Stephanie Nadine Hollauf, BSc

Analysis of putative nonribosomal peptide synthetases from the moss metagenome

MASTERARBEIT

zur Erlangung des akademischen Grades

Master of Science

Masterstudium Molekulare Mikrobiologie

eingereicht an der

Technischen Universität Graz

Betreuerin

Univ.-Prof. Dipl.-Biol. Dr.rer.nat. Gabriele Berg

Institut für Umweltbiotechnologie

Graz, Oktober 2017

EIDESSTATTLICHE ERKLÄRUNG

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Das in TUGRAZonline hochgeladene Textdokument ist mit der vorliegenden Masterarbeit identisch.

Datum

Unterschrift

Acknowledgements

First, I would like to thank Univ.-Prof. Dipl.-Biol. Dr.rer.nat. Gabriele Berg for giving me the opportunity to perform my master's thesis at the Institute of Environmental Biotechnology of the University of Technology Graz. It was a great pleasure working in this novel field of research and I am sincerely thankful for her support during my time at institute.

I would also like to express my heartfelt gratitude to M.Sc. B.E. Dr.rer.nat Christina Andrea Müller for her excellent supervision, her guidance and encouraging words when sometimes things got frustrating. Without her untiring commitment, the thesis presented here would not have been possible.

Special thanks also to Ing. Angelika Schaefer, for performing and evaluating the HPLC analyses, which were more than a little tricky. Likewise, I want to thank Ing. Barbara Fetz for all her help in the laboratory. The institute would not be the same without her.

Further I want to thank my colleagues at the institute for ungrudgingly taking me in and making me part of this "research family". Especially I would like to thank Tanja Nottendorfer and Isabella Wrolli for all their help with experiments and M.Sc. B.Sc. Melanie Obermeier for the great scientific and private talks we had. It was a great pleasure working with them.

At this point I would also like to express my heartfelt thanks to my friends. Especially Mag. Pharm. Teresa Pirker, B.Sc. Anna Pietschnig and M.Sc. Christina Laireiter were always there to motivate me and to remember me not to lose sight of the desired outcome. I am so thankful to have them.

Last but not least, I am deeply grateful for the support and motivation of my parents Marianne and Gerhard Hollauf. Without them I would not be where I am now.

Abstract

Occurrence of new antibiotic resistances in bacteria, high cancer rates and man-made pollution of the environment – those are only some of the challenges mankind is currently facing. Consequently, the need for novel antibiotics, chemotherapeutics, agrochemicals or biocontrol agents is ever growing. The exploration of biodiversity for the development of new biological products, which is called bioprospecting, has a successful history. In this regard, bioprospecting of plant microbiomes opens up possibilities for the discovery of new biologicals and enzymes, since plant microbiomes are nearly untapped sources for novel microbial functions. *Sphagnum* mosses are known to produce a wide variety of bioactive secondary metabolites and harbour a highly diverse microbial community. Since many prominent classes of bioactive compounds, like antibiotics, anti-cancer agents, antifungals or siderophores, are synthesized by nonribosomal peptide synthetases (NRPSs), the diversity of NRPS-assigned sequences in a *Sphagnum* moss metagenomic fosmid library was explored using a PCR based approach. Thirteen novel NRPS-assigned sequences with amino acid sequence identities up to 91% to annotated but not yet characterized synthetases were identified.

The aim of this study was to perform a comprehensive analysis of those novel NRPSs. Besides activity-based screenings for antimicrobials and siderophores, the NRPS-harboring fosmid clones were analyzed for novel chemical entities with thin layer chromatography (TLC) and high-performance liquid chromatography (HPLC). Furthermore, one NRPS clone was chosen for whole-insert sequencing and sequence analysis with the goal to retrieve the NRPS-like gene for subsequent heterologous expression and characterization. The insert sequence of this clone was also analyzed for other putative open reading frames (ORFs) and gene clusters.

The activity screenings led to the identification of a putative antimicrobial-producing clone harboring a NRPS system that might be responsible for the synthesis of a lipopeptide-type antibiotic. Moreover, sequence analysis of the selected clone led to the identification of 84 putative open reading frames coding for various enzymes with potential for industrial applications. The NRPS-like gene retrieved from this sequence has strong similarities with homopoly(amino acid) synthetizing NRPSs. Homopoly(amino acids) have enormous potential for various industrial applications, for example as food preservative or as biocontrol agents. Possible substrates for this unusual NRPS were postulated based on sequence information and results from the performed HPLC analysis. Further studies need to be performed to characterize this synthetase and its putative homopolymer. This synthetase would be the third characterized NRPS known to produce homopoly(amino acids) and would also be the first characterized homopoly(amino acids) synthetizing NRPS from a metagenome.

The findings in this study validate that bioprospecting plant-associated microbiomes for novel natural compounds and enzymes has a promising future.

Kurzzusammenfassung

Ein erhöhtes Vorkommen von Antibiotikaresistenzen und tumorösen Erkrankungen sowie die von uns Menschen verursachte Verschmutzung der Umwelt sind nur einige der Herausforderungen die sich die Menschheit gegenübersieht. Als Konsequenz steigt auch der Bedarf an neuen Antibiotika, Chemotherapeutika, Chemikalien für die Landwirtschaft oder Produkten für die Biokontrolle. Bioprospektion, die Erforschung der Biodiversität für die Entwicklung und Kommerzialisierung neuer Produkte, hat eine erfolgreiche Geschichte. In diesem Zusammenhang hat auch die Bioprospektion von Pflanzenmikrobiomen großes Potential für die Entdeckung neuer bioaktiver Stoffe und Enzyme, vor allem da Pflanzenmikrobiome hinsichtlich mikrobieller Funktionen nahezu unerforscht sind. Moose der Gattung *Sphagnum* produzieren eine weite Palette an bioaktiven sekundären Metaboliten und beherbergen ein überaus diverses Mikrobiom. Viele interessante bioaktive Stoffe, wie Antibiotika oder Siderophore, werden von nichtribosomalen Peptidsynthetasen (NRPSs) synthetisiert. Daher wurde die Diversität von NRPSs Sequenzen in einer *Sphagnum* Fosmid-Metagenombibliothek mittels PCR erforscht. Dreizehn neue NRPSs Sequenzen mit einer maximalen Identität von 91% zu annotierten, aber nicht charakterisierten Synthetasen konnten so identifiziert werden.

