 Agarose gel electrophoresis of PCR products resulting from BGC 482.1 transcription analysis of *Burkholderia* LB cultures. PCR products for amplification of a C gene part (**A**), a B1 gene part (**B**) and an A gene part (**C**) using cDNA and RNA as template. Lanes B, isolated RNA and respective cDNA from empty *Burkholderia* FERM BP-3421 strain as template; lanes 3, 6 and 9, isolated RNA and respective cDNA from *Burkholderia*/pCC1FOS_Cl482_ori1600 clones 3, 6 and 9 as templates; lanes -, negative controls with H_2O instead of template; lanes +, positive controls with pCC1FOS_Cl482 as template; lanes M, GeneRuler 1kb DNA Ladder; lane EL, EasyLadder I

As transcripts of the analyzed lasso peptide specific genes were at least found for BGC 482.1 but no lasso peptide could be detected, it was hypothesized that in complex medium such as LB the lasso peptide might have been covered by media components, especially if the lasso peptide was produced in very low amounts. Therefore, further heterologous expression experiments were conducted in minimal medium to avoid media components covering a potentially produced lasso peptide.

The same *Burkholderia* strains carrying pCC1FOS_Cl40_ori1600 and pCC1FOS_Cl482_ori1600 respectively as well as the empty control strain were grown in M9 minimal media. After three days, RNA was isolated and reverse transcribed to cDNA. Initially,

PCR was performed using primers targeting the C genes of both BGCs. As a PCR product could again only be detected for BGC 482.1, the additional lasso peptide specific genes B1 and A were only checked for the same BGC. Analysis of the PCR products on an agarose gel revealed that all genes were transcribed in clones three and six (Fig. 19A and B). Although very light bands could be observed for the RNA controls of B1 and A gene amplification, the much stronger bands for cDNA indicated that the genes were transcribed.

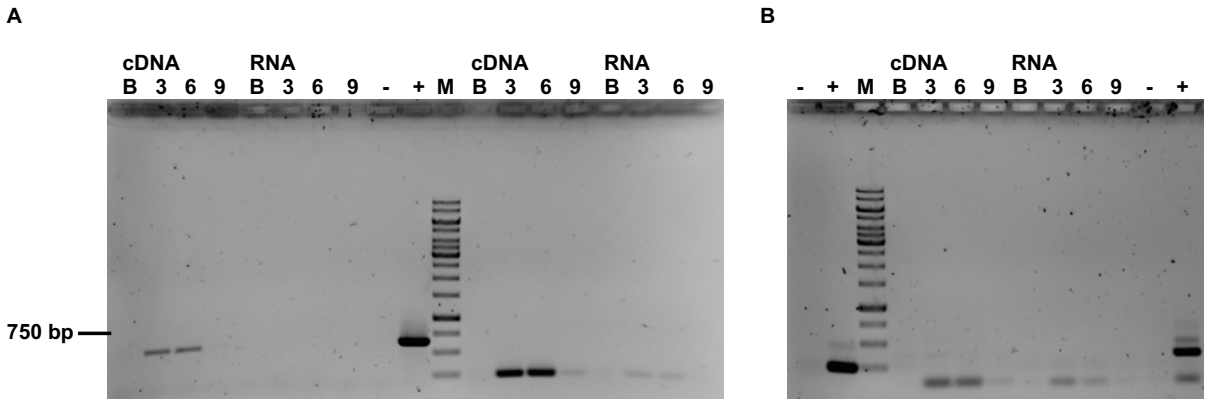


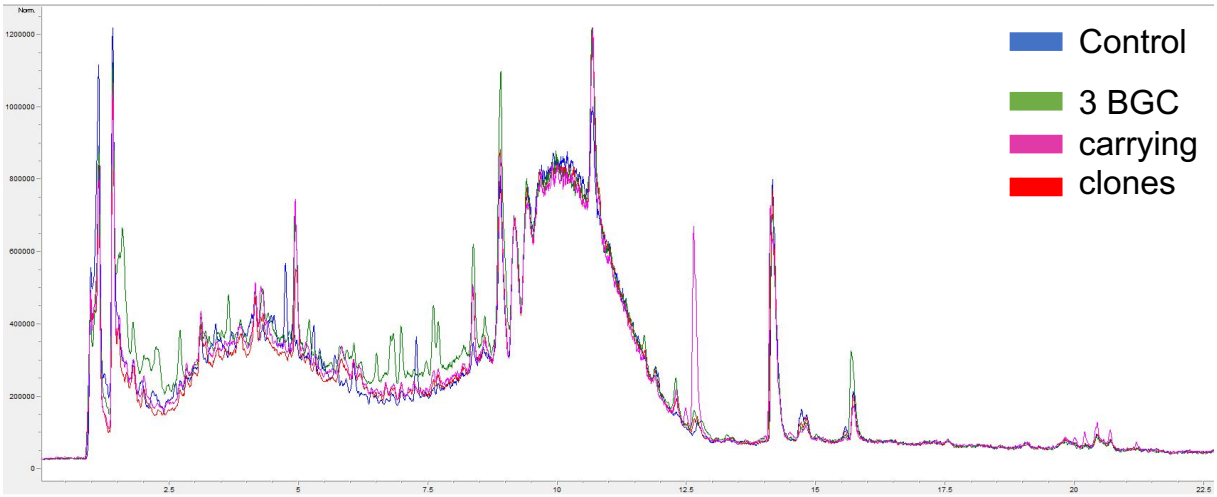
Fig. 19 Agarose gel electrophoresis of PCR products resulting from BGC 482.1 transcription analysis of *Burkholderia* M9 cultures. PCR products for amplification of a C gene part (**A,** gel part left from marker), a B1 gene part (**A**, gel part right from marker) and an A gene part (**B**) using cDNA and RNA as template. Lanes B, isolated RNA and respective cDNA from empty *Burkholderia* FERM BP-3421 strains as template; Lanes 3, 6 and 9, isolated RNA and respective cDNA from *Burkholderia*/pCC1FOS_Cl482_ori1600 clones 3, 6 and 9 as templates; lanes -, negative controls with H_2O instead of template; lanes +, positive controls with pCC1FOS_Cl482 as template; lanes M, GeneRuler 1kb DNA Ladder

Although transcripts could also only be detected for *Burkholderia*/pCC1FOS_Cl482_ori1600 but not for *Burkholderia*/pCC1FOS_Cl40_ori1600 M9 cultures, both were nevertheless investigated for the production of a lasso peptide. The culture supernatants and cell pellets of the 3-day M9 cultures were extracted separately and the extracts were analyzed via HPLC-MS, which did not lead to the detection of respective lasso peptides.

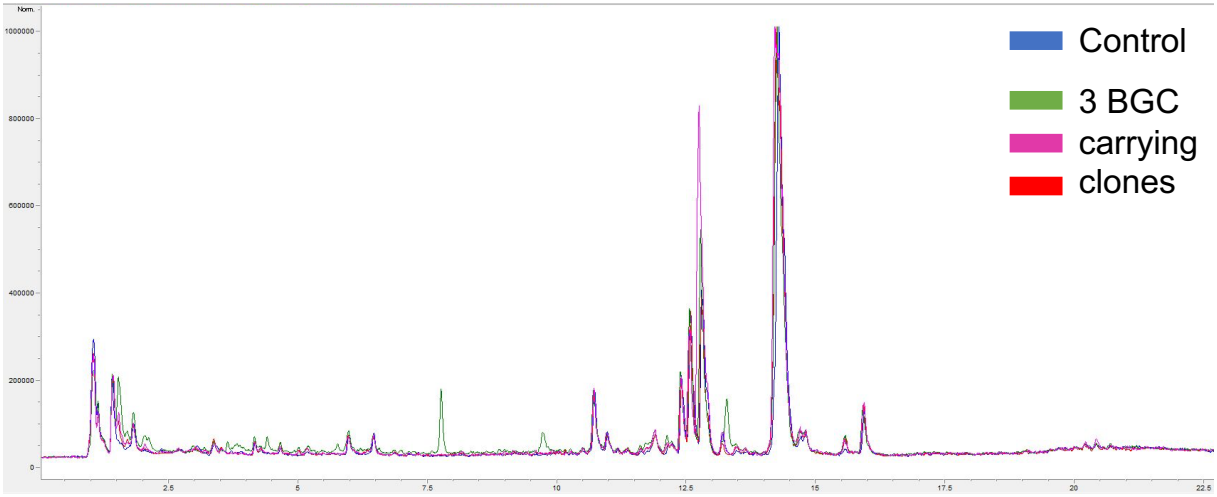
Prolongation of the M9 culturing time to five days followed by XAD-16 resin and butanol extraction of the supernatants as well as MeOH extraction of the pellets did not lead to the detection of a lasso peptide either. In Fig. 20 the respective mass chromatograms of BGC 482.1 extracts are exemplarily shown, which are representative for the aforementioned expression experiments. Additional peaks that were only observed in BGC carrying clones but not in the empty control strain were each investigated manually for corresponding masses. The masses causing the additional peaks could be excluded as potential lasso peptides for one or both of the following reasons: i) the respective mass was also found in the control although with a lower signal intensity, ii) the mass was singly charged and far too small for a lasso peptide. The strong

signal (red) for the pellet extraction revealed to be a contamination derived from the HPLC-MS device (Fig. 20 C).

A Five-day M9 cultures extracted with XAD-16 resin



B Five-day M9 cultures extracted with butanol



C Five-day M9 cultures, pellet MeOH extraction

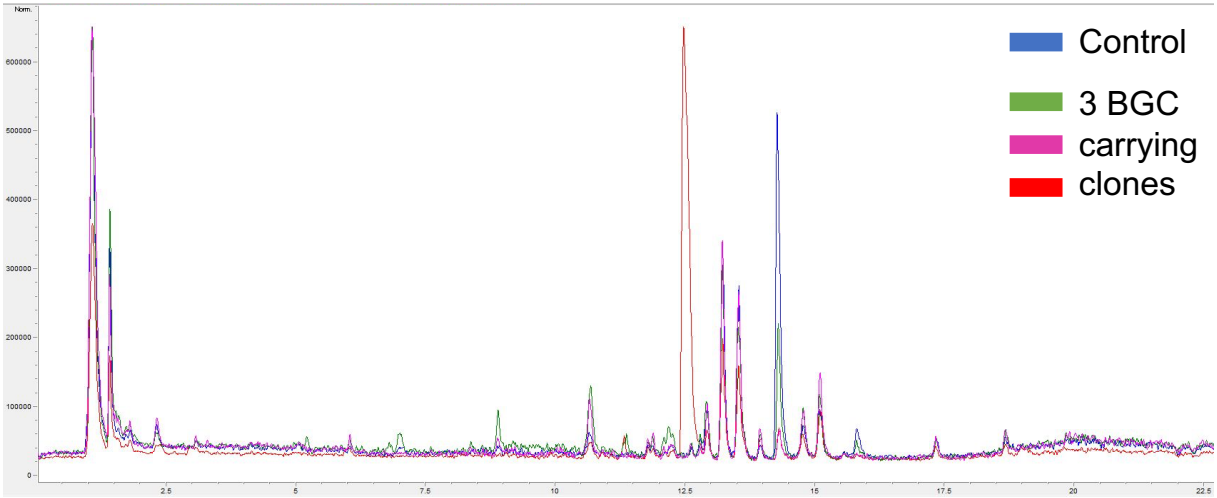


Fig. 20 HPLC-MS mass chromatograms of extracts derived from *Burkholderia*/pCC1FOS_Cl482_ori1600 and FERM BP-3421 control cultures. Five-day M9 cultures extracted with XAD-16 resin **(A)**, butanol **(B)** and pellet MeOH extraction **(C)**. Blue, FERM BP-3421 control culture extracts; remaining colors, extracts of the three BGC carrying clones

The same strains grown for three days on LB, 2S4G and M9 agar followed by butanol extraction did also not lead to the detection of a lasso peptide (data not shown).

The main part of sections 5.5.6 to 5.5.10 was published in (Negri et al. 2022).

5.5.6 Using HMW metagenomic DNA for the direct amplification of BGCs via PCR and subsequent cloning into expression vectors

Isolated high-quality HMW DNA of sufficient purity can be used to access BGCs by direct amplification via PCR for subsequent cloning into expression vectors. The 113 BGCs detected on hybrid assembled contigs greater than 25 kb derived from the metagenome sequencing were investigated for BGCs suitable for amplification. The lasso peptide BGC 44.1 (Fig. 21) was chosen and the putative lasso peptide specific genes A, B1, B2, C and D were identified using bioinformatics methods as described before (Table 19, S4). The BGC proved to be a suitable candidate for amplification as except for gene D all genes showed the same orientation, which allowed their cloning in front of a promoter.

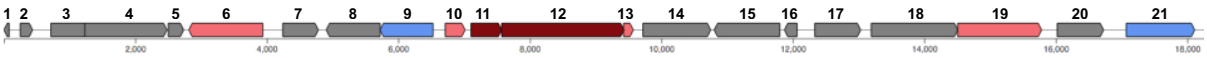


Fig. 21 antiSMASH output of lasso peptide BGC 44.1 with numbered genes

Table 19 Identification of putative lasso peptide genes of BGC 44.1 using Blastx results, gene sizes and antiSMASH annotations. Blastx results, gene sizes and antiSMASH annotations with high similarities to known B1 genes highlighted in blue; modified from (Negri et al. 2022)

Metagenomic lasso peptide BGC 44.1				
Gene	**Blastx results**	**Gene size (bp)**	**antiSMASH annotation**	**Putative lasso peptide gene**
8	ABC transporter permease [*Acidobacteriia bacterium*]	822	-	D1 gene
9	ABC transporter ATP-binding protein [*Acidobacteriia bacterium*]	810	ABC transporter ATP-binding protein	D2 gene
10	PqqD family peptide modification chaperone [*Acidobacteriia bacterium*]	303	PF05402	B1 gene
11	lasso peptide biosynthesis B2 protein [*Acidobacteriia bacterium*]	471	PF13471	B2 gene
12	asparagine synthetase B [*Acidobacteriia bacterium*]	1863	Asn_synthase	C gene
13	hypothetical protein DMG36_15005 [*Acidobacteria bacterium*]	150	predicted lasso peptide	A gene

Specific primers with restriction site overhangs compatible for subsequent cloning were used to amplify the genes A, B1, B2 and C by PCR (3034 bp) and amplicons were verified on an agarose gel (Fig. 22).

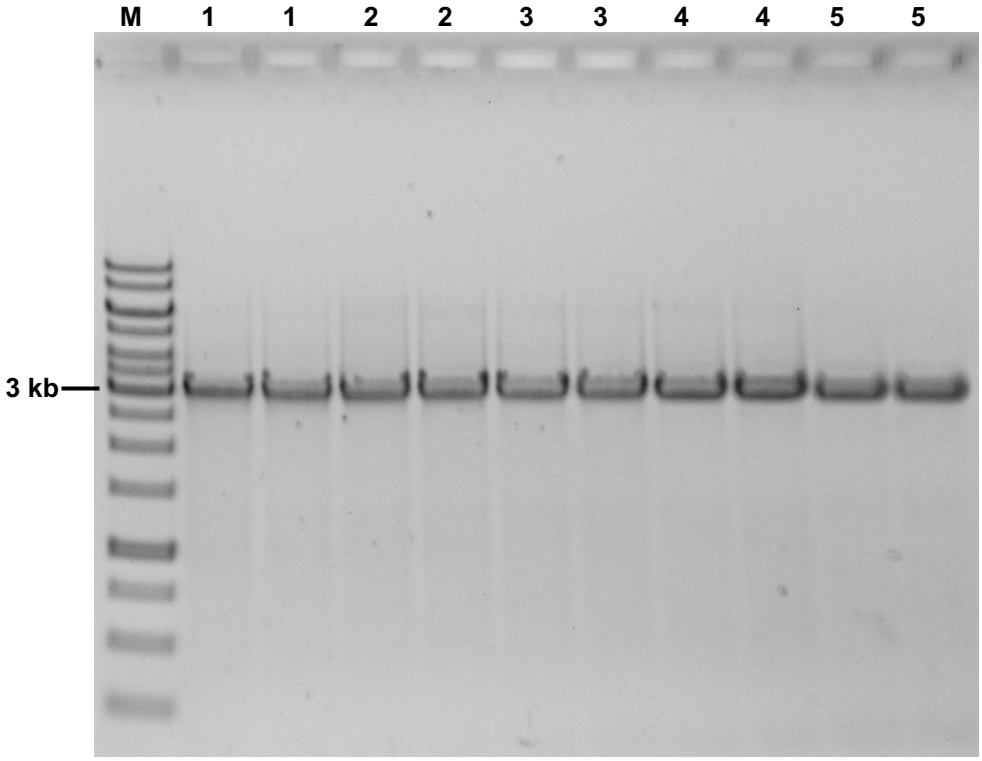


Fig. 22 Agarose gel electrophoresis of PCR products resulting from the amplification of BGC 44.1 lasso peptide genes A, B1, B2 and C. Each PCR reaction performed in duplicates. Lanes 1-5, decreasing amounts of HMW metagenomic DNA used as templates: 114 ng, 54 ng, 30 ng, 15 ng and 3 ng; lane M, GeneRuler 1kb DNA Ladder

The amplified BGC was ligated into the expression vector pSK019 to yield pSK019_Cl44 and the construct was subsequently transferred to *E. coli*. Fourteen individual transformants were screened for the presence of the BGC via PCR and the same primers used for BGC amplification. Analysis of the PCR products on an agarose gel revealed multiple positive clones showing the 3034 bp amplicon (Fig. 23).

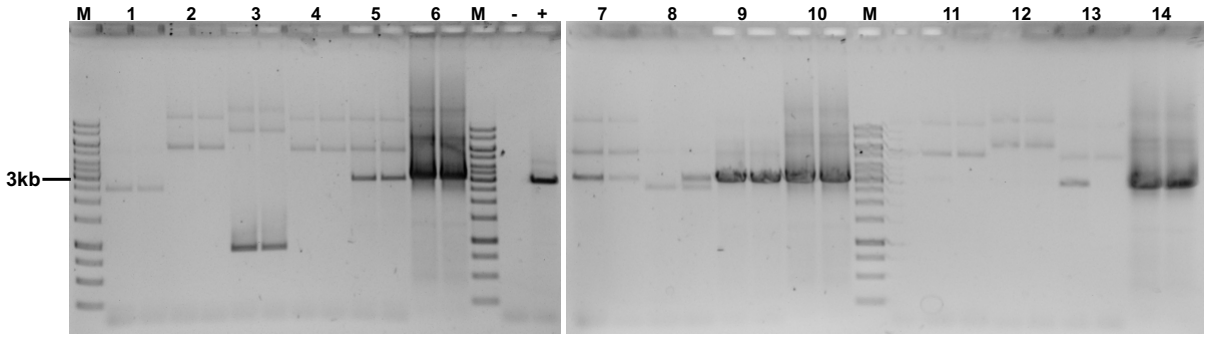


Fig. 23 Agarose gel electrophoresis of PCR products resulting from the screening of *E. coli* pSK019_Cl44 transformants. Each PCR reaction performed in duplicates. Lanes 1-14, plasmids of individual *E. coli* pSK019_Cl44 transformants as templates; lane -, negative control with H_2O instead of template; lane +, positive control with metagenomic DNA as template; lane M, GeneRuler 1kb DNA Ladder

Plasmids of clones 5, 6, 7, 8, 9, 10 and 14 showed the correct amplicon size for at least one PCR reaction and were digested with PvuI and subsequently analyzed on an agarose gel for further verification (Fig. 24A). Plasmids of clones 6, 10 and 14 showed the expected fragments of 3570 bp, 3185 bp and 2813 bp. In a next step, using specific primers with overhangs the translationally coupled genes D1 and D2 of BGC 44.1 were amplified by PCR (1814 bp) and verified on an agarose gel (Fig. 24B).