Das Ziel der hier durchgeführten Studie war es, diese neuen Moos-Metagenom NRPSs umfassend zu analysieren. Neben Aktivitätsscreenings, wo nach antimikrobiellen Stoffen sowie Siderophoren gesucht wurde, wurden die NRPS-Klone mittels Dünnschichtchromatographie (TLC) und Hochleistungsflüssigkeitschromatographie (HPLC) auf die Produktion neuer chemischer Verbindungen untersucht. Überdies wurde ein NRPS-Klon für die komplette Sequenzierung des Metagenom-DNA Inserts ausgewählt, mit dem übergeordneten Ziel das NRPS-Gen für eine spätere Expression und Charakterisierung zu gewinnen. Zusätzlich wurde das Insert auf das Vorhandensein möglicher andere Gene und Gen-Cluster analysiert.

Die aktivitätsbasierten Screenings führten hierbei zur Identifikation eines Klons mit antimikrobieller Aktivität, der möglicherweise ein NRPS-System für die Synthese eines Lipopeptid-Antibiotikums besitzt. Die durchgeführte Sequenzanalyse des ausgewählten Klons führte zur Identifikation von 84 offenen Leserahmen, die für diverse Enzyme mit industriellem Potenzial kodieren. Durch die Sequenzanalyse konnte ein NRPS-Gen gewonnen werden, das starke Ähnlichkeiten mit Homopolyaminosäuren-synthetisierenden NRPS aufweist. Homopolyaminosäuren haben enormes industrielles Potenzial, zum Beispiel als Konservierungsmittel. Mögliche Substrate für diese ungewöhnliche Synthetase wurden anhand der Sequenzinformation und der durchgeführten HPLC-Analyse bestimmt. Weitere Studien zur Charakterisierung dieser Synthetase und ihres möglichen Homopolyaminosäure-Produkts sollten durchgeführt werden. Würde sich die Annahme, dass es sich bei dieser Synthetase um eine Homopolyaminosäuren-Synthetase handelt, bestätigen, wäre dies erst die dritte charakterisierte Synthetase dieser Art.

Ebenso wäre sie die erste charakterisierte Homopolyaminosäuren-Synthetase aus einem Metagenom. Die Ergebnisse dieser Studie bestätigen, dass die Bioprospektion von Pflanzenmikrobiomen großes Potenzial für die Entdeckung neuer bioaktiver Stoffe und Enzyme besitzt.

Table of Contents

1. INTRODUCTION 1

1.1 A need for new natural products 1

1.2 The plant microbiome – A great source for novel metabolites 1

1.3 Metagenomics as a tool for bioprospecting 2

1.4 Nonribosomal peptide synthetases: natural-product assemblers 4

1.4.1 How nonribosomal peptides are synthesized 4

1.4.2 Antibiotics, siderophores & co: examples of NRPS products 6

1.4.3 Nonribosomal peptide assembly lines can be engineered 9

1.5 Connecting the plant microbiome, metagenomics & NRPSs Current state of research 11

2. AIM OF THE THESIS 13

3. MATERIAL AND METHODS 14

3.1 Bacterial strains, fungal strains and vector systems 14

3.1.1 *Escherichia coli* strains and vectors 14

3.1.2 NRPS-containing fosmid clones (EPI300 pCC2FOS_NRPS) 14

3.1.3 Bacterial and fungal strains for antimicrobial activity assays 15

3.2 Vector DNA isolation 16

3.3 Restriction analysis of vectors 16

3.4 Agarose gel electrophoresis 17

3.5 Preparation of electrocompetent *E. coli* cells 17

3.6 Siderophore production 17

3.6.1 Media and solutions 17

3.6.2 Generation of the strain *E. coli* K-12 *trfA* A250V *ΔentF* harboring pCC2FOS_NRPS 18

3.6.2.1 Verification that *ΔentF* lacks siderophore production 18

3.6.2.2 Polymerase chain reaction to control the *trfA* gene on pRS48 18

3.6.2.3 Site-directed mutagenesis of the *trfA* gene on pRS48 19

3.6.2.4 Transformation of pRS48 *trfA* A250V into pir-116 19

3.6.2.5 Transformation of pRS48 *trfA* A250V into *E. coli ΔentF* 20

3.6.2.6 Verification of the transposon integration into *ΔentF* 20

3.6.2.7 Transformation of pCC2FOS_NRPS into *E. coli* K 12 *trfA* A250V *ΔentF* 21

3.6.2.8 Enhancing the copy number of pCC2FOS_NRPS in *trfA* A250V *ΔentF* 21

3.6.3 Growth of EPI300 pCC2FOS_empty and *trfA* A250V *ΔentF* pCC2FOS_empty in MM9 22

3.6.4 Siderophore production in liquid culture 22

3.6.5 Siderophore production on CAS agar plates 23

3.7 Antimicrobial activity assays 23

3.7.1 Preparation of the test samples 23

3.7.2 Well-diffusion assay using soft agar 24

3.7.2 Agar disc-diffusion and agar well-diffusion assay 24

3.8 Thin-Layer Chromatography 26

3.9 High Performance Liquid Chromatography 26

3.9.1 HPLC coupled with UV-detection 26
3.9.2 HPLC coupled with mass spectrometry (MS) 27
3.10 Sequence analysis of fosmid clone 6-H4 27
3.10.1 Shotgun-sequencing, primer walking and sequence analysis 27
3.10.2 Generation of pET28a(+) 6-H4_Synth expression vector and transformation 28
3.10.3 Expression of 6-H4_Synth in BAP1 29
3.10.4 SDS-PAGE analysis 30

4. RESULTS 31
4.1 Siderophore production of the NRPS fosmid clones 31
4.1.1 Successful generation of *E. coli* K-12 *trfA* A250V *ΔentF* pCC2FOS_NRPS 31
4.1.2 Copy number enhancement in *trfA* A250V *ΔentF* pCC2FOS is not detectable 33
4.1.3 Reduced growth of EPI300 and *ΔentF* when grown in iron-deficient medium 34
4.1.4 Investigation of fosmid clone siderophore production using liquid CAS assay 34
4.1.5 Siderophore production of the NRPS fosmid clones on CAS agar plates 36
4.2 Analyzing the NRPS clones for production of antimicrobials 37
4.2.1 Clone 2-F4 exhibited antimicrobial activity in one round of testing 37
4.2.2 The 2-F4 synthetase shows similarity to antimicrobial synthesizing NRPSs 38
4.3 TLC and HPLC analysis for further product detection 40
4.3.1 Clone 6-H4 showed differences on TLC plates and unique peaks in HPLC/UV 40
4.3.2 Preparative TLC and HPLC/MS revealed unique products of clone 6-H4 43
4.4 Sequence analysis of fosmid clone 6-H4 48
4.4.1 Clone 6-H4 harbors 84 putative open reading frames coding for various enzymes 48
4.4.2. The 6-H4 NRPS (ORF 27) has an unusual structure and is part of a gene cluster 52
4.4.3. Heterologous expression of the 6-H4 synthetase led to insoluble protein 53
4.4.4. The 6-H4 synthetase is a membrane-related protein 56
4.4.5. The 6-H4 synthetase has similarities with homopoly(amino acid) synthetases 57

5. DISCUSSION 62
5.1 Analysis of siderophore production 62
5.2 The antimicrobial activity of clone 2-F4 65
5.3 TLC and HPLC analysis for metagenomic library screening 66
5.4 Sequence analysis of the NRPS-containing fosmid clone 6-H4 68

6. CONCLUSION 74

REFERENCES 75

LIST OF TABLES 83

LIST OF FIGURES 84

LIST OF ABBREVIATIONS 85

ATTACHMENT 87

1. Introduction

1.1 A need for new natural products

A higher occurrence of new antibiotic resistances in bacteria, high cancer rates and man-made pollution of land and water are only some of the challenges mankind is currently facing (Woolhouse & Farrar 2014; Torre et al. 2015; Sharma 2012). Consequently, there is an ever growing need for new antibiotics, anti-cancer drugs, agrochemicals, food additives and biocontrol agents, that are highly effective, but possess low toxicity and only have a minor environmental impact (Mateo et al. 2001). The mining of biodiversity for new secondary metabolites and their commercialization – which is called bioprospecting – has a successful history, since up to 50% of all approved drugs currently on the market are natural products or derivatives thereof (Strobel & Daisy 2003; Newman & Cragg 2016). In this regard, bioprospecting of plant microbiomes opens up new possibilities for the discovery of new natural products.

1.2 The plant microbiome – A great source for novel metabolites

Microorganisms represent the richest source of biologically active compounds that find applications in various industrial fields (Gunatilaka 2006). Unlike other known producers of small molecule natural products, microbes occupy all animate and inanimate niches on earth: deserts, alpine, arctic and Antarctic areas, the ocean and the subsurface, even hot springs and caves (Whitman et al. 1998; Barton & Jurado 2007). Correspondingly, plants harbor specific microbial communities as well and are nowadays considered as metaorganisms that establish a tight symbiotic relationship with their microbiome. Plant-associated microbes fulfill essential functions for plant growth and health, as they can improve nutrient acquisition, stimulate germination, fend off plant pathogens, enhance stress resistance and influence the general plant fitness (Berg et al. 2014, Berg et al. 2016).

It is estimated that there are roughly 300,000 species of higher plants (McChesney et al. 2007) harboring different specific niches for microbes. However, while the rhizosphere has been well-investigated (Philippot et al. 2013), the phyllosphere and the endosphere are nearly untapped sources for novel microbial functions and natural products (Vorholt 2012; Strobel & Daisy 2003). Each plant itself is colonized by more than a thousand microbial species, but researchers who want to investigate those microbes for their potential to produce biotechnologically interesting metabolites face a huge problem: many of those microorganisms found in the environment are not cultivable under laboratory conditions (Müller et al. 2016).

1.3 Metagenomics as a tool for bioprospecting

The enormous diversity of uncultured microorganisms observed in the environment (about 99% of all microbes) provides a rich repository for novel natural compounds (Nováková & Farkašovský 2013). To get access to this enormous treasure chest of industrially interesting secondary metabolites, a cultivation-independent approach based on the construction of complex DNA libraries in a heterologous host was developed, so-called metagenomics (Handelsman et al. 1998). Bypassing the need for isolation and cultivation, the culture-independent analysis of the collective genome of an assemblage of microorganisms from a particular habitat has proven to be a powerful tool, not only for exploring and comparing microbial ecology (Xu 2006) but also for identifying novel enzymes and bioactive molecules (Rondon et al. 2000; Voget et al. 2003; Courtois et al. 2003; Iqbal et al. 2012).

Two different strategies are used to examine metagenomic libraries for secondary metabolites and biocatalysts: function-based and sequence-based screenings (Fig. 1).

With function-driven analysis, biomolecules are detected based on their activity using specifically designed screening assays and are further characterized by molecular-biological and biochemical methods (Sleator et al. 2008). Although this strategy allows quick identification of new natural products and enzymes, it is dependent on the heterologous expression of genes in the foreign host. Absence of building blocks, incompatibility with transcription, different codon usage or the requirement for post-translational modifications are a few examples of expression-related limitations – problems that could be partially overcome by using different host expression systems (Culligan et al. 2014; Uchiyama & Miyazaki 2009). In contrast, sequence-based approaches employ generally shot gun sequencing in combination with PCR screening or colony blot hybridization to identify genes coding for target proteins (Gabor et al. 2007; Nováková & Farkašovský 2013). The DNA primers or probes used for those screenings are derived from conserved regions of already known genes. Here is also were the limitation of this approach lies: only variants of known protein families can be detected (Simon & Daniel 2011).

Due to the fact that from both screening approaches mainly the function-driven analysis of metagenomic libraries can reveal truly novel enzymes, it is not surprising that the majority of characterized enzymes derived from the metagenome originated from activity based screenings (Ferrer et al. 2009; Tuffin et al. 2009). But the frequency of identified compounds is actually quite low. For example, Yun et al. (2004) screened a soil-derived metagenomic library for amylases and only one clone from among approximately 30,000 recombinant clones showed the desired activity.