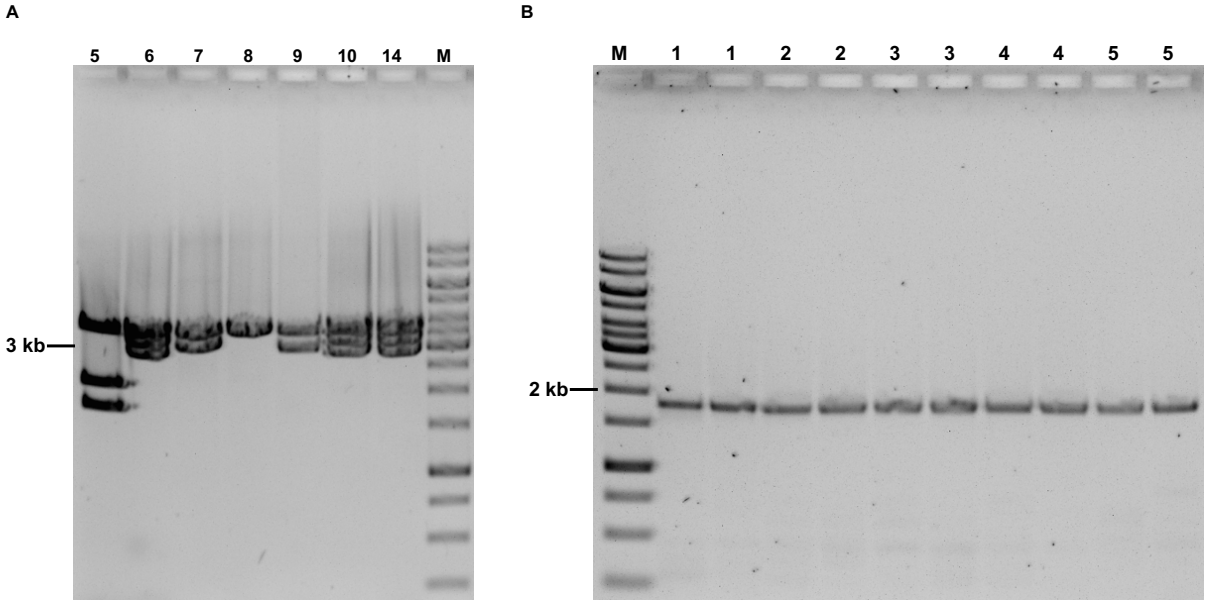


Fig. 24 A) Agarose gel electrophoresis of PvuI digested pSK019_Cl44 candidates. Lanes 5-10 and 14, digested plasmids of PCR-positive clones. **B) Agarose gel electrophoresis of PCR products resulting from the amplification of BGC 44.1 lasso peptide gene D.** Each PCR reaction performed in duplicates. Lanes 1-5, decreasing amounts of HMW metagenomic DNA used as templates: 114 ng, 54 ng, 30 ng, 15 ng and 3 ng; lanes M, GeneRuler 1kb DNA Ladder

The amplified D genes were inserted into pSK019_Cl44 to construct pSK019_Cl44_D. This was done by Gibson Assembly using the attached overhangs of the amplified D genes to insert them downstream of the lasso peptide genes of the linearized pSK019_Cl44 thereby inverting

them to adjust the gene orientation. The construct was transferred to *E. coli* DH10ß cells and individual clones were screened for the presence of the D gene via PCR (data not shown). The plasmids of three PCR-positive clones were digested with PvuI and visualized on an agarose gel (Fig. 25) for further confirmation. Clones 1, 3 and 5 showed the expected fragments of 3570 bp, 3142 bp, 2813 bp and 1821 bp.

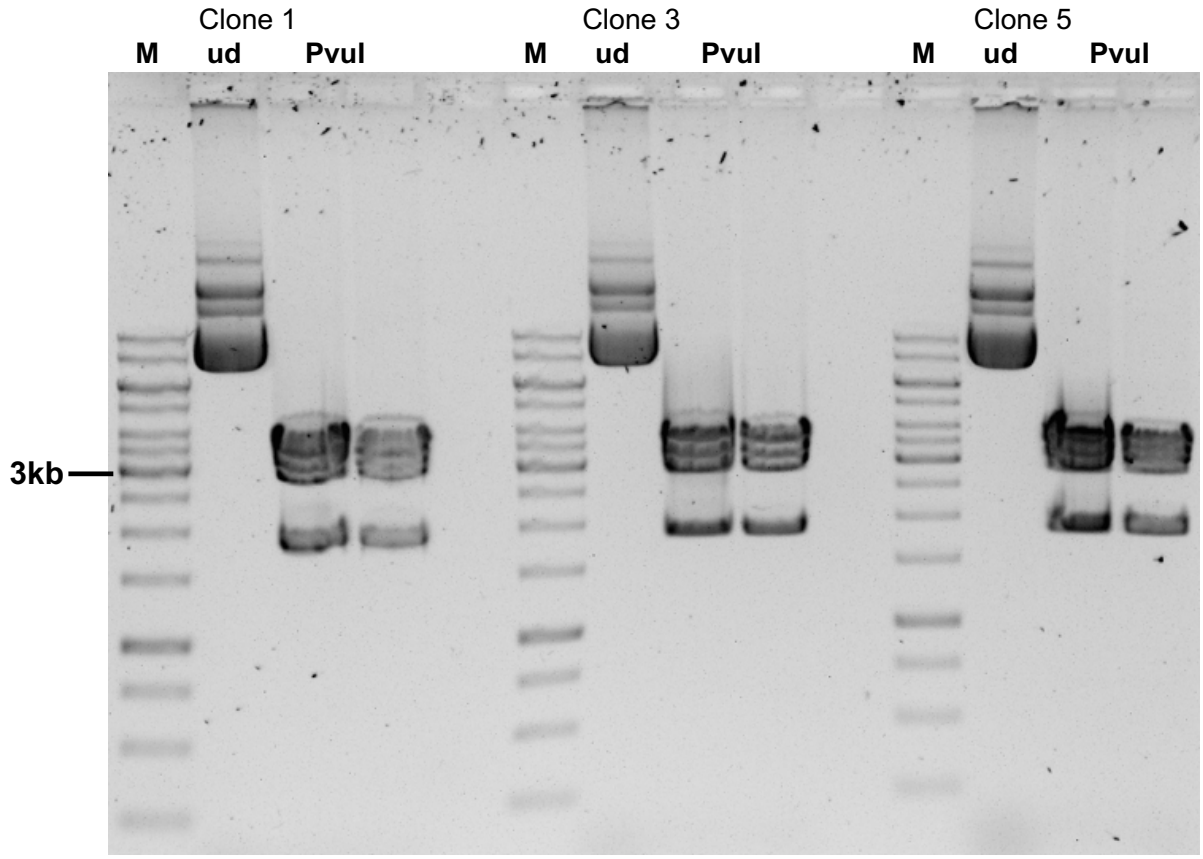


Fig. 25 Agarose gel electrophoresis of PvuI digested pSK019_Cl44_D candidates. Lanes ud, undigested plasmids; lanes PvuI, 15 µl and 5 µl of each PvuI digested plasmid; lanes M, GeneRuler 1kb DNA Ladder

The construct pSK019_Cl44_D containing the complete lasso peptide BGC 44.1 was finally confirmed by sequencing of the insert (Fig. 26).

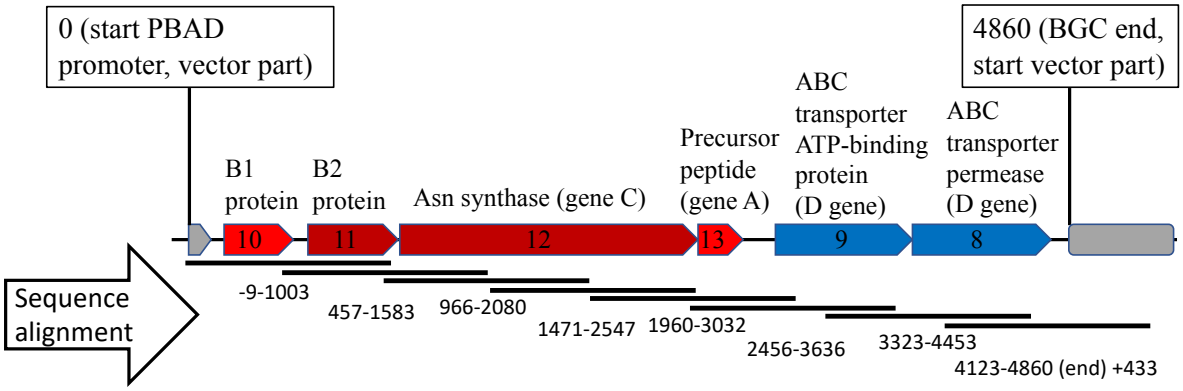


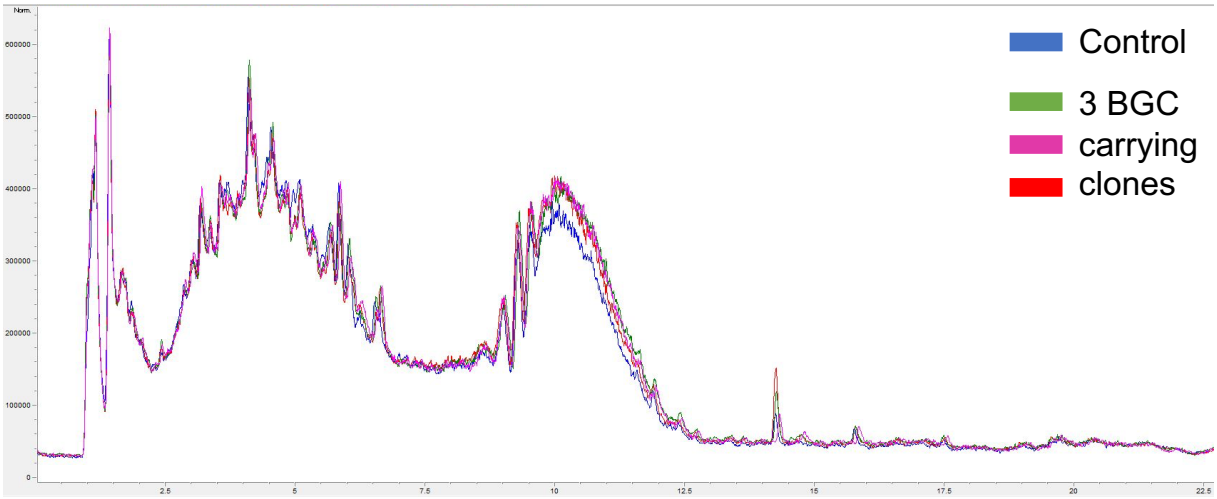
Fig. 26 Sequencing confirmation for lasso peptide BGC 44.1 cloned into an expression vector. Sequencing of lasso peptide BGC 44.1 using specific sequencing primers that generate sequences with overlap. Numbers of the alignment refer to the start and end point of each generated sequence that matches the reference sequence. Vector derived PBAD promoter determined as starting point of the reference sequence. Gene annotation via bioinformatics analysis as described; adopted from (Negri et al. 2022)

5.5.7 Heterologous expression experiments of BGC 44.1 in *E. coli*

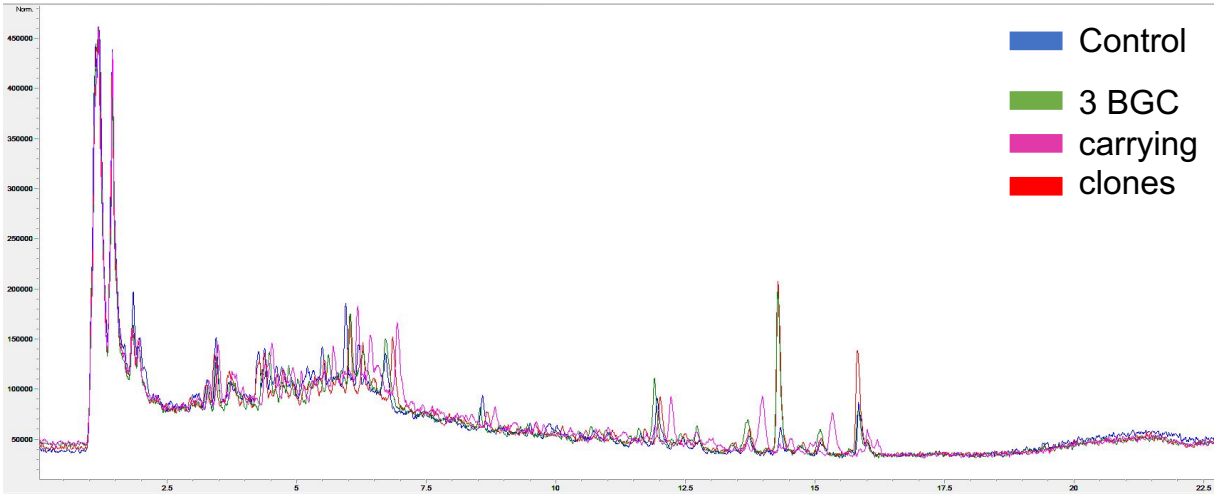
In a first attempt, the three confirmed *E. coli* DH10ß clones carrying pSK019_Cl44_D as well as the *E. coli* DH10ß control strain carrying the empty pSK019 vector were used for heterologous expression experiments.

The strains were grown in M9 as well as M20 minimal media respectively each supplemented with 100 mM arabinose for induction of the PBAD promoter. After three days, the culture supernatants were first extracted with XAD-16 resin and subsequently with butanol, while the cell pellets were extracted with MeOH. In Fig. 27 the respective mass chromatograms for M9 culture extracts are exemplarily shown, which are representative for BGC 44.1 heterologous expression experiments in *E. coli*. Additional peaks that were only observed for BGC carrying clones but not for the empty vector control were each investigated and evaluated manually as described before, which did not lead to the detection of a potential lasso peptide mass.

A 3-day M9 cultures extracted with XAD-16 resin



B 3-day M9 cultures extracted with butanol



C 3-day M9 cultures, pellet MeOH extraction

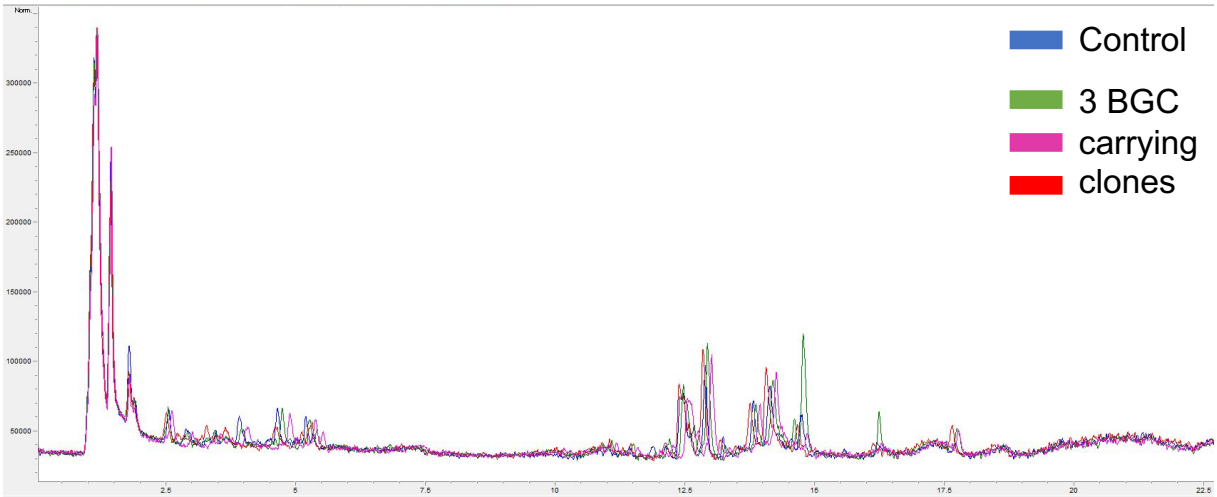


Fig. 27 HPLC-MS mass chromatograms of extracts derived from *E. coli*/pSK019_Cl44_D clones as well as the *E. coli*/pSK019 control strain. 3-day M9 cultures extracted with XAD-16 resin **(A)**, butanol **(B)** and pellet MeOH extraction **(C)**. Blue, *E. coli*/pSK019 control culture extracts; remaining colors, extracts of the three BGC carrying clones

5.5.8 Transfer of pSK019_Cl44_D to *Burkholderia* sp. FERM BP-3421

Since no lasso peptide could be detected upon heterologous expression experiments with *E. coli*, *Burkholderia* sp. FERM BP-3421 was chosen as an alternative heterologous expression host, as the pSK019 vector also carried the genetic elements for replication in *Burkholderia*. pSK019_Cl44_D was transferred to *Burkholderia* FERM BP-3421 and plasmids of three individual transformants were isolated to check their integrity via restriction digestion. For that purpose, the plasmids were retransferred to *E. coli* first as the plasmid copy number in *Burkholderia* is not sufficient for restriction digestion analysis. Plasmids from two transformants of each of the three retransferred plasmids were isolated, digested with PvuI and subsequently analyzed via agarose gel electrophoresis revealing the presence of all expected fragments (Fig. 28).

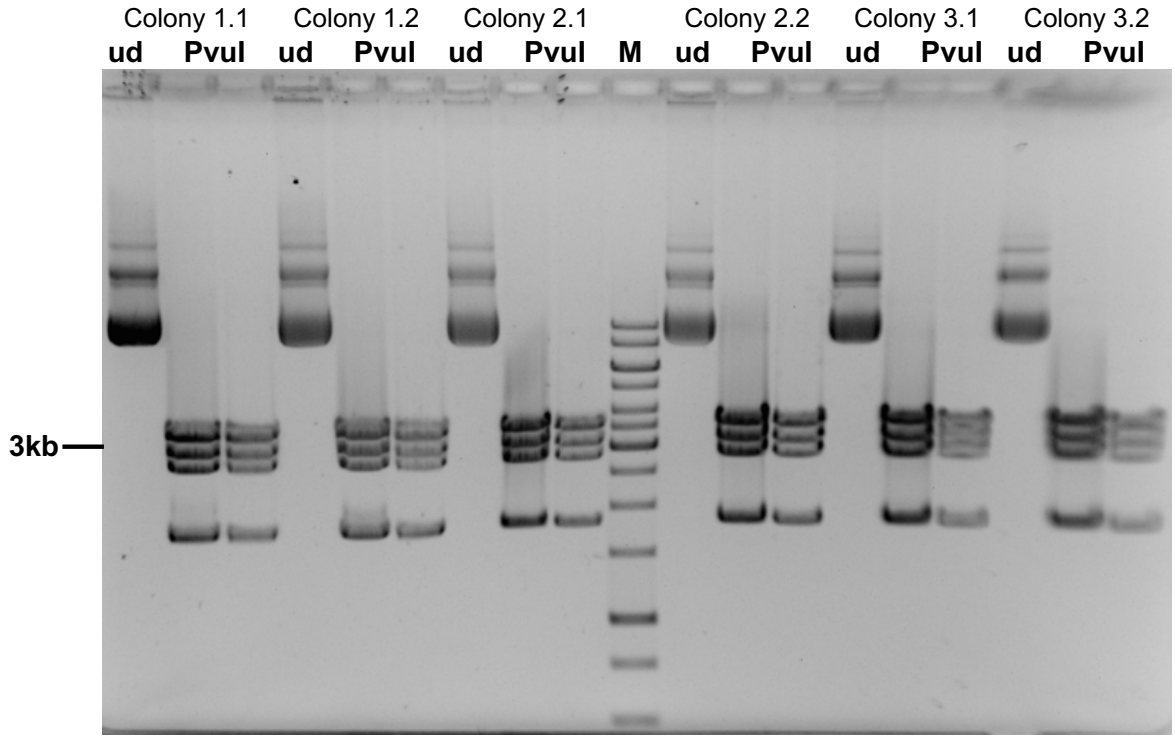


Fig. 28 Agarose gel electrophoresis of *Burkholderia* derived PvuI digested pSK019_Cl44_D plasmids. Plasmids isolated from two *E. coli* transformants of each of the three retransferred *Burkholderia* plasmids. Lanes ud, undigested plasmids; lanes PvuI, 15 µl and 5 µl of each PvuI digested plasmid; lane M, GeneRuler 1kb DNA Ladder

5.5.9 Heterologous expression experiments and transcription analysis of BGC 44.1 in *Burkholderia* sp. FERM BP-3421

The three confirmed *Burkholderia* clones carrying pSK019_Cl44_D as well as the control strain carrying the empty pSK019 vector were cultured in M9 and M20 media respectively both supplemented with 100 mM arabinose for induction of the PBAD promoter. Samples were

taken after two and three days and stored at -80 °C for later RNA isolation. After three days, the cultures were extracted and analyzed equivalent to the procedure for the *E. coli* cultures. Since no lasso peptide could be detected, RNA was isolated from the two- and three-day samples to check if all lasso peptide genes were transcribed. For that purpose, multiple primer pairs were designed to amplify regions of the lasso peptide genes B1, B2 to C, C and D from reverse transcribed cDNA while using the respective RNAs as a control. Agarose gel electrophoresis revealed PCR products for B1 gene (Fig. 29A) and B2 to C gene (Fig. 29B) regions for samples from 2- and 3-day cultures.

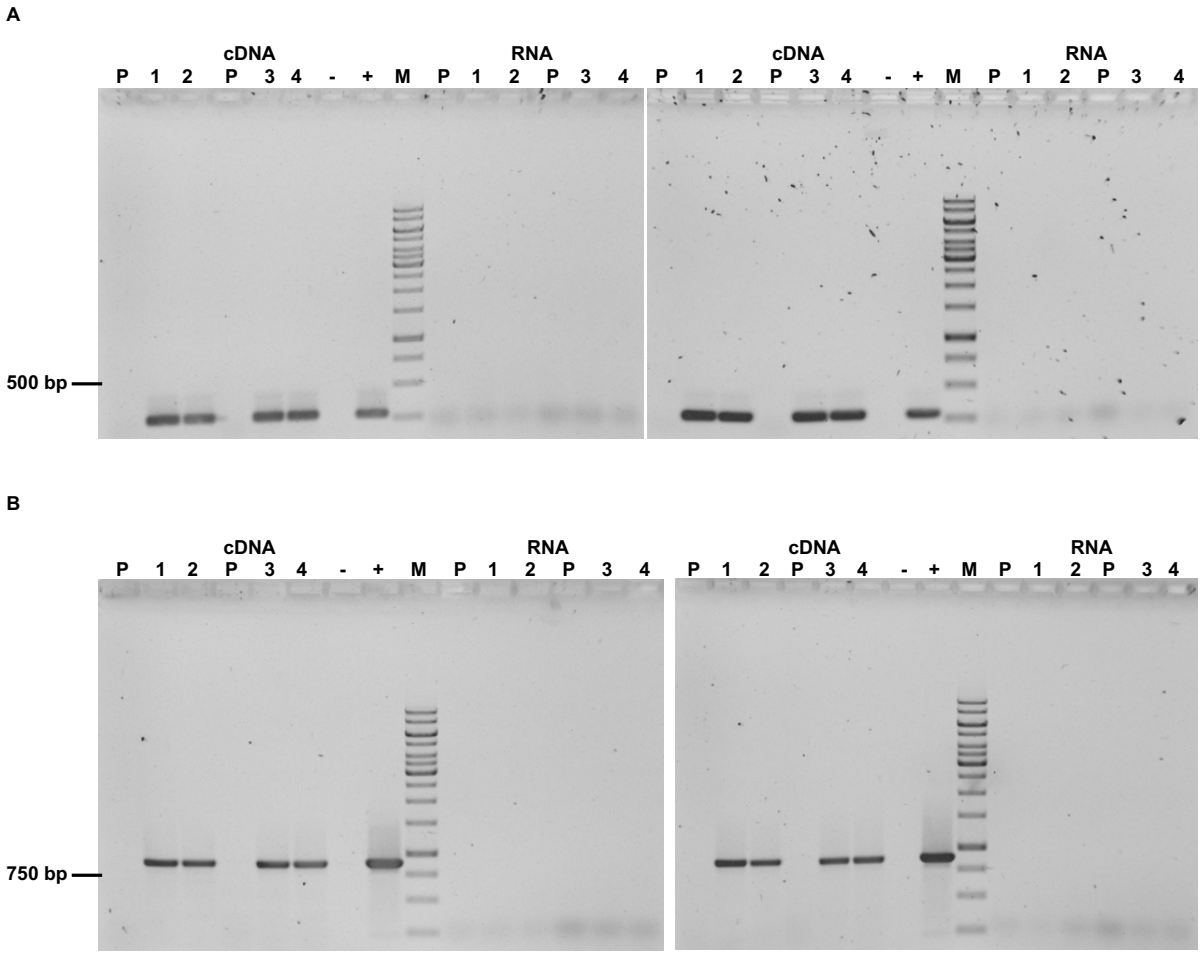


Fig. 29 Agarose gel electrophoresis of PCR products resulting from B1, B2 and C gene transcription analysis in *Burkholderia*. A) PCR products for amplification of a B1 gene part using cDNA and RNA from 2-day (left) and 3-day (right) cultures. **B)** PCR products for amplification of a B2 to C gene part using cDNA and RNA from 2-day (left) and 3-day (right) cultures. Lanes P, RNA and respective cDNA isolated from *Burkholderia*/pSK019 as template; lanes 1 and 2, RNA and respective cDNA isolated from two *Burkholderia*/pSK019_Cl44_D clones grown in M9 medium as templates; lanes 3 and 4, RNA and respective cDNA isolated from two *Burkholderia*/pSK019_Cl44_D clones grown in M20 medium as templates; lanes -, negative controls with H_2O instead of template; lanes +, positive controls with pSK019_Cl44_D as template

No PCR products could be detected for the C (Fig. 30A) and D gene (Fig. 30B) regions indicating that either transcription stopped in the middle of the BGC or the RNA was degraded from the 3-prime end during the culturing process.

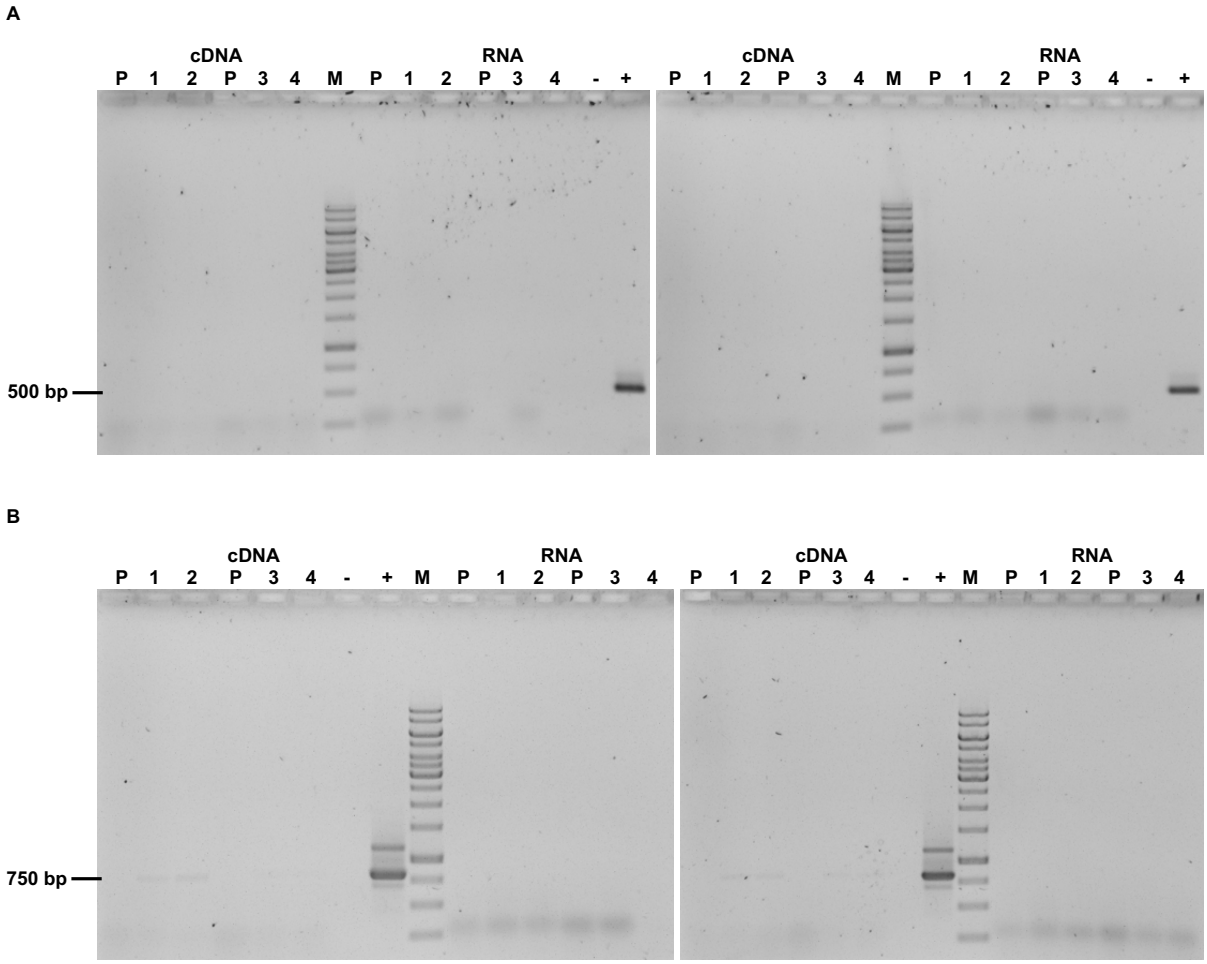


Fig. 30 Agarose gel electrophoresis of PCR products resulting from C and D gene transcription analysis in *Burkholderia*. A) PCR products for amplification of a C gene part using cDNA and RNA from 2-day (left) and 3-day (right) cultures. **B)** PCR products for amplification of a D gene part using cDNA and RNA from 2-day (left) and 3-day (right) cultures. Lanes P, RNA and respective cDNA isolated from *Burkholderia*/pSK019 as template; lanes 1 and 2, RNA and respective cDNA isolated from two *Burkholderia*/pSK019_Cl44_D clones grown in M9 medium as templates; lanes 3 and 4, RNA and respective cDNA isolated from two *Burkholderia*/pSK019_Cl44_D clones grown in M20 medium as templates; lanes -, negative controls with H_2O instead of template; lanes +, positive controls with pSK019_Cl44_D as template

To investigate whether the transcript was degraded due to a too long culturing time, both *E. coli* and *Burkholderia* FERM BP-3421 carrying pSK019_Cl44_D were first grown to a certain optical density (OD) and subsequently induced for a shorter period of time before samples for RNA isolation were taken. *E. coli* was induced for 3 hours in LB medium, while *Burkholderia* was induced overnight in 2S4G medium. RNA was isolated from both and reverse transcribed to cDNA followed by PCR targeting the same regions as described before. Agarose gel electrophoresis of the PCR products revealed that for *Burkholderia* FERM BP-3421 no

difference in transcription occurred as products could be detected for gene regions B1 and B2 to C (Fig. 31A), while the C gene was not detectable (Fig. 31B). In contrast two of the three *E. coli* clones showed products for all gene regions revealing that the BGC was completely transcribed (Fig. 31A to C).

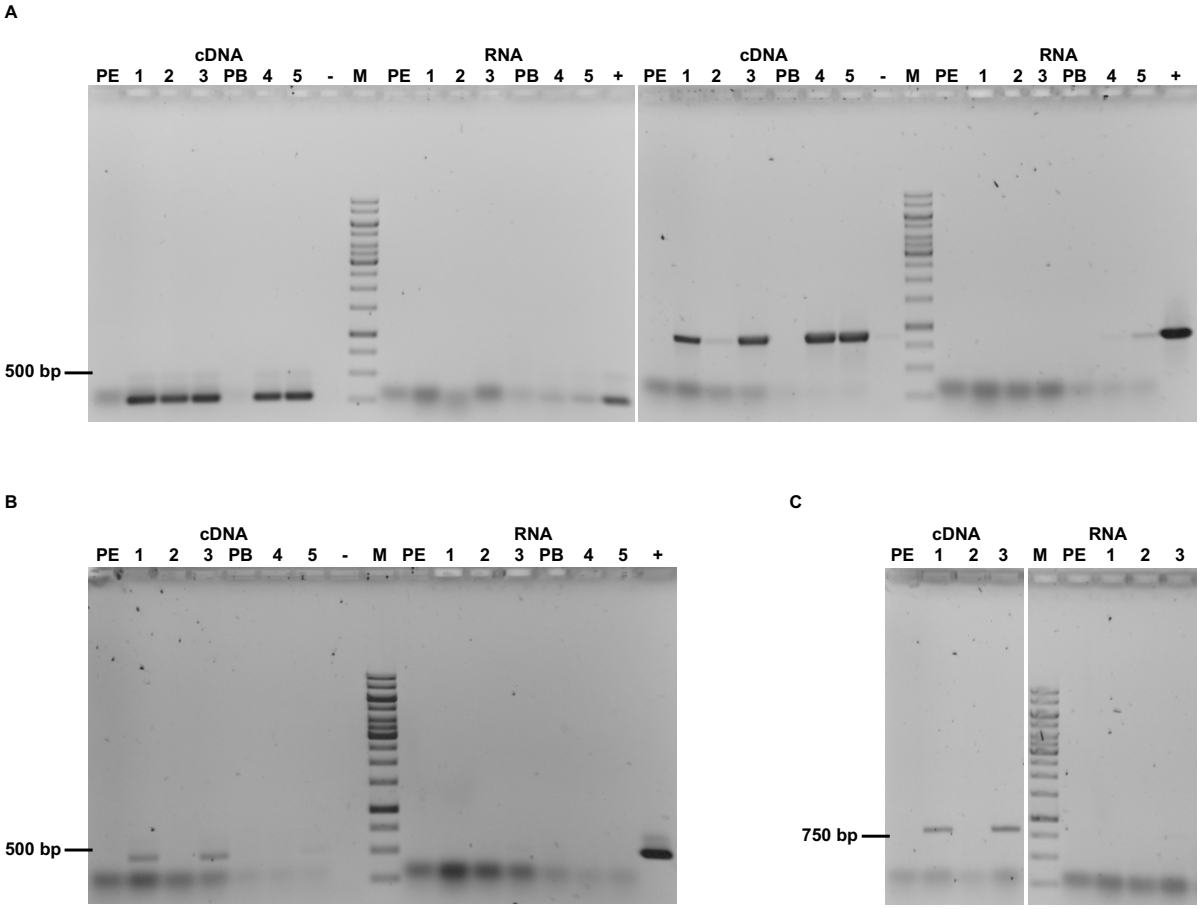


Fig. 31 Agarose gel electrophoresis of PCR products resulting from BGC 44.1 transcription analysis for short induction time in *E. coli* and *Burkholderia*. PCR products for amplification of a B1 gene part (**A**, **left**), a B2 to C gene part (**A**, **right**) and a C gene part (**B**) using cDNA and RNA from *E. coli* and *Burkholderia* cultures. **C)** PCR products for amplification of a D gene part using cDNA and RNA from *E. coli* cultures. Lanes PE, RNA and respective cDNA isolated from DH10ß/pSK019 as template; lanes PB, RNA and respective cDNA isolated from *Burkholderia*/pSK019 as template; lanes 1, 2 and 3, RNA and respective cDNA isolated from three DH10ß/pSK019_Cl44_D clones as templates; lanes 4 and 5, RNA and respective cDNA isolated from two *Burkholderia*/pSK019_Cl44_D clones as templates; lanes -, negative controls with H_2O instead of template; lanes +, positive controls with pSK019_Cl44_D as template

Since the *E. coli* LB cultures showed transcripts for all genes, half of the cultures were extracted after 3 hours of induction, while the other halves were grown for an additional three days and were then equally extracted. The samples are currently being analyzed using an alternative detection method.

5.5.10 Assembly of a large NRPS BGC distributed over three different fosmids via transformation-associated recombination (TAR) cloning

The generated sequencing information of the metagenomic fosmid library and its corresponding metagenomic DNA opened up the possibility of not only recovering small BGC classes such as RiPPs, but also other classes that are typically larger in size even with limited library size. The BGCs detected on contigs derived from the fosmid library and metagenome sequencing were investigated for larger clusters suitable for recovery. The approximately 58 kb NRPS BGC 76.1 (Fig. 32) that was detected on a 138.907 bp contig derived from the direct metagenome sequencing proved to be suitable for recovery as it was also almost completely covered by different contigs derived from the fosmid library sequencing.

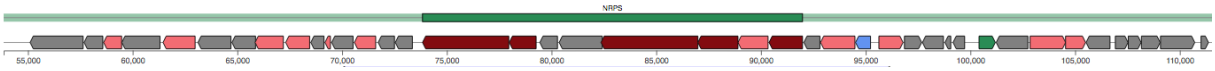


Fig. 32 antiSMASH output of NRPS BGC 76.1

In order to screen the fosmid library for clones covering the complete NRPS BGC, specific primers targeting the left, middle and right part of the BGC were designed (Fig. 38A). In the following, the recovery of a clone from the library that was positive for the middle part of the BGC is exemplarily shown. The recovery of clones positive for the left and right part was performed equivalently. Screening the pools of 2000 clones of the fosmid library with primers specific for the middle part of the BGC and subsequent analysis of the PCR products on agarose gels revealed pool 24 to be positive as it showed amplicons of the correct size of 594 bp (Fig. 33A). A serial dilution PCR strategy was applied to recover the positive clone from the pool of 2000 clones. For that purpose, the pool was diluted, which enabled the inoculation of 96 wells of four 24-well plates each with approx. 400 cells. After cells were grown overnight, 12 wells each were pooled and fosmids isolated, followed by screening for positive pools via PCR. Analysis of PCR products on an agarose gel revealed a positive pool consisting of wells 49-60 (Fig. 33B).

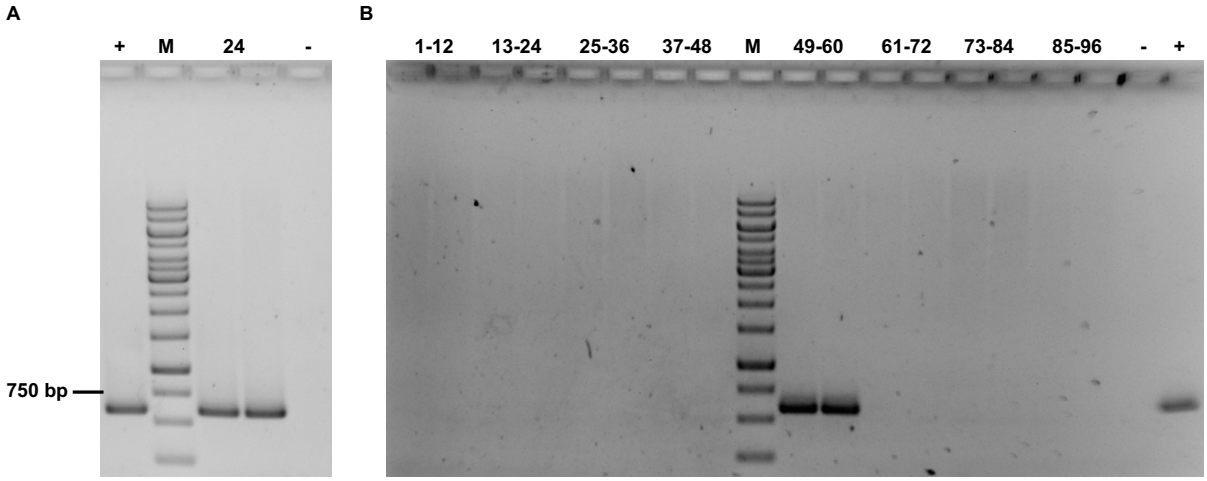


Fig. 33 Agarose gel electrophoresis of PCR products resulting from the screening of the fosmid library for the middle part of BGC 76.1. A) Screening of pools of 2000 clones. Lane 24, fosmid DNA of pool number 24 as template. **B)** Screening of dilutions of positive pool number 24. Lane 1-12, fosmid DNA of pooled wells 1-12 as template; lane 13-24, fosmid DNA of pooled wells 13-24 as template; lane 25-36, fosmid DNA of pooled wells 25-36 as template; lane 37-48, fosmid DNA of pooled wells 37-48 as template; lane 49-60, fosmid DNA of pooled wells 49-60 as template; lane 61-72, fosmid DNA of pooled wells 61-72 as template; lane 73-84, fosmid DNA of pooled wells 73-84 as template; lane 85-96, fosmid DNA of pooled wells 85-96 as template; lanes M, GeneRuler 1kb DNA Ladder; lanes +, positive control with total fosmid library DNA as template; lanes -, negative controls with H_2O instead of template

Individual wells 49-60 were subsequently separately screened via PCR. Analysis of the PCR products via agarose gel electrophoresis revealed a positive signal for well 55 (Fig. 34).