Other reasons for the low amount of newly discovered compounds are not only the aforementioned hurdles of heterologous expression in a foreign host, but also that prominent classes of natural products, like linear or cyclic polyketides and peptides, are often synthesized by huge polyketide synthase or nonribosomal peptide synthetase gene clusters, that frequently exceed the cloning capacity of commonly used fosmid vectors (Nováková & Farkašovský 2013).

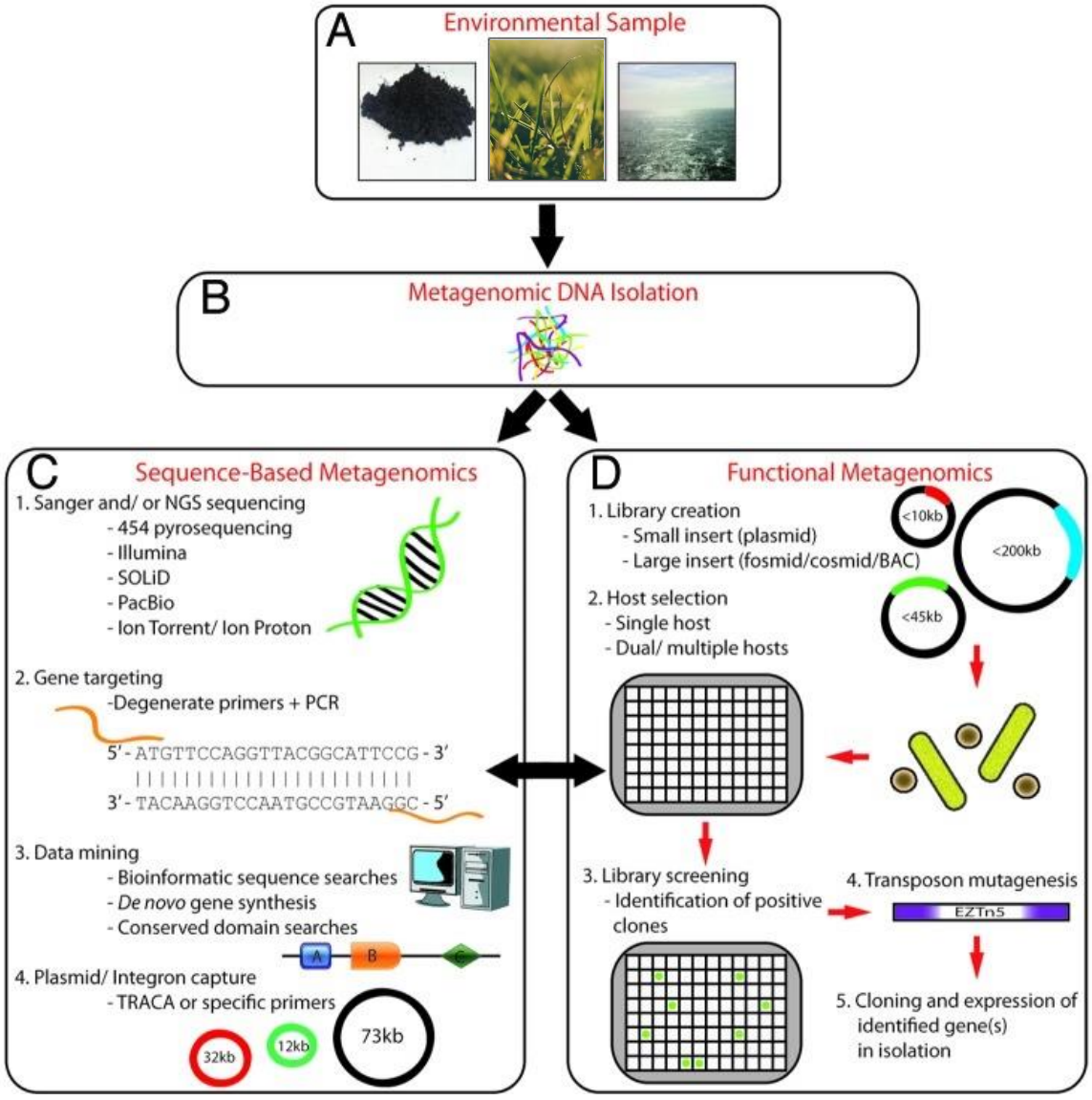


Figure 1: Overview of metagenomic library screening. *An environmental sample (e.g. soil, plants or ocean) is collected (A) and the total metagenomic DNA is isolated (B). The metagenomic DNA can either directly be subjected to sequence-based analysis (Sanger or next generation sequencing, targeting of genes using degenerate primers, in-silico data mining, complete gene synthesis, transposon-aided capture [TRACA] to find novel genes etc.) or a library is generated first and then screened for genes of interest (C). Functional metagenomics (D) involves the creation of library (small-insert or large-insert library) using a specific vector system (plasmid, fosmid, cosmid, bacterial artificial chromosome) and a host (or two/multiple hosts) suitable for the aims of the project. The metagenomic clones are then screened to identify a phenotype of interest and transposon mutagenesis can further be performed on positive clones to identify the gene or genes of interest. (Modified from Culligan et al. 2014)*

1.4 Nonribosomal peptide synthetases: natural-product assemblers

Nonribosomal peptides represent a remarkably diverse group of bioactive compounds, that exhibit complex chemical structures and have enormous biotechnological and pharmaceutical potential, as they include antibiotics, cytostatics, immunosuppressants, siderophores, toxins and pigments amongst others (Wang et al. 2014; Strieker et al. 2010; Amoutzias et al. 2016). The structural diversity of nonribosomal peptides can partly be attributed to the incorporation of not only proteinogenic amino acids, but also of many unusual, nonproteinogenic residues – so far more than 300 different precursors have been identified (Konz & Marahiel 1999). They are synthesized via a conserved thiotemplate mechanism by large multi-modular enzyme complexes termed nonribosomal peptide synthetases (NRPSs) (Walsh 2008).

As Wang et al. (2014) reported, NRPSs are widely distributed across all three domains of life. The majority of NRPS assigned gene clusters were found in bacterial species - especially in the phyla *Proteobacteria*, *Actinobacteria*, *Cyanobacteria* and *Firmicutes* - and fungi, but also animals, plants and archaea harbor biosynthetic pathways of NRPSs, although less frequently (Wang et al. 2014).