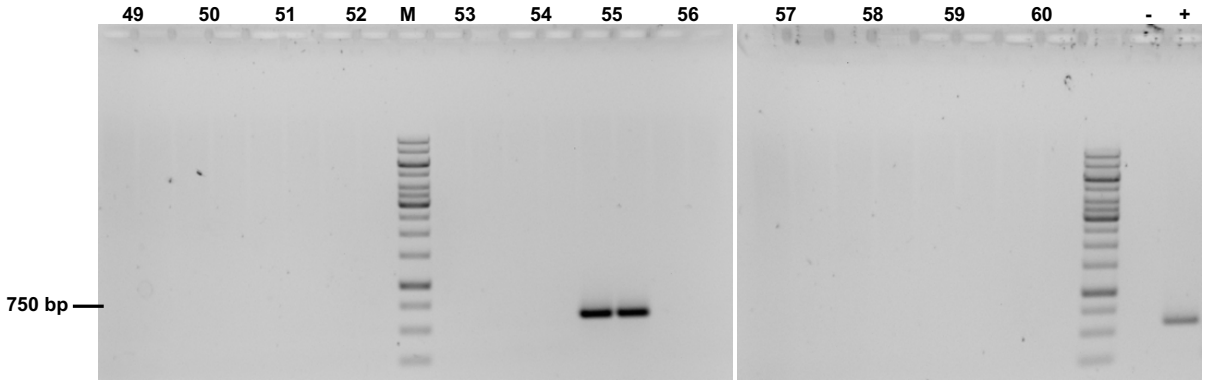


Fig. 34 Agarose gel electrophoresis of PCR products resulting from the screening of fosmid library wells for the middle part of BGC 76.1. Lanes 49-60, fosmid DNA of wells 49-60 as templates; lanes M, GeneRuler 1kb DNA Ladder; lane +, positive control with total fosmid library DNA as template; lane -, negative control with H_2O instead of template

Cells of well 55 were diluted, plated and grown overnight. 144 individual colonies were picked and grown in three 48 well plates. 16 wells each containing individual clones were pooled and the resulting nine pools were screened again leading to the identification of a positive pool consisting of 16 clones (Fig. 35A). Finally, the 16 clones were screened individually leading to the identification of a single clone positive for the middle part of BGC 76.1 (Fig. 35B).

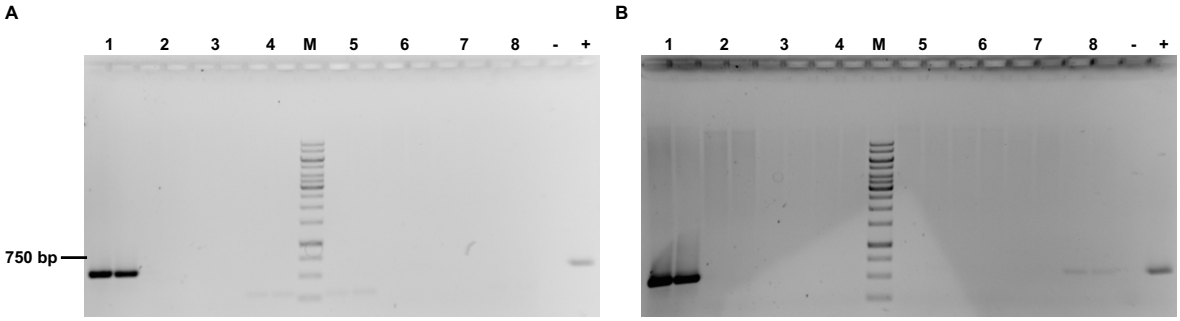


Fig. 35 Agarose gel electrophoresis of PCR products resulting from the screening of individually picked clones for the middle part of BGC 76.1. A) Screening of pools consisting of 16 individually picked clones each. Lanes 1-8, fosmid DNA of pools 1-8 as templates. **B)** Screening of individually picked clones. Lanes 1-8, fosmid DNA of individual clones 1-8 as templates. Lanes M, GeneRuler 1kb DNA Ladder; lanes +, positive controls with total fosmid library DNA as template; lanes -, negative controls with H_2O instead of template

After isolating two clones from the library that were positive for the left and right part of the BGC using the same strategy, the three fosmids were end-sequenced from both sides. Alignment of the resulting sequences with the NRPS BGC containing 138.907 bp contig derived from the metagenome sequencing revealed that the three fosmids together covered the complete NRPS BGC with shared sequence overlap between each fosmid (Fig. 36).

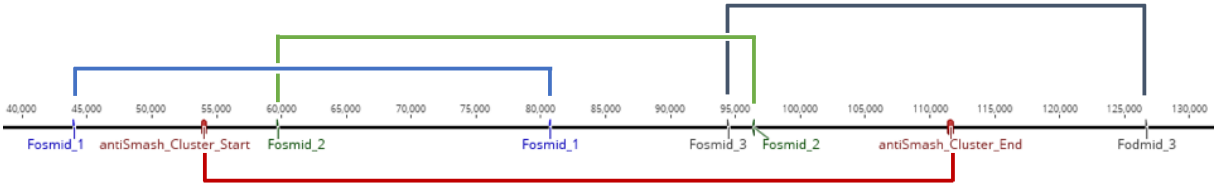


Fig. 36 Alignment of end-sequences of each fosmid with the NRPS BGC carrying contig. Red, NRPS BGC on contig; blue, contig part carried by fosmid 1; green, contig part carried by fosmid 2; grey, contig part carried by fosmid 3

To assemble the BGC from the three fosmids via TAR cloning, the fosmids needed to be linearized first. For that purpose, each fosmid was digested with a unique restriction enzyme, which released three large overlapping fragments that together still covered the complete cluster. Digested fosmids 1 and 2 shared 18.570 bp and digested fosmids 2 and 3 shared 1.990 bp of BGC overlap. The BGC parts were assembled to a complete cluster using TAR cloning in yeast (Fig. 38A). Multiple yeast clones were screened for the presence of the completely captured BGC (pCAP03_Cl76) via colony PCR using the same three primer pairs that were used for fosmid library screening before. Analysis of PCR products on agarose gels revealed yeast clone 157 to be positive for all three BGC parts (Fig. 37).

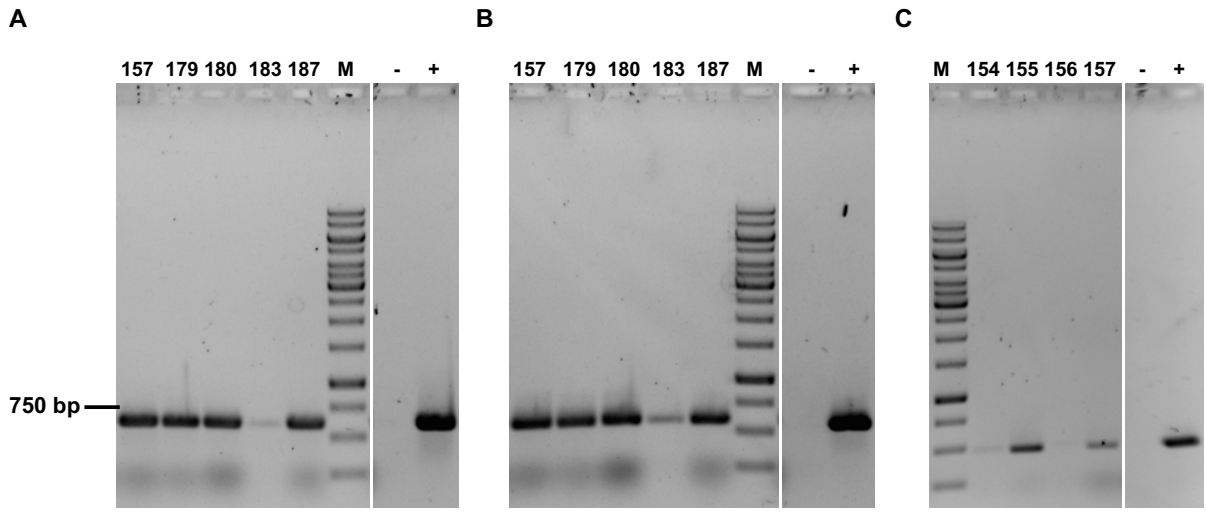


Fig. 37 Agarose gel electrophoresis of colony PCR products resulting from the screening of yeast clones for the three parts of BGC 76.1. Screening for the left **(A)** and middle **(B)** part of the BGC. Lanes 157, 179, 180, 183, 187, individual yeast clones as templates. **C)** Screening for the right part of the BGC. Lanes 154, 155, 156, 157, individual yeast clones as templates. Lanes M, GeneRuler 1kb DNA Ladder; lanes +, positive controls with total fosmid library DNA as template; lanes -, negative controls with H_2O instead of template

The plasmid of yeast clone 157 was isolated and subsequently transferred to *E. coli* followed by its isolation. The correct assembly of the cluster was confirmed by cutting the final plasmid into two fragments (SmiI) of approximately 57 kb and 16 kb via restriction digestion and subsequent sequencing of the two fragments via Nanopore. The resulting Nanopore reads were filtered for size ranges (15.5 kb – 16.5 kb and 56 kb – 58 kb respectively) matching the two fragments and subsequently aligned with the SmiI digested pCAP03_Cl76 reference sequence that was constructed using the metagenome sequencing data (Fig. 38B). The majority of size filtered reads aligned gapless to the two reference sequence parts thereby confirming the accurate assembly of the NRPS BGC 76.1 from three different fosmids to a complete cluster.

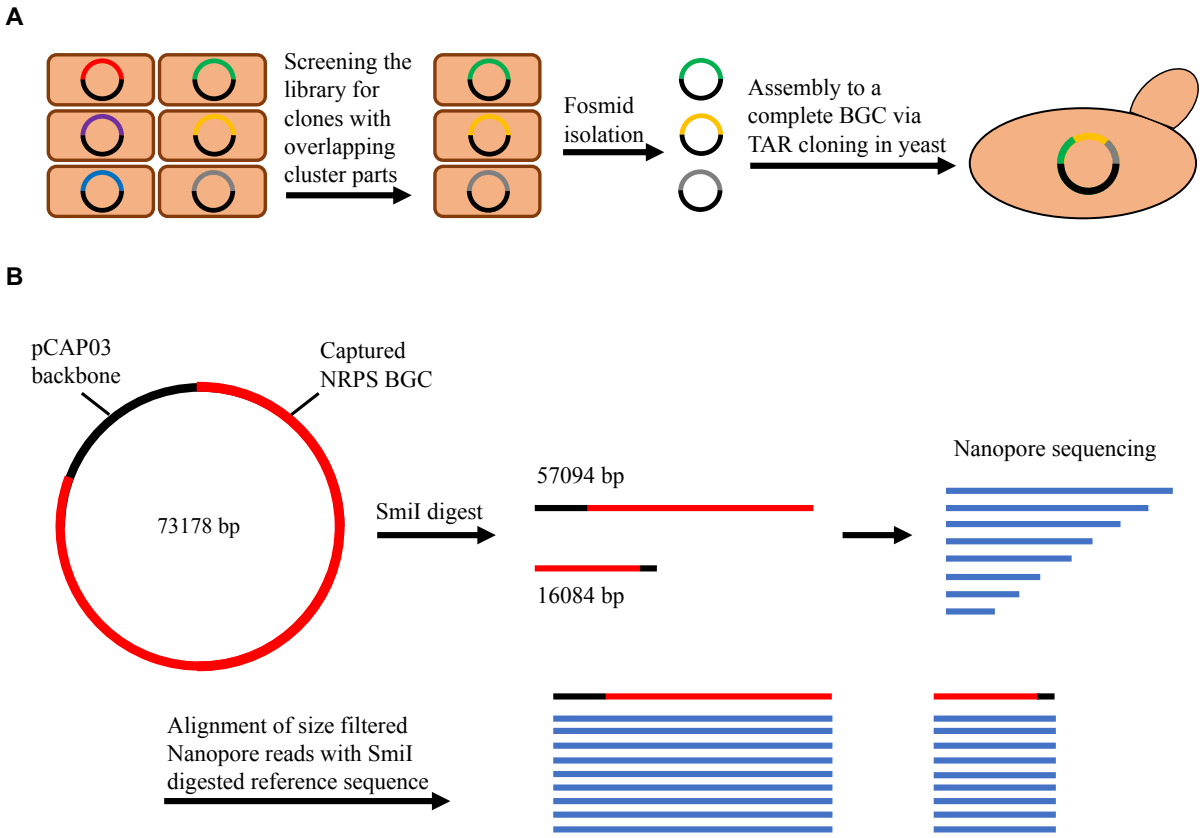


Fig. 38 Workflow for the assembly of a large NRPS BGC from cluster parts on different fosmids via TAR cloning. A) Screening of the fosmid library for clones carrying fosmids with overlapping BGC parts and subsequent isolation of the respective clones. Fosmid isolation and assembly of the cluster parts into a complete cluster via TAR cloning in yeast. **B)** SmiI digest of the assembled NRPS BGC generates two fragments for subsequent Nanopore sequencing. Size filtering of the generated reads and alignment with SmiI digested reference sequence; modified from (Negri et al. 2022)

5.5.11 Heterologous expression experiments of NRPS BGC 76.1

The NRPS BGC carrying TAR vector pCAP03_Cl76 was transferred to *Streptomyces albus* Del 14 (*S. albus*) (Myronovskyi et al. 2018) via triparental conjugation to conduct heterologous expression experiments. Three individual exconjugants were screened for the presence of the BGC via PCR and the same primers that were used previously to screen the fosmids for the left, middle and right part of the BGC. Analysis of the PCR products via agarose gel electrophoresis revealed that a strong band could only be detected for the left part, while the middle part showed very light bands and the right part was completely missing (Fig. 39). Since all three exconjugants showed the same pattern, it was hypothesized that specific parts of the BGC were deleted by the host, which is why *Streptomyces coelicolor* (*S. coelicolor*) M1152 (Gomez-Escribano and Bibb 2011) was tested as an alternative host.

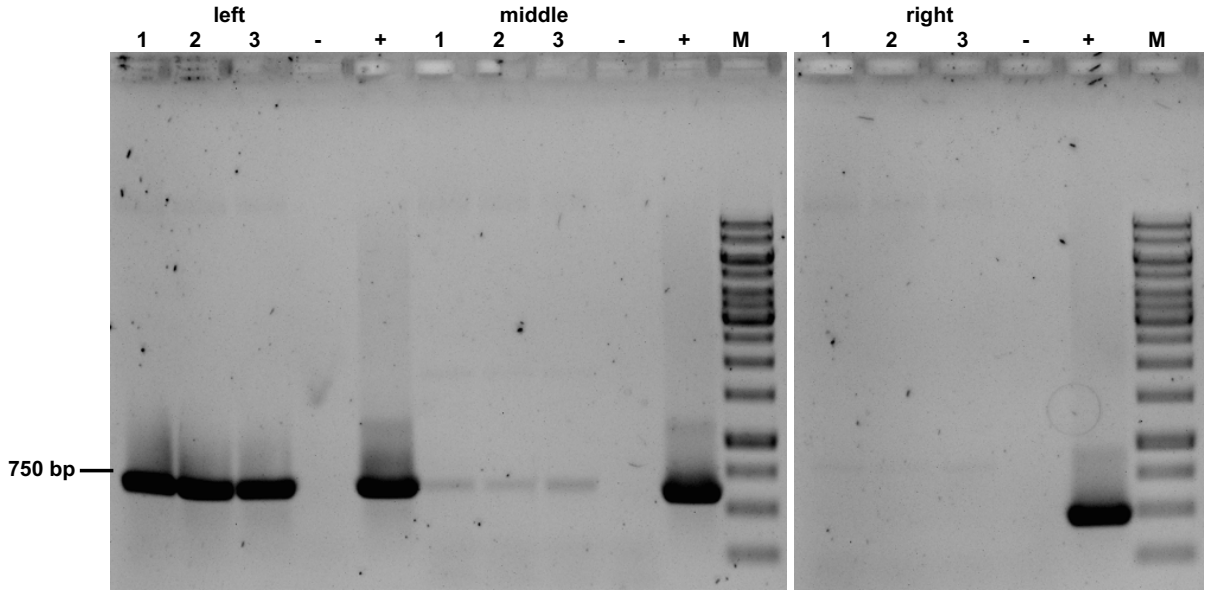


Fig. 39 Agarose gel electrophoresis of PCR products resulting from the screening of *S. albus* exconjugants for BGC 76.1. Screening for the left, middle and right part of the BGC. Lanes 1-3, *S. albus* genomic DNA (gDNA) of three individual exconjugants as templates; lanes M, GeneRuler 1kb DNA Ladder; lanes +, positive controls with pCAP03_Cl76 as template; lanes -, negative controls with H_2O instead of template

pCAP03_Cl76 was transferred to *S. coelicolor* via triparental conjugation and five exconjugants were screened for the presence of the BGC via PCR. Analysis of respective PCR products on an agarose gel revealed that all exconjugants were positive for the three BGC parts (Fig. 40).

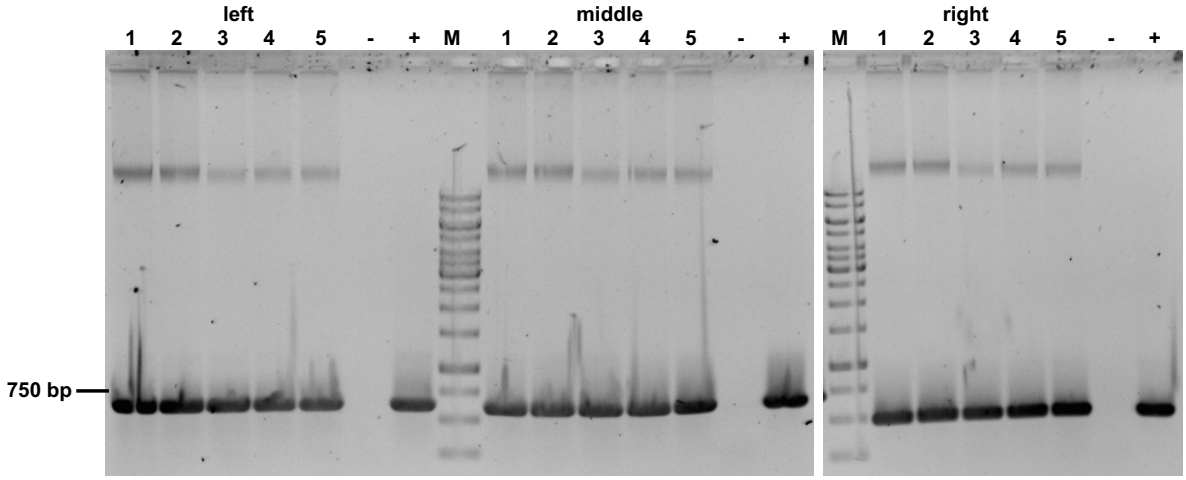


Fig. 40 Agarose gel electrophoresis of PCR products resulting from the screening of *S. coelicolor* exconjugants for BGC 76.1. Screening for the left, middle and right part of the BGC. Lanes 1-5, *S. coelicolor* gDNA of five individual exconjugants as templates; lanes M, GeneRuler 1kb DNA Ladder; lanes +, positive controls with pCAP03_Cl76 as template; lanes -, negative controls with H_2O instead of template

However, the *Streptomyces* exconjugants did not necessarily contain a BGC integrated into every chromosome as they can harbor multiple chromosomes within their cells at a time and a single integration already confers the resistance that is selected for. Spores are generally

considered to only contain a single chromosome and thus spores that were generated from the five positive exconjugants could have been composed of mixtures with and without the BGC integrated into the chromosome. Therefore, the exconjugant spores were diluted and plated again to yield clones derived from a single spore. Colonies derived from single spores were subsequently investigated for the presence of the BGC via PCR. Agarose gel electrophoresis of the PCR products revealed that all clones derived from single spores lacked the middle and right part of the BGC (data not shown) as it was the case for *S. albus*. These results indicated that parts of the BGC were also deleted in *S. coelicolor* although in a slower process as the respective bands of screened exconjugants only disappeared after further rounds of culturing that were necessary to generate spores derived from single spores. The same pattern was observed after conjugation of the construct into *Streptomyces lividans* (data not shown). Other actinobacteria tested such as *Amycolatopsis japonicum* showed missing BGC parts directly after conjugation (data not shown) as it was observed for *S. albus*.

S. coelicolor was chosen for heterologous expression experiments to test for the detectability of a BGC specific compound even when the BGC is deleted while culturing. To increase the chances for compound production, culturing steps were minimized to shorten the time span for BGC deletion. Therefore, spores generated from three of the five positive *S. coelicolor* exconjugants as well as the control strain carrying the empty pCAP03 vector were directly inoculated in R5 media and M8 media respectively without preparing a preculture. The R5 cultures were grown for seven days, while the M8 cultures were grown for nine days. To confirm the hypothesis of the BGC being deleted with increasing culturing time, an aliquot was taken after two and seven days for DNA isolation and subsequent PCR. Analysis of the PCR products for the left, middle and right part of the BGC revealed that after two days clones 2 and 3 were still positive for all three BGC parts, while neither of them was positive anymore for all three parts after seven days (Fig. 41).