The initial research into NRPSs started in the early 1960s (Felnagle et al. 2008) and since then substantial insights into the mechanism on how synthetases assemble peptides have been gained. The following subchapters will give an overview of biosynthesis via NRPSs and will also provide examples of natural products assembled by those synthetases.

1.4.1 How nonribosomal peptides are synthesized

The structural framework of nonribosomal peptide synthetases is built by a various number of modules (Challis & Naismith 2004; Hur et al. 2012). NRPSs can be divided into type I NRPSs, were the modules are organized on a single polypeptide, and type II NRPSs, were they are in several interacting proteins, with the former seen mainly in many fungal and the latter in many bacterial species (Konz & Marahiel 1999; Hur et al. 2012). Furthermore, NRPSs can be classified into three categories, depending on how the biosynthesis of the nonribosomal peptide occurs. Type A NRPSs use the linear strategy where the number and the order of the modules in the synthetase matches the number and sequence of the amino acids in the peptide product. In Type B, domains or modules of the NRPS are used more than once in an iterative way, leading to a peptide consisting of repeated amino acid sequences. Nonlinear NRPSs – Type C – synthesize peptides with amino acid sequences that do not correlate with the arrangement of the modules on the synthetase (Mootz et al. 2002). Examples for all three possibilities of biosynthesis are shown in Figure 2.

The modules of NRPSs themselves consist of different catalytic domains responsible for the activation, incorporation and modification of a single building block into the natural product (Felnagle et al. 2008).

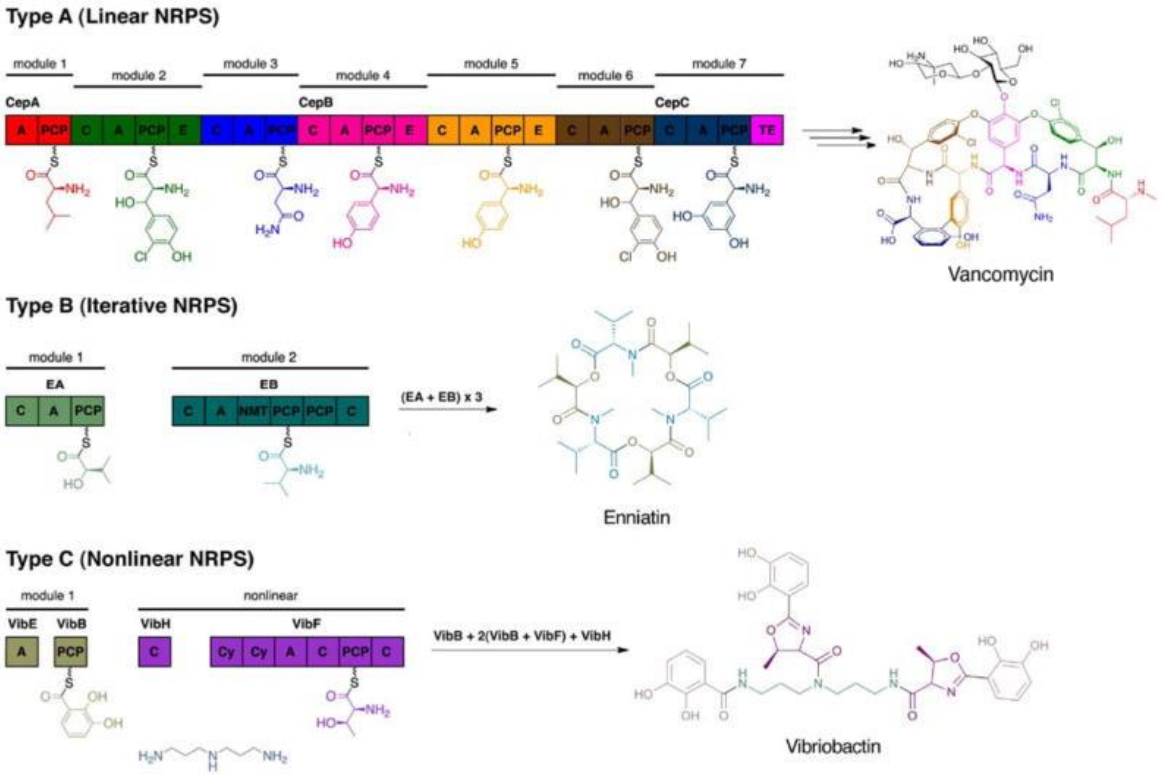


Figure 2: The three strategies of nonribosomal peptide synthesis. *The glycopeptide antibiotic vancomycin is assembled by a linear NRPS (Type A), while enniatins (cyclohexadepsipeptides) are synthesized via the iterative (Type B) and the siderophore vibriobactin via the nonlinear way (Type C). (Taken from Hur et al. 2012)*

Central domain of all NRPSs is the thiolation domain, also called peptidyl carrier protein (PCP). PCP acts as a scaffold, tethering the activated building blocks while they are modified and condensed by other domains of these modular biosynthetic enzymes (Lai et al. 2006). Together with the PCP domain, the adenylation domain (A) and the condensation domain (C) constitute the three core domains of a typical NRPS module (Hur et al. 2012). The A domain is responsible for substrate specificity and activation of the amino acid building block by formation of an amino acyl intermediate, a process that consumes ATP (Felnagle et al. 2008). This activated intermediate is then transferred and covalently bound to the thiol group of the 4'-phosphopantetheine arm that is attached to the PCP posttranslationally by a phosphopantetheinyl transferase (PPTase) (Lambalot et al. 1996; Lai et al. 2006). After undergoing all necessary modifications, the formation of a C-N bond with the amino acyl substrate tethered to the adjacent PCP is catalyzed by the C domain (Stachelhaus et al. 1998). The first (initiation) module of a linear NRPS usually lacks a C domain, due to the fact that there is no preceding substrate for peptide

bond formation (Felnagle et al. 2008). The last module, termed termination module, usually contains an additional thioesterase domain (TE) for the release of the assembled peptide from the final PCP (Hur et al. 2012). Upon release the linear peptide can undergo branching or cyclization (Kleinkauf & Von Döhren 1996; Konz & Marahiel 1999). Besides those core domains there are often additional tailoring domains, which are responsible for the modifications of the used building blocks, like epimerase (E), methyltransferase (MT), oxidation (Ox), and cyclization (Cy) domains (Hur et al. 2012). Also the released nonribosomal peptide can be further modified, for example by enzymes like glycosyltransferases or oxygenases – enhancing the structural diversity of those peptides even more (Hur et al. 2012; Konz & Marahiel 1999).

1.4.2 Antibiotics, siderophores & co: examples of NRPS products

Vancomycin, Enterobactin and Cyclosporin are extensively studied examples of nonribosomally synthesized peptides from microbial origin. These molecules and their synthesis pathways enable a good insight into the molecular diversity and genetic machinery of NRPSs.

Vancomycin, a glycopeptide type antibiotic, is produced by the actinomycete *Amycolatopsis orientalis* (Nagarajan 1991). Discovered in the 1950's (McCormick et al. 1954), this antibiotic is on the World Health Organization's List of Essential Medicines and is used as a last resort antibiotic in the treatment against multiple-drug resistant gram-positive bacterial infections, like methicillin-resistant *Staphylococcus aureus* (C. Liu et al. 2011; Williams & Bardsley 1999). However, resistance to vancomycin has emerged as well (Howden et al. 2010; Park et al. 2014). The mode of action of vancomycin is the blocking of cell wall biosynthesis by binding to the D-Ala-D-Ala terminus of the pentapeptides of the growing cell wall (Fig. 3), which likely prevents transpeptidation/transglycosylation, resulting in a weakened cell wall and bacterial lysis (Felnagle et al. 2008; Nagarajan 1991). Other antibiotics of this family are teicoplanin, chloroeremomycin, and balhimycin (Felnagle et al. 2008).

Synthesized by the Type I strategy of NRPS synthesis (Fig. 2), vancomycin, like all glycopeptide antibiotics, has a linear heptapeptide backbone and features oxidative cross-linking and glycosylation between the side chains of the peptide amino acids (Felnagle et al. 2008). It has to be noted that the NRPS enzymes are actually three in the number and are annotated as CepA, CepB, and CepC (Fig 2). Each of those enzymes is responsible for the incorporation of one to three amino acids. Besides leucine and asparagine (Nagarajan 1991), the backbone of vancomycin contains three nonproteinogenic amino acids that need to be assembled before the NRPS-catalyzed synthesis: β-hydroxytyrosine (β-HT; positions two and six), 4-hydroxyphenylglycine (HPG; positions four and five), and 3,5-dihydroxyphenylglycine (DPG; position seven) (Puk et al. 2004).

Oxidative cross-linking was thought to follow the heptapeptide synthesis, but it was later suggested that this process is an integral part of peptide biosynthesis (Zerbe et al. 2004). Biosynthesis is complete, when the oxidative cross-linked and chlorinated heptapeptides are glycosylated. Vancomycin biosynthesis represents an excellent example of nonribosomal peptide synthesis to generate structural diversity: incorporation of unique precursors, integrated modifying domains (epimerases), auxiliary enzymes, as well as downstream modifications after the peptide assembly (Felnagle et al. 2008).

Following the example of a nonribosomally synthesized antibiotic, another group of interesting natural products synthesized by NRPSs are small high-affinity iron chelating compounds, termed **siderophores**.

Iron (Fe) is an essential element for almost all life, playing not only a role in growth and development, but also enzymatic processes, oxygen metabolism, DNA and RNA synthesis and electron transfer (Kaplan & Ward 2013; Symeonidis & Marangos 2012). Besides those functions, Fe plays a crucial role in the pathogenicity of microorganisms (Skaar 2010). Because of the low bioavailability of Fe in the environment, microbes have developed specific strategies to access the insoluble form of this element (Fe^{3+}), for example by using the aforementioned siderophores. Depending on their functional group, siderophores are divided into three main families: catecholates, hydroxamates, and carboxylates (Ahmed & Holmström 2014).

Enterobactin (Fig. 3d), a siderophore secreted e.g. by *Escherichia coli* and *Salmonella typhimurium*, belongs to the chatecholate family (Raymond et al. 2003) and is synthesized by an iterative NRPS (Y. Liu et al. 2011). Three gene products are necessary for the initial stage of enterobactin synthesis, EntC (isochorismate synthetase), EntB (2,3-dihydro-2,3-dihydroxybenzoate synthetase) and EntA (2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase) (Crosa & Walsh 2002). Those synthetases are responsible for the conversion of chorismic acid to 2,3-dihydroxybenzoic acid (DHB) (Raymond et al. 2003). One molecule of enterobactin is then assembled from three molecules each of DHB and L-serine by the enzymes EntD, EntE, EntF, and the aryl carrier protein found at the C-terminus of EntB (ArCP). The stand-alone PCP-domain of EntB, as well as the PCP domain of the 142-kDA synthetase EntF (four domains: C-A-PCP-TE) are primed with phosphopantetheine by EntD. The building block DHB is activated by EntE and then transferred to the primed EntB (Fig. 3a). The A-domain in EntF activates and binds L-serine and the C-domain of this synthetase forms a peptide bond between the DHB intermediate bound to EntB and the L-serine intermediate bound to the PCP domain of EntF (Fig. 3b). Enterobactin is released after three cycles of loading and condensation by the TE-domain of EntF (Fig. 3c). The TE domain is also responsible for intermolecular cyclization of

the product (Ehmann et al. 2000). Other genes in the 22 kb enterobactin gene cluster are responsible for transport of ferric enterobactin and the intracellular release of iron from the siderophore (Crosa & Walsh 2002).