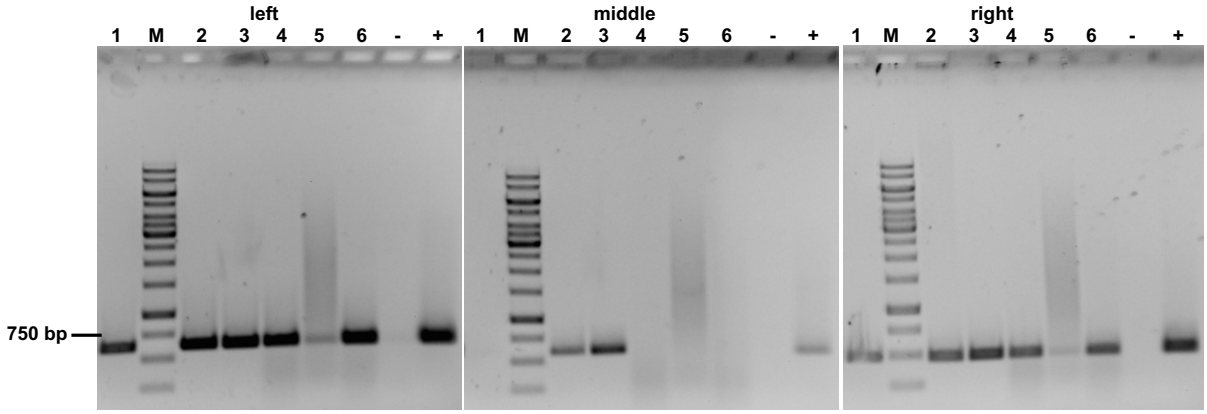
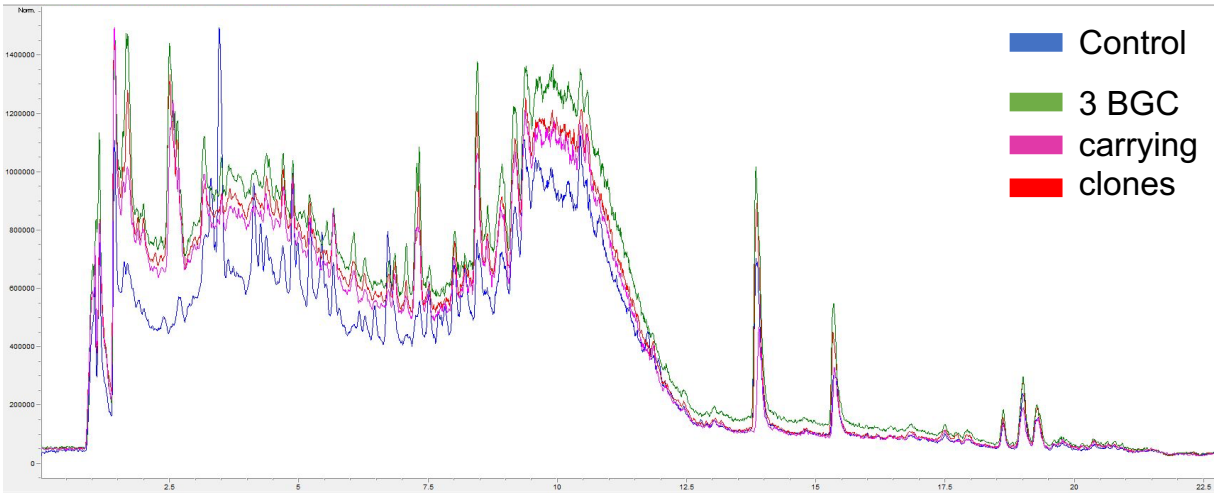


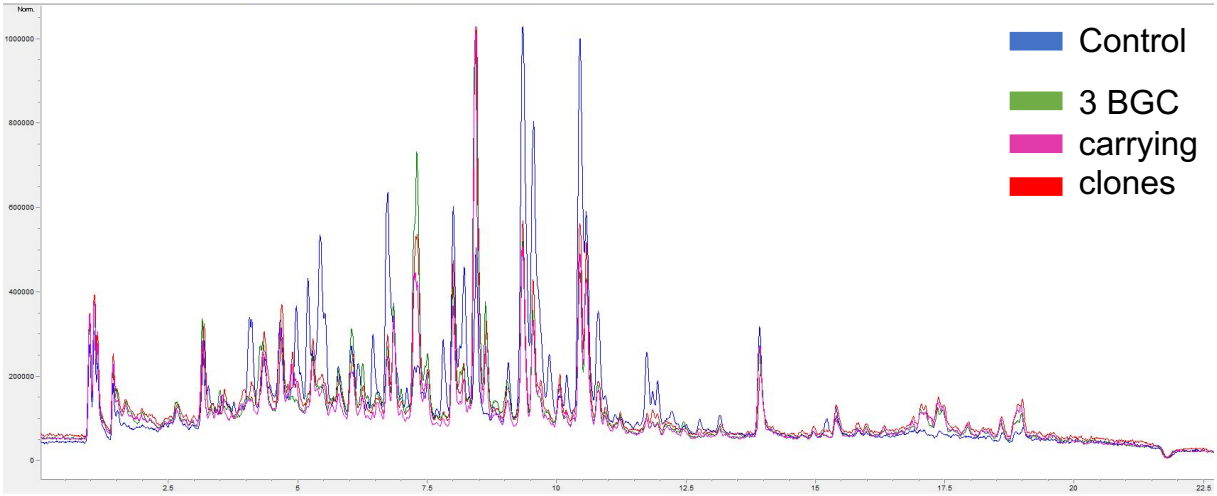
Fig. 41 Agarose gel electrophoresis of PCR products resulting from the investigation of *S. coelicolor* exconjugants for the presence of BGC 76.1 after two and seven days of culturing. Screen for the left, middle and right part of the BGC. Lanes 1-3, *S. coelicolor* gDNA of three individual exconjugants cultured for two days as templates; lanes 4-6, *S. coelicolor* gDNA of the same exconjugants cultured for seven days as templates; lanes M, GeneRuler 1kb DNA Ladder; lanes +, positive controls with pCAP03_Cl76 as template; lanes -, negative controls with H_2O instead of template

After seven days (R5 cultures) and nine days (M8 cultures), half of the culture supernatants were extracted with ethyl acetate and subsequently with butanol and the other halves were extracted with XAD-16 resin. Cell pellets were extracted with 50% methanol/50% acetone. The extracts were analyzed via HPLC-MS and the resulting UV and mass chromatograms of the exconjugant extracts were compared to those of the empty control strain. In Fig. 42 the respective mass chromatograms for R5 cultures extracted with XAD-16 resin (A) and ethyl acetate (B) as well as the pellet extracts (C) are exemplarily shown, which are representative for further heterologous expression experiments. Additional peaks that were only observed in BGC carrying clones but not in the empty vector control were each investigated and evaluated manually as described before. All the investigated masses could also be detected in the control, although with a lower signal intensity.

A 7-day R5 cultures extracted with XAD-16 resin



B 7-day R5 cultures extracted ethyl acetate



C 7-day R5 cultures, pellet extraction

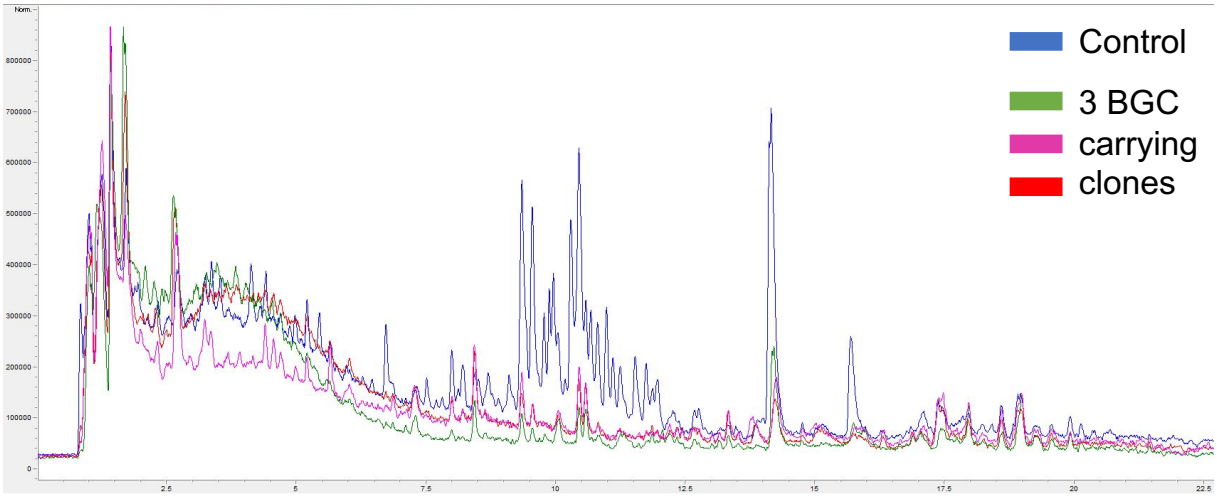


Fig. 42 HPLC-MS mass chromatograms of extracts derived from *S. coelicolor*/pCAP03_Cl76 and *S. coelicolor*/pCAP03 (empty vector control) cultures. 7-day R5 cultures extracted with XAD-16 resin **(A)**, ethyl acetate **(B)** and pellet extraction **(C)**. Blue, *S. coelicolor*/pCAP03 extracts; remaining colors, extracts of the three BGC carrying clones

Since the BGC seemed to be deleted in all Actinobacteria tested, a non-actinobacterial host was attempted for heterologous expression of the BGC. Therefore, the phosphopantetheinyl transferase (PPtase) gene carrying *E. coli* GB05-MtaA strain (Fu et al. 2012) was transformed with pCAP03_Cl76. A control was generated by transforming the strain with the empty pCAP03 vector. Three clones carrying pCAP03_Cl76 as well as the control strain carrying the empty pCAP03 vector were cultured in LB and M9 media respectively. After three days, culture supernatants and cell pellets were extracted separately using different extraction methods and extracts were analyzed via HPLC-MS, which did not reveal any new peaks for the BGC carrying strains.

6. Discussion

The rise of multi-drug resistant bacteria combined with the decreasing discovery rate of new antimicrobial drugs in the recent past demands for new antibiotics. Microorganisms from soil have already been a rich source of antibiotics and other medically relevant compounds for decades. Nevertheless, a large proportion of the biosynthetic diversity of soil bacteria remains unexplored due to their uncultivable nature under conventional laboratory conditions. However, this group of so far uncultivated microbes contains many bacteria that are very distantly related to the cultivated ones. Furthermore, their BGCs similarly often only show distant relation to characterized BGCs indicating that the corresponding encoded secondary metabolites also differ significantly from the known ones. Therefore, culture-independent metagenomic natural product discovery approaches provide a promising solution to access this previously missed potential and thus can lead to the discovery of new antibiotic classes for example. For that purpose, in a first step the metagenomic BGCs need to be identified and recovered from respective soil metagenomes to subsequently access the encoded compounds via heterologous expression. However, this first step often consists of laborious and time-consuming procedures such as PCR-based screening approaches of metagenomic libraries.

In this thesis, different soil sampling sites were investigated for their biosynthetic potential, followed by the development of an efficient strategy to identify and recover metagenomic BGCs for subsequent heterologous expression experiments aiming at the production of the encoded molecules.

The Schönbuch forest provides an optimal sampling location for comparing the biosynthetic potential of different soils as it harbors the three soil types podzol, cambisol and stagnosol in close proximity to each other, which ensures that they are all subjected to the same environmental conditions such as temperature or humidity. In this study, amplicon sequencing as well as Illumina shotgun sequencing were used to assess the biosynthetic potential of the different soil types. Analysis of generated amplicon sequences revealed that generally all samples harbored a huge unexplored biosynthetic potential that could not be captured completely as shown by rarefaction analysis (Mantri et al. 2021). Other studies in this field similarly investigated the biosynthetic capacity of soils from various locations using amplicon sequencing of KS and A domains. In contrast to the here conducted study, their soil samples did not consist of different soil types of the same location but of numerous topsoils from locations that were both, very close and very distant to each other. However, their main findings were similar, as they reported high numbers of unique biosynthetic domains for the majority of

the samples and in some cases similarly observed that rarefaction curves of biosynthetic domains did not reach saturation (Charlop-Powers et al. 2014; Charlop-Powers et al. 2015; Charlop-Powers et al. 2016). Thus, the biosynthetic capacity of the different soil types from the Schönbuch forest that was assessed by amplicon sequencing is in accordance with the previously reported unexplored biosynthetic diversity of soil metagenomes despite the major differences in the soil samples. In contrast to the studies by Charlop-Powers et al., within the here conducted study shotgun sequencing was additionally applied for the same samples, which revealed that the majority of domains detected using this method were not detected by amplicon sequencing. Thus, the combination of both methods revealed an even higher biosynthetic potential as a higher number of unique biosynthetic domains were detected as compared to each of the methods alone. Therefore, it can be concluded that a single method is most likely not sufficient for capturing the full biosynthetic capacity due to an individual bias of the respective method. However, it also has to be considered that none of the methods reached saturation because of limiting sequencing depth. The rarefaction analysis further revealed that with increasing sequencing depth, the soil types cambisol and stagnosol showed a higher number of OBUs as compared to podzol. Since preliminary experiments (data not shown) have shown that the isolation of pure metagenomic DNA from stagnosol samples is generally more challenging because of higher clay contents, cambisol was chosen for further experiments conducted in this study.

Previous studies have shown that the biosynthetic capacity of soil metagenomes can be captured with the construction of metagenomic libraries (Parsley et al. 2011; Reddy et al. 2012; Santana-Pereira et al. 2020), which is often the first step in culture-independent natural product discovery that finally aims at the recovery and heterologous expression of BGCs (Katz et al. 2016). The generation of metagenomic libraries as well as alternative methods aiming at the identification and recovery of complete BGCs such as PCR, next-generation sequencing or DNA cloning require the isolation of high-quality HMW DNA. However, this crucial procedure is especially challenging for soils as they contain numerous contaminants such as humic acids. The latter exhibit similar characteristics to DNA and are therefore isolated together with metagenomic DNA in many cases (Amorim et al. 2008; Sar et al. 2018), which can hinder the use of the aforementioned methods (Nair et al. 2014; Verma et al. 2017). Therefore, as a first step a protocol was developed to isolate HMW DNA of high quality. The purity of the metagenomic DNAs isolated from the A horizon of cambisol was assessed by 260/280 and 260/230 absorbance ratios. A 260/280 ratio not lower than 1.8 and a 260/230 ratio between 2.0 and 2.2 is considered to indicate pure DNA. While an abnormal 260/280 ratio is generally

known to indicate impurities such as protein or phenol contamination, for soil metagenomic DNA the 260/230 ratio can be used to assess contamination with humic acids (Ning et al. 2009). However, it has to be noted that the absorbance ratios generally serve as a first orientation, while using the DNA for desired downstream applications finally reveals if the purity is sufficient. Each of the three isolated HMW metagenomic DNAs showed an almost optimal 260/280 value, while the more critical 260/230 value varied between the isolations and was used to assess the different degrees of humic acid contamination. The first isolated DNA (metagenomic DNA 1) showed the highest degree of humic acid contamination, but nevertheless proved to be of sufficient quality for Nanopore sequencing. Further purification of that same DNA for Illumina sequencing yielded highly pure DNA, likely free of humic acid contamination (metagenomic DNA 2). The third metagenomic DNA that was isolated using a different method needed to be of a high molecular weight and pure at the same time to construct the main part of the metagenomic library. HMW metagenomic DNA 3 still showed slight humic acid contamination, but significantly improved as compared to HMW metagenomic DNA 1. As mentioned earlier, the quality of metagenomic DNAs isolated from different soil types and sampling locations can vary significantly due to differences in soil composition and thus concentration of contaminants. Therefore, the purity of metagenomic DNA obtained by different isolation methods cannot be compared and evaluated if two different soil types were used. This has also been shown in a past study that isolated metagenomic DNA from eight different soil samples consisting of six different types using a single method (Wnuk et al. 2020). The purity of the DNAs isolated by Wnuk et al. were equally assessed by both absorbance ratios. In contrast to the 260/280 ratios, the 260/230 ratios varied significantly depending on the soil type. Interestingly, one of the eight samples was forest soil of the cambisol type, which allows for a comparison with the value obtained within this thesis. After isolation and final purification of the DNA, Wnuk et al. measured a 260/230 ratio of 1.4. Although they did not aim for HMW DNA, the value still remained lower than the here measured value for the same soil type, which further confirms that comparatively pure HMW DNA was isolated within this thesis.

Agarose gel electrophoresis of metagenomic DNA isolated within this study confirmed its high molecular weight as a single band migrating above the lambda DNA/HindIII 23 kb band with almost no smear below was detected. Overall, these fulfilled prerequisites of high-quality metagenomic DNA enabled its use for the amplification of complete BGCs via PCR, the generation of a high-quality metagenomic fosmid library as well as the generation of high-quality long- and short-read sequencing data. These applications were the basis for the

following three strategies that enabled the recovery of complete BGCs from the cambisol metagenome.

The quality of the metagenomic fosmid library was assessed by releasing the inserts of ten randomly isolated fosmids via restriction digestion, which confirmed that large inserts have been captured. This procedure has also been applied in similar studies to investigate the insert size of soil metagenomic fosmid libraries (Costa et al. 2021). For the development of the first strategy, the metagenomic fosmid library was subsequently sequenced via Nanopore and Illumina. The long- and short-read sequencing data generated from the metagenomic fosmid library also confirmed to be of a high quality, especially for Nanopore sequencing as more than 367 thousand long-reads greater than 25 kb in size could be generated, which was crucial for the development of the SNRCM approach. The strategy of identifying complete BGCs for potential heterologous expression upon sequencing of the fosmid library proved successful as shown by the 34 complete BGCs detected on hybrid assembled contigs greater than 40 kb and more than 100 additional ones on contigs greater than 25 kb. The developed method SNRCM enabled the direct and fast identification of clones carrying complete BGCs by aligning complete BGCs detected on hybrid assembled contigs with single Nanopore reads. The efficiency could be validated with the identification of 15 BGCs on single fosmids among the 34 complete BGCs detected on contigs. The here developed approach proved to be more efficient than previous methods such as the classical screening of metagenomic libraries via PCR that remains a time-consuming multi-step process. For instance, in two past studies the cosmids of clones of interest from a soil metagenomic library were isolated, transferred to another host for heterologous expression and in one case additionally sequenced before it was revealed that the BGCs encoded on the cosmids were incomplete. The encoded compounds could therefore only be obtained after rescreening of the library for clones carrying the missing parts and subsequent assembly to a complete cluster via TAR cloning (Feng et al. 2010; Feng et al. 2011). In a similar study of the same group, cosmids of a soil metagenomic library carrying type II PKS genes were identified by a PCR screening using PKS specific degenerate primers. Positive cosmids were transferred to *Streptomyces albus* and resulting clones were screened for the heterologous production of new compounds (Bauer et al. 2010). Although in this case one clone produced a new compound and therefore seemed to carry a complete BGC, the screening process consisting of heterologous expression and extraction remains rather inefficient. The here developed method overcomes these problems as the identification of clones carrying complete BGCs does not even necessitate prior clone recovery. The method could be validated with the isolation of two lasso peptide BGCs on single fosmids and the subsequent

identification and resequencing of the lasso peptide specific genes, which confirmed completeness of the clusters. Although SNRCM only allows the recovery of BGCs that do not surpass the insert size of 40 kb from a single fosmid, natural products encoded by typically smaller BGCs such as RiPPs have been shown to contain medically relevant molecules including antibiotics (Waisvisz et al. 1957; Schmidt et al. 2005; Scholz et al. 2011). After the recovery of respective clones, these BGCs can directly be tested for heterologous expression in *E. coli* as no modification of the fosmids is required for that purpose.