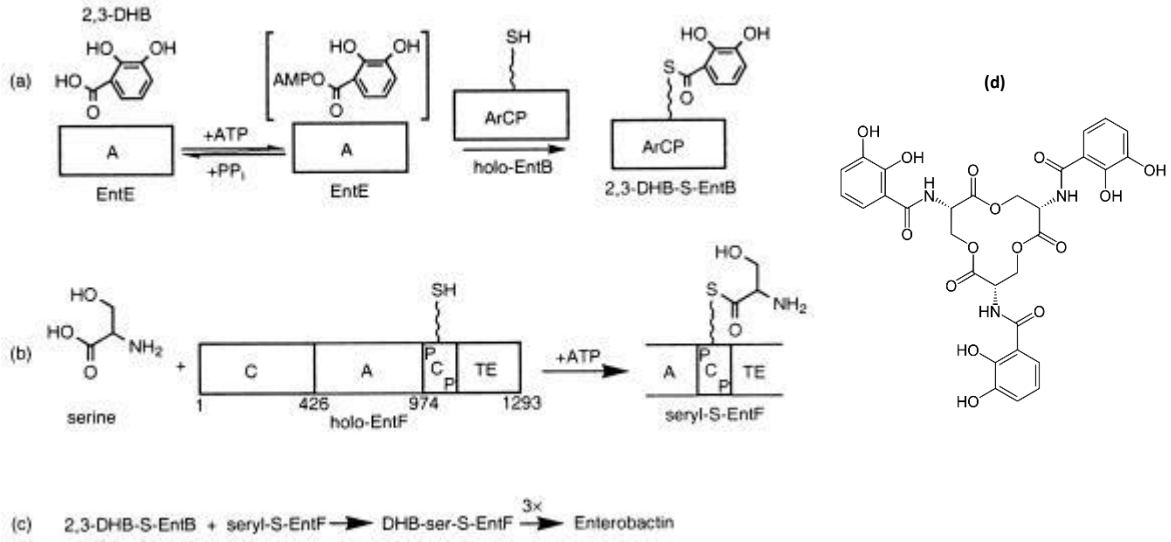


Figure 3: Scheme of enzymes involved in enterobactin synthesis. *(a) DHB is activated by EntE and loaded onto the primed ArCP domain of EntB. (b) L-serine is activated by the A-domain of EntF and then transferred onto the PCP domain of the same synthetase. (c) The C-domain of EntF condenses the serine intermediate with the DHB intermediate bound to EntB (Taken from Ehmann et al. 2000). After three cycles of loading and condensation, enterobactin (d) is released.*

The vibriobactin molecule shown in Figure 2 is also a nonribosomally synthesized catechol-type siderophore produced by the bacterium *Vibrio cholerae*. Unlike enterobactin, the vibriobactin biosynthetic and transport genes are located in two separate gene clusters (Crosa & Walsh 2002). Pyoverdines, fluorescent siderophores produced by *Pseudomonas spp.*, are mixed catecholate-hydroxamate siderophores synthesized by NRPS (Visca et al. 2007).

Moving away from iron-chelating compounds, the last more detailed explained nonribosomal peptide in this chapter will be **cyclosporin A** (Fig. 4). Cyclosporins are undecapeptides produced by the fungi *Tolypocladium inflatum* (Dreyfuss et al. 1976).

While cyclosporin A was initially investigated as a possible fungicide, its potent anti-inflammatory and immunosuppressant activity (due to suppression of T-cell activation) impelled its use in transplant surgeries to prevent graft-versus-host disease, as well as in the treatment of autoimmune diseases (Felnagle et al. 2008).

Cyclosporin A is assembled by a single Type A NRPS with eleven modules. Two points of divergence from the standard C-A-PCP architecture can be found in the immunosuppressants module architecture (Felnagle et al. 2008).

The modules two to five, as well as seven to eleven harbor and extra methyltransferase (M) domain, responsible for the N-methylation of six of the eight proteinogenic amino acids of cyclosporin A. Additional to the proteinogenic amino acids, cyclosporin A features three nonproteinogenic amino acids: D-Ala, (4*R*)-4-[(*E*)-2-butyl]-4-methyl-L-threonine, and L-2-aminobutyric acid. The second divergence is the occurrence of two additional C-domains, the first module initiates in a C-domain (instead of an A-domain) and the last module terminates in a C-domain (instead of a TE-domain). It is thought that one or both of these C-domains play a role in the cyclisation and the release of cyclosporin A from the synthetase (Felnagle et al. 2008).

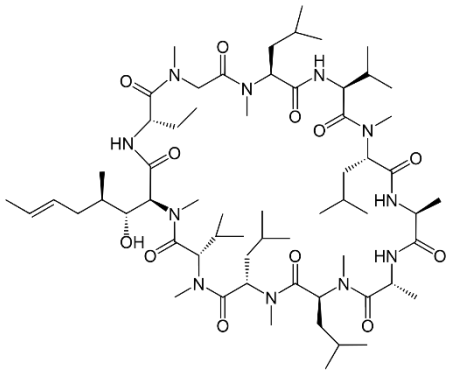


Figure 4: Structure of cyclosporin A.

Further examples of nonribosomally synthesized peptides with medicinal properties are listed by Felnagle et al. 2008. Also assembled by NRPSs are **enniatins** (Fig. 2) produced by *Fusarium* species, showing various biological activities, from ionophoric and antiparasitic effects to cytotoxic activity against cancer cells, which therefore may play a significant role in the future (Firáková et al. 2007).

The aforementioned examples give a short view of the vast molecular diversity of microbially synthesized nonribosomal peptides and illustrate their promising application as biotechnological and medicinal products.

1.4.3 Nonribosomal peptide assembly lines can be engineered

Due to the structural complexity of many nonribosomal peptides, major challenges associated with total- and semi-synthesis of those natural products exist. Therefore, there is great interest in the development of bioengineering approaches to increase product yields and generate novel modified peptides with improved bioactivity and physicochemical properties. Winn et al. (2016) reviewed the major advances in engineering nonribosomal peptide biosynthesis that occurred in the last decade, depicting different potential targets and starting points, like engineering of precursors and tailoring enzymes, NRPS module/domain exchange or the modification of adenylation domains.

One example is the introduction of halogen substituents into nonribosomal peptide assembly lines, which can significantly impact the bioactivity of compounds. When the *Streptomyces coeruleorubidus* peptide antibiotic pacidamycin was coexpressed with a halogenase originating from the pyrrolnitrin biosynthesis pathway, a novel halogenated (chlorinated) analogue was

generated. The halogenation of the tryptophan residue of pacidamycin occurred prior to the incorporation by the NRPS (Roy et al. 2010; Winn et al. 2016). But also exogenous tailoring enzymes from other pathways can be used to diversify the final peptide structure (Winn et al. 2016).