As *E. coli* is also known to be capable of lasso peptide production, such as the strains producing the antibacterial lasso peptide microcin J25 (Bayro et al. 2003), the two recovered metagenomic lasso peptide BGCs on fosmids were first tested for heterologous expression in *E. coli*. For that purpose, *E. coli* was cultured in the commonly used M9 minimal medium to facilitate the detection of a lasso peptide potentially produced in low amounts. Since this attempt did not lead to the detection of respective lasso peptides, an alternative host that was potentially more suitable for lasso peptide production than *E. coli* was chosen for further heterologous expression attempts. The strain *Burkholderia* sp. FERM BP-3421 appeared to be an especially suitable heterologous host for the following reasons. In general, representatives of the *Burkholderiaceae* family have been proven in the past to be naturally talented producers of natural products as compounds of approx. 66 classes with various bioactivities have been discovered from bacteria of this family (Kunakom and Eustáquio 2019). Additionally, the genus *Burkholderia* is known to be capable of lasso peptide production such as the strain *Burkholderia thailandensis* E264, which produces capistruin (Knappe et al. 2008). Furthermore, *Burkholderia* sp. FERM BP-3421 has been previously shown to be able to heterologously produce capistruin in higher amounts than *E. coli* (Kunakom and Eustáquio 2020). To enable usage of this strain, the *Burkholderia* replicon (ori1600) was integrated into the backbone of the lasso peptide BGC carrying fosmids, followed by transfer to FERM BP-3421. Upon cultivation and transcription analysis, it was revealed that only BGC 482.1 was transcribed under various conditions, while BGC 40.1 was not. To activate transcription for the latter, the next logical step would normally be to clone the BGC in front of an inducible or constitutive promoter. However, since for BGC 40.1 the precursor peptide (A) gene could not be detected, this approach is not applicable. Therefore, in future experiments, the BGC could be tested in other hosts for transcription from the native promoter. This could be done by further genetically modifying the fosmid backbone to enable replication in other hosts such as *Streptomyces*. Although transcripts for lasso peptide specific genes were detected for BGC 482.1 heterologous expression experiments in *Burkholderia*, no respective lasso peptide could be found. As discussed in more in detail in the

following, a major challenge for the detection of heterologously produced lasso peptides is a low production yield in many cases. Therefore, different optimizations have been applied in the past to increase the production yield of various lasso peptides in *E. coli* that could also be tested for the metagenomic BGCs. Especially for BGC 482.1, where transcripts of lasso peptide genes were detected, a too low production amount might have hampered the detection of the compound. Since all genes show the same orientation, in a next step, the BGC could be cloned in front of a strong promoter to increase the production yield. It has been shown in past studies that cloning the A gene in front of a strong inducible promoter, while putting the remaining lasso peptide genes in front of a constitutive promoter can increase the production yield of lasso peptides heterologously expressed in *E. coli* (Hegemann et al. 2013a). This modification was routinely used for the heterologous expression of several lasso peptides in *E. coli* to increase the production yields (Cheung-Lee et al. 2019b; Cheung-Lee et al. 2019a; Cheung-Lee et al. 2020). Therefore, such modifications can be considered when cloning BGC 482.1 into an expression vector. Additionally, several studies have shown that separating multiple precursor peptide genes from a single lasso peptide BGC and expressing them in individual constructs increases the production yield (Hegemann et al. 2013b; Zimmermann et al. 2014; Hegemann et al. 2014). The effect could be nicely shown in a study that heterologously expressed a lasso peptide BGC of *Caulobacter segnis* in *E. coli*. The single BGC codes for the three different lasso peptides caulosegnin I, II, and III. Upon construction of single precursor constructs and expression in *E. coli* the production yield of caulosegnin I increased 15-fold as compared to the expression of the natural BGC containing all three precursors (Hegemann et al. 2013a). Since BGC 482.1 also carries more than one A gene, this modification could be included in future cloning experiments with the generation of two constructs carrying A1 and A2 respectively to further optimize the production yield. While these optimizations are specifically useful for BGC 482.1, further possible approaches to increase the production yield are representatively discussed for lasso peptide BGC 44.1 in the following.

The high-quality HMW metagenomic DNA isolated in this study opened up additional possibilities for directly recovering complete BGCs from soil metagenomes. It enabled long- and short-read sequencing of the DNA, followed by hybrid assembly into large contigs, which led to the detection of 113 BGCs on contigs greater than 25 kb. This allowed to use the second strategy for complete BGC recovery, i.e. the direct amplification of detected BGCs of interest via PCR, such as the here amplified and cloned lasso peptide BGC 44.1. Generally, this method enables cloning of the amplified BGCs into different expression vectors for different hosts as the required restriction sites for cloning can be easily attached, which further expands the

options of heterologous expression. Similarly, this method is limited to the recovery of small BGCs. However, the combination of both methods yielded a large number of small BGCs available for recovery and heterologous expression, although only a tiny part of the metagenome was covered due to a comparatively small library size and limiting sequencing depth. Upscaling of both is expected to yield a larger number of complete BGCs to choose from. Access to a large number of BGCs allows to prioritize for BGCs potentially encoding medically relevant molecules. Lasso peptides with antimicrobial activity for example often contain self-resistance conferring ABC transporter (D) genes within their BGCs (Hegemann et al. 2021), which makes the selection and recovery of those BGCs reasonable. The lasso peptide BGC 44.1 recovered in this study also contained a D gene, thus indicating potential antimicrobial activity. The system used for heterologous expression of BGC 44.1, i.e. induced expression from the PBAD promoter of pSK019 in *Burkholderia*, has been shown to be functional with the heterologous expression of the lasso peptide capistruin in a previous study (Kunakom and Eustáquio 2020). Nevertheless, using this system for heterologous expression of BGC 44.1 in *Burkholderia* as well as in *E. coli* did not lead to the detection of the respective lasso peptide. The transcription analysis in *Burkholderia* unexpectedly revealed that transcripts could only be detected for gene regions of B1 and B2 to C but not for the further downstream gene regions C and D. This finding was consistent for different cultivation times, indicating that the transcription might stop in the middle of the BGC for unknown reasons. In contrast, *E. coli* showed transcripts for all gene regions confirming the complete transcription of the BGC. However, it is noticeable that the more downstream lying genes C and D that could not be detected in *Burkholderia* show much lighter bands than the more upstream lying gene regions B1 and B2 to C in *E. coli*. Considering the fact that the *E. coli* origin of replication (ori) on pSK019 confers a higher copy number than the *Burkholderia* ori of the same plasmid, it can be hypothesized that the amount of transcript resulting from gene regions C and D is too low to be detectable in *Burkholderia*. To confirm this hypothesis, in future experiments the BGC can be cloned into a plasmid conferring a higher copy number in *Burkholderia*, followed by a transcription analysis. If the hypothesis proves to be true, the low amount of transcript would have led to an equally low translation rate in the following, which might explain the undetectability of the lasso peptide in *Burkholderia*. In contrast, the detected transcripts for all genes in *E. coli* imply that transcription was not the limiting factor here. Possible issues associated with translation include the native ribosome binding sites (RBSs) of the metagenomic BGC that might not have been recognized in *E. coli* or only with low efficiency. The latter case again would have led to very low production amounts preventing compound

detection. A past study has shown that the yield for heterologous expression of capistruin in *E. coli* could be significantly increased upon deletion of the intergenic region of genes A and B and subsequent introduction of an *E. coli* optimized RBS (Pan et al. 2012). Applying that same technique to the previously described single precursor construct of caulosegnin I led to an 11-fold higher yield in *E. coli* (Hegemann et al. 2013a) and was further used for the improved heterologous expression of other lasso peptides in *E. coli* (Zimmermann et al. 2014). The beneficial effect could be especially shown in a study that heterologously expressed several lasso peptide BGCs from different organisms in *E. coli* using the native BGCs as well as modified ones carrying the described RBS. A comparison of the yields revealed that lasso peptide production was increased in almost all cases with the modified BGCs. Remarkably, three of the lasso peptides were only detectable by using the constructs containing the *E. coli* RBS (Hegemann et al. 2013b). Future studies can therefore aim at exchanging the native RBSs of the metagenomic lasso peptide genes with *E. coli* RBSs. Since the cloned BGC 44.1 does not show the classical ABCD but a B1B2CAD organization, one RBS could be introduced upstream of the B1 gene and another one upstream of the D gene. In case this procedure is conducted by de novo synthesis of the BGC, the codon usage can be adjusted to *E. coli* at the same time as it has been previously done for the heterologous expression of the *Burkholderia ubonensis* derived lasso peptide ubonodin (Cheung-Lee et al. 2020) as well as the citrocin lasso peptide from *Citrobacter pasteurii* and *Citrobacter braakii* (Cheung-Lee et al. 2019b). Alternatively, the metagenomic BGC can be tested for expression from a constitutive promoter in *Streptomyces* as exemplified by a previous study that could heterologously produce the lasso peptide albusnodin in *Streptomyces* species but not in *E. coli* (Zong et al. 2018).

However, the issue could also be an analytical one. Especially if the lasso peptide was produced in very low amounts, standard high-performance liquid chromatography-mass spectrometry (HPLC-MS) might have failed in detecting the lasso peptide. Therefore, future studies can make use of alternative detection methods and techniques, which are currently applied to the extracts of the transcript positive *E. coli* strains. Alternatively, cultures could be supplemented with labeled amino acids that are known to be incorporated into the lasso peptide based on the predicted core peptide sequence, which would facilitate the detection of a produced lasso peptide.

The generated sequencing data of the metagenomic fosmid library as well as the corresponding metagenome allowed the use of a third strategy that enabled the efficient recovery of a larger BGC that was distributed over different fosmids. In past studies, the recovery of larger BGCs was performed by screening a library for clones of interest using degenerate primers for

biosynthetic domains as a first step. The library was then screened for clones carrying the corresponding and overlapping parts of the respective BGC of interest, followed by assembly of the cluster parts to a complete BGC (Hover et al. 2018; Wu et al. 2019; Stevenson et al. 2021). This method is dependent on the complete coverage of every potential BGC of interest within the library, which requires the labor-intensive construction of a saturating metagenomic library consisting of millions of clones. In contrast, the significantly smaller library size in this study is sufficient for recovering complete BGCs distributed over multiple fosmids, as the generated sequencing data directly reveals BGCs that are completely covered within the library. Even if a BGC of interest reveals to be only partially captured within the library, the sequencing data of the corresponding metagenome can be used to amplify missing parts for subsequent assembly to a complete cluster. Overall, these sequencing data can be utilized as a reference sequence to facilitate the recovery of respective BGCs as it has been demonstrated for the NRPS BGC in this study. Similarly, upscaling of the sequencing depth and clone number is expected to yield a greater number of large BGCs for subsequent prioritization.

Although a very small part of the NRPS BGC was not covered by the fosmid library contigs, end-sequencing of isolated fosmids positive for the left, middle and right part of the BGC revealed that they completely covered the cluster. This indicates that the library might not have been sequenced deep enough to cover it a hundred percent. The correct assembly of the NRPS BGC could be confirmed by Nanopore sequencing. However, upon transfer of the captured BGC into multiple established heterologous *Streptomyces* hosts, it could be observed that parts of the cluster were either immediately deleted or upon cultivation of the organisms. A possible explanation for this consistent pattern in all three *Streptomyces* species could be the rather distant relation of the source organism, presumably of the phylum *Acidobacteria*, and the host organisms belonging to *Actinobacteria*, which might hinder maintenance of the BGC. One such difference could be the difference in GC content. While the used *Streptomyces coelicolor* host for example has a genomic GC content of approx. 72 % (Borodina et al. 2005), the entire NRPS BGC 76.1 has a GC content of only 55.6 %. Alternatively, the BGC might code for a compound that is toxic to *Streptomyces*, thereby establishing a selection pressure for the loss of the BGC. Therefore, a non-actinobacterial host, namely *E. coli* GB05-MtaA (Fu et al. 2012) that carries a phosphopantetheinyl transferase (PPtase) gene necessary for NRP biosynthesis was tested for heterologous expression in two different media using different extraction methods, which did not lead to the detection of a new compound.

The pCAP03 TAR vector used to capture the BGC contains the genetic elements for transfer via conjugation and integration of the plasmid into several *Actinomycete* genomes such as the

here used *Streptomyces* species. Since it was revealed that these organisms were not suitable for that specific metagenomic NRPS BGC, in future experiments TAR cloning can be repeated using alternative available TAR capture vectors that would allow to test for expression in other hosts that might be more suitable such as *B. subtilis* or other proteobacterial host organisms (Zhang et al. 2019). Additionally, also here the BGC could be codon optimized for various hosts in the future aiming at the stable maintenance and heterologous expression of the BGC. Further engineering such as the introduction of constitutive promoters in front of relevant genes will be challenging for this particular BGC as the contained genes might not have been accurately identified and annotated by antiSMASH due to its distant relation to known BGCs. This was for instance indicated by the presence of incomplete modules and the identification of an additional main biosynthetic gene upon Blast analysis that was missed by antiSMASH (data not shown).

Therefore, in future efforts it might be more promising to focus on the identification of more suitable and phylogenetically more related hosts such as species of the *Acidobacteria* phylum, which would enhance the chances of the native promoters being recognized by the respective heterologous host (Huo et al. 2019). Species of the *Acidobacteria* phylum would be especially valuable for the heterologous expression of various soil metagenome derived BGCs as this phylum belongs to one of the most frequent phyla in soil (Giguere et al. 2021).

The overall discussed techniques and options for the successful heterologous expression of metagenomic BGCs will also be useful to increase actual expression rates, which currently represent one of the main bottlenecks in the production of natural products from metagenomes. This was especially exemplified in a recent study that heterologously expressed a metagenomic BGC in *Streptomyces albus*. The corresponding final product of the pathway metathramycin was initially missed because of very low production rates and could only be discovered by applying a combination of additional techniques such as pathway engineering and bioactivity guided fractionation (Stevenson et al. 2021). The latter will be especially useful for low abundance compounds with expected bioactivity, such as the lasso peptide encoded by BGC 44.1 that contains an ABC transporter gene indicating potential antimicrobial activity.

In summary, in this thesis an efficient approach for the identification and recovery of novel natural product BGCs from soil metagenomes was developed. Recovered metagenomic BGCs were tested in different heterologous expression experiments under various conditions, which yielded valuable information that can be used to adjust future experiments to finally achieve the heterologous production of novel compounds. Especially the successful heterologous

production of metagenomic lasso peptides in the future is expected to be promising as most known antimicrobial lasso peptides showed antibacterial activity against a broad spectrum of pathogens (Cheng and Hua 2020).

Overall, this thesis contributes to facilitate and accelerate future soil metagenomic natural product discovery efforts that can potentially lead to new therapeutically relevant compounds such as antibiotics.

7. Supplements

Table S1 Details of metagenomic fosmid library sequencing output for Illumina and Nanopore sequencing and hybrid assembly of both; adopted from (Negri et al. 2022)

Illumina sequencing	
Approximate amount of generated paired reads	more than 92 million
Size of short-read cumulative data	greater than 26 gigabases
Nanopore sequencing	
Size of long-read cumulative data	greater than 27 gigabases
Amount of Nanopore reads greater than 25 kb	more than 367.000
Hybrid assembly	
Amount of assembled contigs greater than 1 kb	2 million
Amount of assembled contigs greater than 25 kb	nearly 16.000

S2 Table 17 including Query Cover, E Value and Percent Identity for Blastx results; modified from (Negri et al. 2022)

Metagenomic lasso peptide BGC 40.1							
Gene	**Blastx results**	**Query Cover**	**E Value**	**Percent Identity**	**Gene size (bp)**	**antiSMASH annotation**	**Putative lasso peptide gene**
8	hypothetical protein [*Acidobacteriia bacterium*]	95%	3,00E-132	60.00%	975	-	-
9	hypothetical protein [*Acidobacteriia bacterium*]	99%	0.0	50.55%	1905	Asn_synthase	C gene
10	hypothetical protein [*Betaproteobacteria bacterium*]	86%	5,00E-28	48.00%	432	-	-
11	lasso peptide biosynthesis B2 protein [*Acidobacteriia bacterium*]	52%	2,00E-15	54.05%	423	PF13471	B2 gene
12	PqqD family peptide modification chaperone [*Acidobacteriia bacterium*]	98%	8,00E-36	60.82%	294	PF05402	B1 gene
13	erythromycin biosynthesis sensory transduction protein eryC1	94%	0.0	83.00%	1116	DegT_DnrJ_EryC1	-

	[*Acidobacteria bacterium*]						
Metagenomic lasso peptide BGC 482.1							
Gene	**Blastx results**	**Query Cover**	**E Value**	**Percent Identity**	**Gene size (bp)**	**antiSMASH annotation**	**Putative lasso peptide gene**
10.1	hypothetical protein DMG37_22385 [*Acidobacteria bacterium*]	90%	7,00E-28	61.45%	276	PF05402	B1 gene
10.2	PqqD family protein [*Acidobacteriia bacterium*]	89%	2,00E-19	53.66%			
11	aminoglycoside phosphotransferase family protein [*Acidobacteriia bacterium*]	99%	1,00E-122	42.61%	1404	-	Kinase
12	lasso peptide biosynthesis B2 protein [*Acidobacteria bacterium*]	99%	2,00E-63	74.19%	375	PF13471	B2 gene
13	hypothetical protein [*Acidobacteriia bacterium*]	95%	0.0	59.97%	1929	Asn_synthase	C gene
14	hypothetical protein [*Acidobacteria bacterium*]	97%	6,00E-04	41.30%	141	-	A1 gene
15	hypothetical protein [*Acidobacteria bacterium*]	97%	1,00E-06	47.83%	141	-	A2 gene
16	hypothetical protein DMG78_32005 [*Acidobacteria bacterium*]	79%	0.0	65.59%	2118	-	-

S3 Table 18 including Query Cover, E Value and Percent Identity for Blastx results; modified from (Negri et al. 2022)

BGC0001176: streptomonomicin biosynthetic gene cluster from *Streptomonospora alba*						
Gene	**Blastx results**	**Query Cover**	**E Value**	**Percent Identity**	**Gene size (bp)**	**antiSMASH annotation**
B1	hypothetical protein LP52_05070 [*Streptomonospora alba*]	98%	3,00E-54	100.00%	255	PF05402
	PqqD family peptide modification chaperone [*Streptomonospora alba*]	98%	3,00E-54	100.00%		
BGC0001356: paeninodin biosynthetic gene cluster from *Paenibacillus dendritiformis C454*						
Gene	**Blastx results**	**Query Cover**	**E Value**	**Percent Identity**	**Gene size (bp)**	**antiSMASH annotation**
B1	lasso peptide biosynthesis PqqD family chaperone [*Paenibacillus dendritiformis*]	99%	5,00E-54	100.00%	300	PF05402
	lasso peptide biosynthesis PqqD family chaperone [*Paenibacillus dendritiformis*]	99%	3,00E-53	97.98%		

S4 Table 19 including Query Cover, E Value and Percent Identity for Blastx results; modified from (Negri et al. 2022)

Metagenomic lasso peptide BGC 44.1							
Gene	**Blastx results**	**Query Cover**	**E Value**	**Percent Identity**	**Gene size (bp)**	**antiSMASH annotation**	**Putative lasso peptide gene**
8	ABC transporter permease [*Acidobacteriia bacterium*]	98%	2,00E-134	80.37%	822	-	D1 gene
9	ABC transporter ATP-binding protein [*Acidobacteriia bacterium*]	99%	4,00E-126	69.52%	810	ABC transporter ATP-binding protein	D2 gene
10	PqqD family peptide modification chaperone [*Acidobacteriia bacterium*]	97%	3,00E-43	71.43%	303	PF05402	B1 gene
11	lasso peptide biosynthesis B2 protein [*Acidobacteriia bacterium*]	96%	2,00E-38	54.61%	471	PF13471	B2 gene
12	asparagine synthetase B [*Acidobacteriia bacterium*]	99%	0.0	64.19%	1863	Asn_synthase	C gene
13	hypothetical protein DMG36_15005 [*Acidobacteria bacterium*]	80%	7,00E-09	65.00%	150	predicted lasso peptide	A gene

8. Acknowledgements

First and foremost I want to thank my supervisor Prof. Dr. Nadine Ziemert for providing me the opportunity to conduct my PhD thesis at her lab and for her guidance during that time. I especially appreciated the support that I experienced during my entire time as a PhD student, which allowed me to learn a lot.

I also want to thank my second supervisor Dr. Evi Stegmann for the additional support she provided during my work on my PhD project.

Furthermore, I want to thank Dr. Günther Muth for always taking the time to discuss challenges and providing his advice.

Thanks to the members of the Ziemert lab for being nice colleagues and always ready to help. Special thanks to Dr. Shrikant Mantri for the pleasant cooperation.

Thanks to Dr. Alessandra Eustáquio for the collaboration and the support she provided.

Thanks to Dr. Libera Lo Presti for her advice on scientific writing.

Thanks to the State Postgraduate Fellowship Programme (Landesgraduiertenförderung) and the DFG for funding.

9. References

Alanjary M, Kronmiller B, Adamek M, Blin K, Weber T, Huson D, Philmus B, Ziemert N (2017) The Antibiotic Resistant Target Seeker (ARTS), an exploration engine for antibiotic cluster prioritization and novel drug target discovery. Nucleic Acids Res 45:W42–W48 . https://doi.org/10.1093/nar/gkx360

Amorim JH, Macena TNS, Lacerda-Junior GV, Rezende RP, Dias JCT, Brendel M, Cascardo JCM (2008) An improved extraction protocol for metagenomic DNA from a soil of the brazilian atlantic rainforest. Genet Mol Res 7:1226–1232 . https://doi.org/10.4238/vol7-4gmr509

Amos GCA, Borsetto C, Laskaris P, Krsek M, Berry AE, Newsham KK, Calvo-Bado L, Pearce DA, Vallin C, Wellington EMH (2015) Designing and implementing an assay for the detection of rare and divergent NRPS and PKS clones in european, antarctic and cuban soils. PLoS One 10:e0138327 . https://doi.org/10.1371/journal.pone.0138327

Arnison PG, Bibb MJ, Bierbaum G, Bowers AA, Bugni TS, Bulaj G, Camarero JA, Campopiano DJ, Challis GL, Clardy J, Cotter PD, Craik DJ, Dawson M, Dittmann E, Donadio S, Dorrestein PC, Entian K-D, Fischbach MA, Garavelli JS, Göransson U, Gruber CW, Haft DH, Hemscheidt TK, Hertweck C, Hill C, Horswill AR, Jaspars M, Kelly WL, Klinman JP, Kuipers OP, Link AJ, Liu W, Marahiel MA, Mitchell DA, Moll GN, Moore BS, Müller R, Nair SK, Nes IF, Norris GE, Olivera BM, Onaka H, Patchett ML, Piel J, Reaney MJT, Rebuffat S, Ross RP, Sahl H-G, Schmidt EW, Selsted ME, Severinov K, Shen B, Sivonen K, Smith L, Stein T, Süssmuth RD, Tagg JR, Tang G-L, Truman AW, Vederas JC, Walsh CT, Walton JD, Wenzel SC, Willey JM, van der Donk WA (2013) Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. Nat Prod Rep 30:108–160 . https://doi.org/10.1039/C2NP20085F

Bachmann BO, Van Lanen SG, Baltz RH (2014) Microbial genome mining for accelerated natural products discovery: is a renaissance in the making? J Ind Microbiol Biotechnol 41:175–184 . https://doi.org/10.1007/s10295-013-1389-9

Baltz RH (2006) Marcel Faber Roundtable: Is our antibiotic pipeline unproductive because of starvation, constipation or lack of inspiration? J Ind Microbiol Biotechnol 33:507–513 . https://doi.org/10.1007/s10295-005-0077-9

Baltz RH (2016) Genetic manipulation of secondary metabolite biosynthesis for improved production in Streptomyces and other actinomycetes. J Ind Microbiol Biotechnol 43:343–370 . https://doi.org/10.1007/s10295-015-1682-x

Bauer JD, King RW, Brady SF (2010) Utahmycins A and B, azaquinones produced by an environmental DNA clone. J Nat Prod 73:976–979 . https://doi.org/10.1021/np900786s

Bayro MJ, Mukhopadhyay J, Swapna GVT, Huang JY, Ma L-C, Sineva E, Dawson PE, Montelione GT, Ebright RH (2003) Structure of Antibacterial Peptide Microcin J25: A 21-Residue Lariat Protoknot. J Am Chem Soc 125:12382–12383 . https://doi.org/10.1021/ja036677e

Bentley SD, Chater KF, Cerdeño-Tárraga A-M, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D, Bateman A, Brown S, Chandra G, Chen CW, Collins M, Cronin A, Fraser A, Goble A, Hidalgo J, Hornsby T, Howarth S, Huang C-H, Kieser T, Larke L, Murphy L, Oliver K, O'Neil S, Rabbinowitsch E, Rajandream M-A, Rutherford K, Rutter S, Seeger K, Saunders D, Sharp S, Squares R, Squares S, Taylor K, Warren T, Wietzorrek A, Woodward J, Barrell BG, Parkhill J, Hopwood DA (2002) Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). Nature 417:141–147 . https://doi.org/10.1038/417141a

Bérdy J (2012) Thoughts and facts about antibiotics: Where we are now and where we are heading. J Antibiot (Tokyo) 65:385–395 . https://doi.org/10.1038/ja.2012.27

Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T (2019) antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res 47:W81–W87 . https://doi.org/10.1093/nar/gkz310

Borodina I, Krabben P, Nielsen J (2005) Genome-scale analysis of Streptomyces coelicolor A3(2) metabolism. Genome Res 15:820–829 . https://doi.org/10.1101/gr.3364705

Brachmann AO, Joyce SA, Jenke-Kodama H, Schwär G, Clarke DJ, Bode HB (2007) A type II polyketide synthase is responsible for anthraquinone biosynthesis in Photorhabdus luminescens. ChemBioChem 8:1721–1728 . https://doi.org/10.1002/cbic.200700300

Brady SF (2007) Construction of soil environmental DNA cosmid libraries and screening for clones that produce biologically active small molecules. Nat Protoc 2:1297–1305 . https://doi.org/10.1038/nprot.2007.195

Brady SF, Chao CJ, Clardy J (2002) New natural product families from an environmental DNA (eDNA) gene cluster. J Am Chem Soc 124:9968–9969 . https://doi.org/10.1021/ja0268985

Brady SF, Chao CJ, Clardy J (2004) Long-chain N-acyltyrosine synthases from environmental DNA. Appl Environ Microbiol 70:6865–6870 . https://doi.org/10.1128/AEM.70.11.6865-6870.2004

Brady SF, Chao CJ, Handelsman J, Clardy J (2001) Cloning and heterologous expression of a

natural product biosynthetic gene cluster from eDNA. Org Lett 3:1981–1984 . https://doi.org/10.1021/ol015949k

Breton RC, Reynolds WF (2013) Using NMR to identify and characterize natural products. Nat Prod Rep 30:501 . https://doi.org/10.1039/c2np20104f

Caboche S, Pupin M, Leclere V, Fontaine A, Jacques P, Kucherov G (2007) NORINE: a database of nonribosomal peptides. Nucleic Acids Res 36:D326–D331 . https://doi.org/10.1093/nar/gkm792

Challis GL (2014) Exploitation of the Streptomyces coelicolor A3(2) genome sequence for discovery of new natural products and biosynthetic pathways. J Ind Microbiol Biotechnol 41:219–232 . https://doi.org/10.1007/s10295-013-1383-2

Charlop-Powers Z, Owen JG, Reddy BVB, Ternei MA, Brady SF (2014) Chemical-biogeographic survey of secondary metabolism in soil. Proc Natl Acad Sci 111:3757–3762 . https://doi.org/10.1073/pnas.1318021111

Charlop-Powers Z, Owen JG, Reddy BVB, Ternei MA, Guimarães DO, de Frias UA, Pupo MT, Seepe P, Feng Z, Brady SF (2015) Global biogeographic sampling of bacterial secondary metabolism. Elife 4: . https://doi.org/10.7554/eLife.05048

Charlop-Powers Z, Pregitzer CC, Lemetre C, Ternei MA, Maniko J, Hover BM, Calle PY, McGuire KL, Garbarino J, Forgione HM, Charlop-Powers S, Brady SF (2016) Urban park soil microbiomes are a rich reservoir of natural product biosynthetic diversity. Proc Natl Acad Sci 113:14811–14816 . https://doi.org/10.1073/pnas.1615581113

Cheng C, Hua Z-C (2020) Lasso peptides: Heterologous production and potential medical application. Front Bioeng Biotechnol 8: . https://doi.org/10.3389/fbioe.2020.571165

Cheung-Lee WL, Cao L, Link AJ (2019a) Pandonodin: A Proteobacterial Lasso Peptide with an Exceptionally Long C-Terminal Tail. ACS Chem Biol 14:2783–2792 . https://doi.org/10.1021/acschembio.9b00676

Cheung-Lee WL, Parry ME, Jaramillo Cartagena A, Darst SA, Link AJ (2019b) Discovery and structure of the antimicrobial lasso peptide citrocin. J Biol Chem 294:6822–6830 . https://doi.org/10.1074/jbc.RA118.006494

Cheung-Lee WL, Parry ME, Zong C, Cartagena AJ, Darst SA, Connell ND, Russo R, Link AJ (2020) Discovery of Ubonodin, an Antimicrobial Lasso Peptide Active against Members of the Burkholderia cepacia Complex. ChemBioChem 21:1335–1340 . https://doi.org/10.1002/cbic.201900707

Church NA, McKillip JL (2021) Antibiotic resistance crisis: challenges and imperatives. Biologia (Bratisl) 76:1535–1550 . https://doi.org/10.1007/s11756-021-00697-x

Clatworthy AE, Pierson E, Hung DT (2007) Targeting virulence: a new paradigm for antimicrobial therapy. Nat Chem Biol 3:541–548 . https://doi.org/10.1038/nchembio.2007.24

Condurso HL, Bruner SD (2012) Structure and noncanonical chemistry of nonribosomal peptide biosynthetic machinery. Nat Prod Rep 29:1099 . https://doi.org/10.1039/c2np20023f

Costa ÂMA, Santos AO, Sousa J, Rodrigues JL, Gudiña EJ, Silvério SC, Rodrigues LR (2021) Improved method for the extraction of high-quality DNA from lignocellulosic compost samples for metagenomic studies. Appl Microbiol Biotechnol 105:8881–8893 . https://doi.org/10.1007/s00253-021-11647-7

Craig JW, Chang F-Y, Kim JH, Obiajulu SC, Brady SF (2010) Expanding small-molecule functional metagenomics through parallel screening of broad-host-range cosmid environmental DNA libraries in diverse proteobacteria. Appl Environ Microbiol 76:1633–1641 . https://doi.org/10.1128/AEM.02169-09

Crits-Christoph A, Diamond S, Butterfield CN, Thomas BC, Banfield JF (2018) Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis. Nature 558:440–444 . https://doi.org/10.1038/s41586-018-0207-y

Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc Natl Acad Sci 97:6640–6645 . https://doi.org/10.1073/pnas.120163297

Delgado MA, Rintoul MR, Farías RN, Salomón RA (2001) Escherichia coli RNA polymerase is the target of the cyclopeptide antibiotic microcin J25. J Bacteriol 183:4543–4550 . https://doi.org/10.1128/JB.183.15.4543-4550.2001

Eustáquio AS, Chang L-P, Steele GL, O′Donnell CJ, Koehn FE (2016) Biosynthetic engineering and fermentation media development leads to gram-scale production of spliceostatin natural products in Burkholderia sp. Metab Eng 33:67–75 . https://doi.org/10.1016/j.ymben.2015.11.003

Feng Z, Kallifidas D, Brady SF (2011) Functional analysis of environmental DNA-derived type II polyketide synthases reveals structurally diverse secondary metabolites. Proc Natl Acad Sci 108:12629–12634 . https://doi.org/10.1073/pnas.1103921108

Feng Z, Kim JH, Brady SF (2010) Fluostatins Produced by the Heterologous Expression of a TAR Reassembled Environmental DNA Derived Type II PKS Gene Cluster. J Am Chem Soc 132:11902–11903 . https://doi.org/10.1021/ja104550p

Fischbach MA, Walsh CT (2006) Assembly-line enzymology for polyketide and

nonribosomal peptide antibiotics: Logic, machinery, and mechanisms. Chem Rev 106:3468–3496 . https://doi.org/10.1021/cr0503097

Frank A, Groll M (2017) The methylerythritol phosphate pathway to isoprenoids. Chem Rev 117:5675–5703 . https://doi.org/10.1021/acs.chemrev.6b00537

Franke J, Ishida K, Hertweck C (2012) Genomics-driven discovery of Burkholderic acid, a noncanonical, cryptic polyketide from human pathogenic Burkholderia species. Angew Chemie Int Ed 51:11611–11615 . https://doi.org/10.1002/anie.201205566

Fu J, Bian X, Hu S, Wang H, Huang F, Seibert PM, Plaza A, Xia L, Müller R, Stewart AF, Zhang Y (2012) Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. Nat Biotechnol 30:440–446 . https://doi.org/10.1038/nbt.2183

Giguere AT, Eichorst SA, Meier D V., Herbold CW, Richter A, Greening C, Woebken D (2021) Acidobacteria are active and abundant members of diverse atmospheric H2-oxidizing communities detected in temperate soils. ISME J 15:363–376 . https://doi.org/10.1038/s41396-020-00750-8

Gillespie DE, Brady SF, Bettermann AD, Cianciotto NP, Liles MR, Rondon MR, Clardy J, Goodman RM, Handelsman J (2002) Isolation of antibiotics Turbomycin A and B from a metagenomic library of soil microbial DNA. Appl Environ Microbiol 68:4301–4306 . https://doi.org/10.1128/AEM.68.9.4301-4306.2002

Ginolhac A, Jarrin C, Gillet B, Robe P, Pujic P, Tuphile K, Bertrand H, Vogel TM, Perrière G, Simonet P, Nalin R (2004) Phylogenetic analysis of polyketide synthase I domains from soil metagenomic libraries allows selection of promising clones. Appl Environ Microbiol 70:5522–5527 . https://doi.org/10.1128/AEM.70.9.5522-5527.2004

Gomez-Escribano JP, Bibb MJ (2011) Engineering Streptomyces coelicolor for heterologous expression of secondary metabolite gene clusters. Microb Biotechnol 4:207–215 . https://doi.org/10.1111/j.1751-7915.2010.00219.x

Gust B, Challis GL, Fowler K, Kieser T, Chater KF (2003) PCR-targeted Streptomyces gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. Proc Natl Acad Sci 100:1541–1546 . https://doi.org/10.1073/pnas.0337542100

Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem Biol 5:R245–R249 . https://doi.org/10.1016/S1074-5521(98)90108-9

Havlicek V, Lemr K, Schug KA (2013) Current trends in microbial diagnostics based on mass

spectrometry. Anal Chem 85:790–797 . https://doi.org/10.1021/ac3031866

Hegemann JD, Jeanne Dit Fouque K, Santos-Fernandez M, Fernandez-Lima F (2021) A bifunctional leader peptidase/ABC transporter protein is involved in the maturation of the lasso peptide cochonodin I from Streptococcus suis. J Nat Prod 84:2683–2691 . https://doi.org/10.1021/acs.jnatprod.1c00514

Hegemann JD, Zimmermann M, Xie X, Marahiel MA (2015) Lasso peptides: An intriguing class of bacterial natural products. Acc Chem Res 48:1909–1919 . https://doi.org/10.1021/acs.accounts.5b00156

Hegemann JD, Zimmermann M, Xie X, Marahiel MA (2013a) Caulosegnins I–III: A Highly Diverse Group of Lasso Peptides Derived from a Single Biosynthetic Gene Cluster. J Am Chem Soc 135:210–222 . https://doi.org/10.1021/ja308173b

Hegemann JD, Zimmermann M, Zhu S, Klug D, Marahiel MA (2013b) Lasso peptides from proteobacteria: Genome mining employing heterologous expression and mass spectrometry. Biopolymers 100:527–542 . https://doi.org/10.1002/bip.22326

Hegemann JD, Zimmermann M, Zhu S, Steuber H, Harms K, Xie X, Marahiel MA (2014) Xanthomonins I-III: A New Class of Lasso Peptides with a Seven-Residue Macrolactam Ring. Angew Chemie Int Ed 53:2230–2234 . https://doi.org/10.1002/anie.201309267

Hendlin D, Stapley EO, Jackson M, Wallick H, Miller AK, Wolf FJ, Miller TW, Chaiet L, Kahan FM, Foltz EL, Woodruff HB, Mata JM, Hernandez S, Mochales S (1969) Phosphonomycin, a New Antibiotic Produced by Strains of Streptomyces. Science (80-) 166:122–123 . https://doi.org/10.1126/science.166.3901.122

Hertweck C (2009) The biosynthetic logic of polyketide diversity. Angew Chemie Int Ed 48:4688–4716 . https://doi.org/10.1002/anie.200806121

Hertweck C, Luzhetskyy A, Rebets Y, Bechthold A (2007) Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. Nat Prod Rep 24:162–190 . https://doi.org/10.1039/B507395M

Hover BM, Kim S-H, Katz M, Charlop-Powers Z, Owen JG, Ternei MA, Maniko J, Estrela AB, Molina H, Park S, Perlin DS, Brady SF (2018) Culture-independent discovery of the malacidins as calcium-dependent antibiotics with activity against multidrug-resistant Gram-positive pathogens. Nat Microbiol 3:415–422 . https://doi.org/10.1038/s41564-018-0110-1

Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hernsdorf AW, Amano Y, Ise K, Suzuki Y, Dudek N, Relman DA, Finstad KM, Amundson R, Thomas BC, Banfield JF (2016) A new view of the tree of life. Nat

Microbiol 1:16048 . https://doi.org/10.1038/nmicrobiol.2016.48

Huo L, Hug JJ, Fu C, Bian X, Zhang Y, Müller R (2019) Heterologous expression of bacterial natural product biosynthetic pathways. Nat Prod Rep 36:1412–1436 . https://doi.org/10.1039/C8NP00091C

Hutchings MI, Truman AW, Wilkinson B (2019) Antibiotics: past, present and future. Curr Opin Microbiol 51:72–80 . https://doi.org/10.1016/j.mib.2019.10.008

I. Tietz J, A. Mitchell D (2016) Using genomics for natural product structure elucidation. Curr Top Med Chem 16:1645–1694 . https://doi.org/10.2174/1568026616666151012111439

Ikeda H, Ishikawa J, Hanamoto A, Shinose M, Kikuchi H, Shiba T, Sakaki Y, Hattori M, Ōmura S (2003) Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis. Nat Biotechnol 21:526–531 . https://doi.org/10.1038/nbt820

Ikeda H, Shin-ya K, Omura S (2014) Genome mining of the Streptomyces avermitilis genome and development of genome-minimized hosts for heterologous expression of biosynthetic gene clusters. J Ind Microbiol Biotechnol 41:233–250 . https://doi.org/10.1007/s10295-013-1327-x

Katz L, Baltz RH (2016) Natural product discovery: past, present, and future. J Ind Microbiol Biotechnol 43:155–176 . https://doi.org/10.1007/s10295-015-1723-5

Katz M, Hover BM, Brady SF (2016) Culture-independent discovery of natural products from soil metagenomes. J Ind Microbiol Biotechnol 43:129–141 . https://doi.org/10.1007/s10295-015-1706-6

Kautsar SA, Blin K, Shaw S, Navarro-Muñoz JC, Terlouw BR, van der Hooft JJJ, van Santen JA, Tracanna V, Suarez Duran HG, Pascal Andreu V, Selem-Mojica N, Alanjary M, Robinson SL, Lund G, Epstein SC, Sisto AC, Charkoudian LK, Collemare J, Linington RG, Weber T, Medema MH (2019) MIBiG 2.0: a repository for biosynthetic gene clusters of known function. Nucleic Acids Res. https://doi.org/10.1093/nar/gkz882

Keatinge-Clay AT (2012) The structures of type I polyketide synthases. Nat Prod Rep 29:1050 . https://doi.org/10.1039/c2np20019h

Kellenberger E (2001) Exploring the unknown. EMBO Rep 2:5–7 . https://doi.org/10.1093/embo-reports/kve014

Kittilä T, Mollo A, Charkoudian LK, Cryle MJ (2016) New structural data reveal the motion of carrier proteins in nonribosomal peptide synthesis. Angew Chemie Int Ed 55:9834–9840 . https://doi.org/10.1002/anie.201602614

Knappe TA, Linne U, Xie X, Marahiel MA (2010) The glucagon receptor antagonist BI-

32169 constitutes a new class of lasso peptides. FEBS Lett 584:785–789 . https://doi.org/10.1016/j.febslet.2009.12.046

Knappe TA, Linne U, Zirah S, Rebuffat S, Xie X, Marahiel MA (2008) Isolation and structural characterization of capistruin, a lasso peptide predicted from the genome sequence of Burkholderia thailandensis E264. J Am Chem Soc 130:11446–11454 . https://doi.org/10.1021/ja802966g

Kopp F, Marahiel MA (2007) Macrocyclization strategies in polyketide and nonribosomal peptide biosynthesis. Nat Prod Rep 24:735 . https://doi.org/10.1039/b613652b

Kunakom S, Eustáquio AS (2020) Heterologous production of lasso peptide capistruin in a Burkholderia host. ACS Synth Biol 9:241–248 . https://doi.org/10.1021/acssynbio.9b00438

Kunakom S, Eustáquio AS (2019) Burkholderia as a Source of Natural Products. J Nat Prod 82:2018–2037 . https://doi.org/10.1021/acs.jnatprod.8b01068

Kuznedelov K, Semenova E, Knappe TA, Mukhamedyarov D, Srivastava A, Chatterjee S, Ebright RH, Marahiel MA, Severinov K (2011) The antibacterial threaded-lasso peptide capistruin inhibits cacterial RNA polymerase. J Mol Biol 412:842–848 . https://doi.org/10.1016/j.jmb.2011.02.060

Lange BM (2016) Online resources for gene discovery and biochemical research with aromatic and medicinal plants. Phytochem Rev 15:489–510 . https://doi.org/10.1007/s11101-015-9450-0

Ling LL, Schneider T, Peoples AJ, Spoering AL, Engels I, Conlon BP, Mueller A, Schäberle TF, Hughes DE, Epstein S, Jones M, Lazarides L, Steadman VA, Cohen DR, Felix CR, Fetterman KA, Millett WP, Nitti AG, Zullo AM, Chen C, Lewis K (2015) A new antibiotic kills pathogens without detectable resistance. Nature 517:455–459 . https://doi.org/10.1038/nature14098

Lombard J, Moreira D (2011) Origins and early evolution of the mevalonate pathway of isoprenoid biosynthesis in the three domains of life. Mol Biol Evol 28:87–99 . https://doi.org/10.1093/molbev/msq177

Lunde CS, Kubo I (2000) Effect of polygodial on the mitochondrial ATPase of Saccharomyces cerevisiae. Antimicrob Agents Chemother 44:1943–1953 . https://doi.org/10.1128/AAC.44.7.1943-1953.2000

MacNeil IA, Tiong CL, Minor C, August PR, Grossman TH, Loiacono KA, Lynch BA, Phillips T, Narula S, Sundaramoorthi R, Tyler A, Aldredge T, Long H, Gilman M, Holt D, Osburne MS (2001) Expression and isolation of antimicrobial small molecules from

soil DNA libraries. J Mol Microbiol Biotechnol 3:301–8

Madhavan A, Sindhu R, Parameswaran B, Sukumaran RK, Pandey A (2017) Metagenome analysis: a powerful tool for enzyme bioprospecting. Appl Biochem Biotechnol 183:636–651 . https://doi.org/10.1007/s12010-017-2568-3

Maksimov MO, Link AJ (2014) Prospecting genomes for lasso peptides. J Ind Microbiol Biotechnol 41:333–344 . https://doi.org/10.1007/s10295-013-1357-4

Maksimov MO, Pan SJ, James Link A (2012) Lasso peptides: structure, function, biosynthesis, and engineering. Nat Prod Rep 29:996 . https://doi.org/10.1039/c2np20070h

Mantri SS, Negri T, Sales-Ortells H, Angelov A, Peter S, Neidhardt H, Oelmann Y, Ziemert N (2021) Metagenomic sequencing of multiple soil horizons and sites in close vicinity revealed novel secondary metabolite diversity. mSystems. https://doi.org/10.1128/mSystems.01018-21

Martens E, Demain AL (2017) The antibiotic resistance crisis, with a focus on the United States. J Antibiot (Tokyo) 70:520–526 . https://doi.org/10.1038/ja.2017.30

Martin-Gómez H, Tulla-Puche J (2018) Lasso peptides: chemical approaches and structural elucidation. Org Biomol Chem 16:5065–5080 . https://doi.org/10.1039/C8OB01304G

McErlean M, Overbay J, Van Lanen S (2019) Refining and expanding nonribosomal peptide synthetase function and mechanism. J Ind Microbiol Biotechnol 46:493–513 . https://doi.org/10.1007/s10295-018-02130-w

Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R (2011) antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res 39:W339–W346 . https://doi.org/10.1093/nar/gkr466

Medema MH, Fischbach MA (2015) Computational approaches to natural product discovery. Nat Chem Biol 11:639–648 . https://doi.org/10.1038/nchembio.1884

Miyanaga A (2019) Michael additions in polyketide biosynthesis. Nat Prod Rep 36:531–547 . https://doi.org/10.1039/C8NP00071A

Moore BS, Hopke JN (2001) Discovery of a new bacterial polyketide biosynthetic pathway. ChemBioChem 2:35–38 . https://doi.org/10.1002/1439-7633(20010105)2:1<35::AID-CBIC35>3.0.CO;2-1

Mungan MD, Alanjary M, Blin K, Weber T, Medema MH, Ziemert N (2020) ARTS 2.0: feature updates and expansion of the Antibiotic Resistant Target Seeker for comparative genome mining. Nucleic Acids Res 48:W546–W552 .

https://doi.org/10.1093/nar/gkaa374

Myronovskyi M, Rosenkränzer B, Nadmid S, Pujic P, Normand P, Luzhetskyy A (2018) Generation of a cluster-free Streptomyces albus chassis strains for improved heterologous expression of secondary metabolite clusters. Metab Eng 49:316–324 . https://doi.org/10.1016/j.ymben.2018.09.004

Nair HP, Vincent H, Bhat SG (2014) Evaluation of five in situ lysis protocols for PCR amenable metagenomic DNA from mangrove soils. Biotechnol Reports 4:134–138 . https://doi.org/10.1016/j.btre.2014.09.008

Negri T, Mantri S, Angelov A, Peter S, Muth G, Eustáquio AS, Ziemert N (2022) A rapid and efficient strategy to identify and recover biosynthetic gene clusters from soil metagenomes. Appl Microbiol Biotechnol. https://doi.org/10.1007/s00253-022-11917-y

Newman DJ, Cragg GM (2020) Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. J Nat Prod 83:770–803 . https://doi.org/10.1021/acs.jnatprod.9b01285

Nichols D, Cahoon N, Trakhtenberg EM, Pham L, Mehta A, Belanger A, Kanigan T, Lewis K, Epstein SS (2010) Use of Ichip for High-Throughput In Situ Cultivation of "Uncultivable" Microbial Species. Appl Environ Microbiol 76:2445–2450 . https://doi.org/10.1128/AEM.01754-09

Ning J, Liebich J, Kästner M, Zhou J, Schäffer A, Burauel P (2009) Different influences of DNA purity indices and quantity on PCR-based DGGE and functional gene microarray in soil microbial community study. Appl Microbiol Biotechnol 82:983–993 . https://doi.org/10.1007/s00253-009-1912-0

Nurk S, Meleshko D, Korobeynikov A, Pevzner PA (2017) metaSPAdes: a new versatile metagenomic assembler. Genome Res 27:824–834 . https://doi.org/10.1101/gr.213959.116

Ogasawara Y, Yackley BJ, Greenberg JA, Rogelj S, Melançon CE (2015) Expanding our understanding of sequence-function relationships of type II polyketide biosynthetic gene clusters: Bioinformatics-guided identification of frankiamicin A from Frankia sp. EAN1pec. PLoS One 10:e0121505 . https://doi.org/10.1371/journal.pone.0121505

Ortega MA, van der Donk WA (2016) New insights into the biosynthetic logic of ribosomally synthesized and post-translationally modified peptide natural products. Cell Chem Biol 23:31–44 . https://doi.org/10.1016/j.chembiol.2015.11.012

Owen JG, Charlop-Powers Z, Smith AG, Ternei MA, Calle PY, Reddy BVB, Montiel D, Brady SF (2015) Multiplexed metagenome mining using short DNA sequence tags

facilitates targeted discovery of epoxyketone proteasome inhibitors. Proc Natl Acad Sci 112:4221–4226 . https://doi.org/10.1073/pnas.1501124112

Palumbi SR (2001) Humans as the World's Greatest Evolutionary Force. Science (80-) 293:1786–1790 . https://doi.org/10.1126/science.293.5536.1786

Pan SJ, Rajniak J, Maksimov MO, Link AJ (2012) The role of a conserved threonine residue in the leader peptide of lasso peptide precursors. Chem Commun 48:1880 . https://doi.org/10.1039/c2cc17211a

Parsley LC, Linneman J, Goode AM, Becklund K, George I, Goodman RM, Lopanik NB, Liles MR (2011) Polyketide synthase pathways identified from a metagenomic library are derived from soil Acidobacteria. FEMS Microbiol Ecol 78:176–187 . https://doi.org/10.1111/j.1574-6941.2011.01122.x

Pfeifer V, Nicholson GJ, Ries J, Recktenwald J, Schefer AB, Shawky RM, Schröder J, Wohlleben W, Pelzer S (2001) A polyketide synthase in glycopeptide biosynthesis. J Biol Chem 276:38370–38377 . https://doi.org/10.1074/jbc.M106580200

PHILLIPS M, LEON P, BORONAT A, RODRIGUEZCONCEPCION M (2008) The plastidial MEP pathway: unified nomenclature and resources. Trends Plant Sci 13:619–623 . https://doi.org/10.1016/j.tplants.2008.09.003

Pimentel-Elardo SM, Grozdanov L, Proksch S, Hentschel U (2012) Diversity of nonribosomal peptide synthetase genes in the microbial metagenomes of marine sponges. Mar Drugs 10:1192–1202 . https://doi.org/10.3390/md10061192

Rastogi N, Abaul J, Goh KS, Devallois A, PhilogÃ¨ne E, Bourgeois P (1998) Antimycobacterial activity of chemically defined natural substances from the Caribbean flora in Guadeloupe. FEMS Immunol Med Microbiol 20:267–273 . https://doi.org/10.1111/j.1574-695X.1998.tb01136.x

Rawlings BJ (2001) Type I polyketide biosynthesis in bacteria (Part A — erythromycin biosynthesis) (1994 to 2000). Nat Prod Rep 18:190–227 . https://doi.org/10.1039/b009329g

Rawlings BJ, Rawlings BJ, Rawlings BJ (1999) Biosynthesis of polyketides (other than actinomycete macrolides). Nat Prod Rep 16:425–484 . https://doi.org/10.1039/a900566h

Reddy BVB, Kallifidas D, Kim JH, Charlop-Powers Z, Feng Z, Brady SF (2012) Natural product biosynthetic gene diversity in geographically distinct soil microbiomes. Appl Environ Microbiol 78:3744–3752 . https://doi.org/10.1128/AEM.00102-12

Reddy BVB, Milshteyn A, Charlop-Powers Z, Brady SF (2014) eSNaPD: A Versatile, Web-Based Bioinformatics Platform for Surveying and Mining Natural Product Biosynthetic

Diversity from Metagenomes. Chem Biol 21:1023–1033 . https://doi.org/10.1016/j.chembiol.2014.06.007

Rix U, Fischer C, Remsing LL, Rohr J (2002) Modification of post-PKS tailoring steps through combinatorial biosynthesis. Nat Prod Rep 19:542–580 . https://doi.org/10.1039/b103920m

Roesch LFW, Fulthorpe RR, Riva A, Casella G, Hadwin AKM, Kent AD, Daroub SH, Camargo FAO, Farmerie WG, Triplett EW (2007) Pyrosequencing enumerates and contrasts soil microbial diversity. ISME J 1:283–290 . https://doi.org/10.1038/ismej.2007.53

Rondon MR, August PR, Bettermann AD, Brady SF, Grossman TH, Liles MR, Loiacono KA, Lynch BA, MacNeil IA, Minor C, Tiong CL, Gilman M, Osburne MS, Clardy J, Handelsman J, Goodman RM (2000) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. Appl Environ Microbiol 66:2541–2547 . https://doi.org/10.1128/AEM.66.6.2541-2547.2000

Russell AH, Truman AW (2020) Genome mining strategies for ribosomally synthesised and post-translationally modified peptides. Comput Struct Biotechnol J 18:1838–1851 . https://doi.org/10.1016/j.csbj.2020.06.032

Rutledge PJ, Challis GL (2015) Discovery of microbial natural products by activation of silent biosynthetic gene clusters. Nat Rev Microbiol 13:509–523 . https://doi.org/10.1038/nrmicro3496

Santana-Pereira ALR, Sandoval-Powers M, Monsma S, Zhou J, Santos SR, Mead DA, Liles MR (2020) Discovery of novel biosynthetic gene cluster diversity from a soil metagenomic library. Front Microbiol 11: . https://doi.org/10.3389/fmicb.2020.585398

Sar A, Pal S, Dam B (2018) Isolation of high molecular weight and humic acid-free metagenomic DNA from lignocellulose-rich samples compatible for direct fosmid cloning. Appl Microbiol Biotechnol 102:6207–6219 . https://doi.org/10.1007/s00253-018-9102-6

Schloss PD, Handelsman J (2003) Biotechnological prospects from metagenomics. Curr Opin Biotechnol 14:303–310 . https://doi.org/10.1016/S0958-1669(03)00067-3

Schmidt EW, Nelson JT, Rasko DA, Sudek S, Eisen JA, Haygood MG, Ravel J (2005) Patellamide A and C biosynthesis by a microcin-like pathway in Prochloron didemni, the cyanobacterial symbiont of Lissoclinum patella. Proc Natl Acad Sci 102:7315–7320 . https://doi.org/10.1073/pnas.0501424102

Scholz R, Molohon KJ, Nachtigall J, Vater J, Markley AL, Süssmuth RD, Mitchell DA,

Borriss R (2011) Plantazolicin, a novel microcin B17/streptolysin S-like natural product from Bacillus amyloliquefaciens FZB42. J Bacteriol 193:215–224 . https://doi.org/10.1128/JB.00784-10

Schwarzer D, Marahiel MA (2001) Multimodular biocatalysts for natural product assembly. Naturwissenschaften 88:93–101 . https://doi.org/10.1007/s001140100211

Seshime Y, Juvvadi PR, Fujii I, Kitamoto K (2005) Discovery of a novel superfamily of type III polyketide synthases in Aspergillus oryzae. Biochem Biophys Res Commun 331:253–260 . https://doi.org/10.1016/j.bbrc.2005.03.160

Seto B (2012) Rapamycin and mTOR: a serendipitous discovery and implications for breast cancer. Clin Transl Med 1: . https://doi.org/10.1186/2001-1326-1-29

Shen B (2003) Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. Curr Opin Chem Biol 7:285–295 . https://doi.org/10.1016/S1367-5931(03)00020-6

Shen B (2000) Biosynthesis of aromatic polyketides. pp 1–51

Skinnider MA, Dejong CA, Rees PN, Johnston CW, Li H, Webster ALH, Wyatt MA, Magarvey NA (2015) Genomes to natural products PRediction Informatics for Secondary Metabolomes (PRISM). Nucleic Acids Res gkv1012 . https://doi.org/10.1093/nar/gkv1012

Stachelhaus T, Hüser A, Marahiel MA (1996) Biochemical characterization of peptidyl carrier protein (PCP), the thiolation domain of multifunctional peptide synthetases. Chem Biol 3:913–921 . https://doi.org/10.1016/S1074-5521(96)90180-5

Stachelhaus T, Mootz HD, Bergendahl V, Marahiel MA (1998) Peptide bond formation in nonribosomal peptide biosynthesis. J Biol Chem 273:22773–22781 . https://doi.org/10.1074/jbc.273.35.22773

Staunton J, Weissman KJ (2001) Polyketide biosynthesis: a millennium review. Nat Prod Rep 18:380–416 . https://doi.org/10.1039/a909079g

Stevenson LJ, Bracegirdle J, Liu L, Sharrock A V., Ackerley DF, Keyzers RA, Owen JG (2021) Metathramycin, a new bioactive aureolic acid discovered by heterologous expression of a metagenome derived biosynthetic pathway. RSC Chem Biol 2:556–567 . https://doi.org/10.1039/D0CB00228C

Süssmuth RD, Mainz A (2017) Nonribosomal peptide synthesis-principles and prospects. Angew Chemie Int Ed 56:3770–3821 . https://doi.org/10.1002/anie.201609079

Tang X, Li J, Millán-Aguiñaga N, Zhang JJ, O'Neill EC, Ugalde JA, Jensen PR, Mantovani SM, Moore BS (2015) Identification of thiotetronic acid antibiotic biosynthetic pathways

by target-directed genome mining. ACS Chem Biol 10:2841–2849 . https://doi.org/10.1021/acschembio.5b00658

Verma SK, Singh H, Sharma PC (2017) An improved method suitable for isolation of high-quality metagenomic DNA from diverse soils. 3 Biotech 7:171 . https://doi.org/10.1007/s13205-017-0847-x

Wagner K-H, Elmadfa I (2003) Biological relevance of terpenoids. Ann Nutr Metab 47:95–106 . https://doi.org/10.1159/000070030

Waisvisz JM, van der Hoeven MG, van Peppen J, Zwennis WCM (1957) Bottromycin. I. A new sulfur-containing antibiotic. J Am Chem Soc 79:4520–4521 . https://doi.org/10.1021/ja01573a072

Walsh CT (2004) Polyketide and nonribosomal peptide antibiotics: Modularity and versatility. Science (80-) 303:1805–1810 . https://doi.org/10.1126/science.1094318

Walsh CT (2008) The chemical versatility of natural-product assembly lines. Acc Chem Res 41:4–10 . https://doi.org/10.1021/ar7000414

Walsh CT, Chen H, Keating TA, Hubbard BK, Losey HC, Luo L, Marshall CG, Miller DA, Patel HM (2001) Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines. Curr Opin Chem Biol 5:525–534 . https://doi.org/10.1016/S1367-5931(00)00235-0

Waschulin V, Borsetto C, James R, Newsham KK, Donadio S, Corre C, Wellington E (2021) Biosynthetic potential of uncultured Antarctic soil bacteria revealed through long-read metagenomic sequencing. ISME J. https://doi.org/10.1038/s41396-021-01052-3

Wexler M, Johnston AWB (2010) Wide host-range cloning for functional metagenomics. pp 77–96

Wnuk E, Waśko A, Walkiewicz A, Bartmiński P, Bejger R, Mielnik L, Bieganowski A (2020) The effects of humic substances on DNA isolation from soils. PeerJ 8:e9378 . https://doi.org/10.7717/peerj.9378

Wohlleben W, Mast Y, Stegmann E, Ziemert N (2016) Antibiotic drug discovery. Microb Biotechnol 9:541–548 . https://doi.org/10.1111/1751-7915.12388

Wu C, Shang Z, Lemetre C, Ternei MA, Brady SF (2019) Cadasides, calcium-dependent acidic lipopeptides from the soil metagenome that are active against multidrug-resistant bacteria. J Am Chem Soc 141:3910–3919 . https://doi.org/10.1021/jacs.8b12087

Yamanaka K, Reynolds KA, Kersten RD, Ryan KS, Gonzalez DJ, Nizet V, Dorrestein PC, Moore BS (2014) Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. Proc Natl Acad Sci 111:1957–1962 .

https://doi.org/10.1073/pnas.1319584111

Zang E, Brandes S, Tovar M, Martin K, Mech F, Horbert P, Henkel T, Figge MT, Roth M (2013) Real-time image processing for label-free enrichment of Actinobacteria cultivated in picolitre droplets. Lab Chip 13:3707 . https://doi.org/10.1039/c3lc50572c

Zhang JJ, Yamanaka K, Tang X, Moore BS (2019) Direct cloning and heterologous expression of natural product biosynthetic gene clusters by transformation-associated recombination. pp 87–110

Zhang L, Hashimoto T, Qin B, Hashimoto J, Kozone I, Kawahara T, Okada M, Awakawa T, Ito T, Asakawa Y, Ueki M, Takahashi S, Osada H, Wakimoto T, Ikeda H, Shin-ya K, Abe I (2017) Characterization of giant modular PKSs provides insight into genetic mechanism for structural diversification of aminopolyol polyketides. Angew Chemie Int Ed 56:1740–1745 . https://doi.org/10.1002/anie.201611371

Zhang Z, Schwartz S, Wagner L, Miller W (2000) A greedy algorithm for aligning DNA sequences. J Comput Biol 7:203–214 . https://doi.org/10.1089/10665270050081478

Zhu S, Fage CD, Hegemann JD, Mielcarek A, Yan D, Linne U, Marahiel MA (2016a) The B1 protein guides the biosynthesis of a lasso peptide. Sci Rep 6:35604 . https://doi.org/10.1038/srep35604

Zhu S, Hegemann JD, Fage CD, Zimmermann M, Xie X, Linne U, Marahiel MA (2016b) Insights into the unique phosphorylation of the lasso peptide paeninodin. J Biol Chem 291:13662–13678 . https://doi.org/10.1074/jbc.M116.722108

Ziemert N, Alanjary M, Weber T (2016) The evolution of genome mining in microbes – a review. Nat Prod Rep 33:988–1005 . https://doi.org/10.1039/C6NP00025H

Ziemert N, Podell S, Penn K, Badger JH, Allen E, Jensen PR (2012) The Natural Product Domain Seeker NaPDoS: A Phylogeny Based Bioinformatic Tool to Classify Secondary Metabolite Gene Diversity. PLoS One 7:e34064 . https://doi.org/10.1371/journal.pone.0034064

Zimmermann M, Hegemann JD, Xie X, Marahiel MA (2014) Characterization of caulonodin lasso peptides revealed unprecedented N-terminal residues and a precursor motif essential for peptide maturation. Chem Sci 5:4032–4043 . https://doi.org/10.1039/C4SC01428F

Zong C, Cheung-Lee WL, Elashal HE, Raj M, Link AJ (2018) Albusnodin: an acetylated lasso peptide from Streptomyces albus. Chem Commun 54:1339–1342 . https://doi.org/10.1039/C7CC08620B