When the structure of the phenylalanine activating A-domain from GrsA – a NRPS involved in gramicidin S synthesis – was determined by Conti et al. in 1997, major insights into the substrate specificity of NRPSs were gained. The active site residues responsible for phenylalanine binding were identified after GrsA was complexed with AMP and L-phenylalanine. This led to deciphering of the NRPS specificity-conferring code (Stachelhaus et al. 1999). The so called "Stachelhaus code" is not only used for the prediction of substrate specificity of A-domains directly from the primary sequence, but also for the redesign of those domains (Kries & Hilvert 2011) by individual or combined point mutations. Changing individual amino acids within the peptides is thought to be less likely to disrupt the synthetase domain architecture and to cause major permutations in the product structure. Adenylation domains can also be promiscuous, meaning they can activate more than one substrate (Winn et al. 2016). For example, the A-domain of the third module of the NRPS FusA (synthesis of fusaricidins) is able to incorporate five different amino acids. Mutations of residues in the FusA adenylation domain binding pocket were performed to shift the production to the antibiotic and antifungal phenylalanine-containing variant – with success (Han et al. 2012).

The trickier way to generate novel synthetic peptides is to change whole domains or modules of NRPSs. The major problem is that even minor changes to the modular structure can effect protein folding and activity (Winn et al. 2016). Daptomycin, a nonribosomally synthesized lipopeptide antibiotic, was the first natural product antibiotic approved for clinical use in over thirty years (Lewis 2013). By exchanging dptD, the third NRPS subunit, with the terminal subunits of the other lipopeptide antibiotics A54145 and CDA (calcium-dependent antibiotic), new derivatives of daptomycin were generated and examined for their antimicrobial activity. Structural similarities of the subunits of those other lipopeptide antibiotics with dptD contributed to this success (Miao et al. 2006).

The engineering of pacidamycin, FusA and the daptomycin assembly line are only three examples of how NRPS can be modified to synthesize customized peptides. Furthermore, advances in synthetic biology could lead to the generation of a far greater number of new NRPS constructs than possible conventional genetic techniques allow (Winn et al. 2016).

Having several engineering tools at hand, the discovery and prospecting of new NRPS sequences and subunits is very promising in order to extend the available repertoire of activities for production of novel synthetic peptides.

1.5 Connecting the plant microbiome, metagenomics & NRPSs Current state of research

To date, metagenomic analyses of a variety of environments has been performed. Not only common habitats, like the soil (Daniel 2005) or water (Gomez-Alvarez et al. 2012), were analyzed, but also glaciers (Simon et al. 2009), hot springs (Kanokratana et al. 2004), marine sponges (Pimentel-Elardo et al. 2012), rumens (Hess et al. 2011) and the human body (Turnbaugh et al. 2007). In 2008, Wang et al. created the first metagenomic fosmid library of endophytes, paving the way for the recovery, identification and characterization of the functional gene repertoire of endophytic microorganisms.

But as already mentioned in chapter 1.2, the plant endosphere and phyllosphere are still a nearly untapped source for novel metabolites and functional genes. While soil (Hamedi et al. 2015), ocean (Hodges et al. 2012) or marine sponges (Pimentel-Elardo et al. 2012) have been investigated for NRPS biosynthetic gene clusters, plant-associated microorganisms rarely were. Using degenerate primers, Zhao et al. (2011) and Janso & Carter (2010) searched for NRPS and polyketide synthase (PKS) genes in endophytic actinomycetes isolated from medical plants in China and tropical plants in Papua New Guinea, on Mborokua Island and Solomon Islands, respectively. Mukherjee et al. (2012) identified a PKS/NRPS hybrid enzyme of *Trichoderma virens* involved in the induced systemic resistance response in maize. *Trichoderma* spp., filamentous fungi, not only colonize the rhizosphere but also live inside the roots as opportunistic symbionts. They are the most widely used biofungicides and produce a wide variety of secondary metabolites (Mukherjee et al. 2012). *Trichoderma* spp. are also known to produce siderophores. Lehner et al. (2013) performed a genome-wide screening for NRPS genes that might be involved in *Trichoderma* siderophore biosynthesis and Rosconi et al. (2013) identified and characterized serobactins, nonribosomally synthesized siderophores from the grass endophyte *Herbaspirillum seropedicae*. Serobactins are the first structurally described siderophores produced by endophytic bacteria. *H. seropedicae* Z67 is able to colonize the inner tissue of crops, such as corn, wheat, sorghum and rice and is of biotechnological interest because of its nitrogen-fixing capabilities, a process highly depending on iron (Rosconi et al. 2013).

The *Sphagnum* bog ecosystem belongs to the oldest vegetation forms and *Sphagnum*-associated microbes produce an enormous variety of bioactive substances. Functional analysis of the *Sphagnum magellanicum* metagenome revealed a high availability of systems responsible for

bioactive compound synthesis, for example quorum-sensing molecules, toxins and, especially, siderophores (Bragina et al. 2014). Besides siderophore production, antifungal and antibiotic activity has been detected in many bacteria isolated from *Sphagnum* species (Opelt et al. 2007; Opelt & Berg 2004). In 2015 Müller et al. mined the *Sphagnum* bog metagenome for NRPS and PKS genes. *In silico* data mining of the *Sphagnum* metagenome revealed 279 NRPS, 346 PKS and 40 NRPS/PKS hybrid genes. Additionally, the PCR-based screening of a *Sphagnum magellanicum* fosmid library led to the identification of thirteen novel NRPS-related sequences with amino acid sequence identities up to 91% to annotated but not yet characterized synthetases. Those identified novel sequences showed, for example, similarities to biosynthetic pathway genes for the toxin Syringomycin, the siderophore Pyoverdin or the well-studied Gramicidin synthetase, suggesting gene clusters for the production of novel siderophores, toxins or antibiotics (Müller et al. 2015).

Despite the phylogenetic study performed so far by Müller et al. (2015), further characterization of the NRPS-clones is necessary. Until now only partial sequences of the identified NRPS-related genes from the *S. magellanicum* metagenome were identified and aligned to available databases. In addition, no expression of the identified genes was tested – it is unknown if the NRPS gene clusters are complete and the NRPS-clones are able to synthesize functional peptides.

2. Aim of the thesis

The overall aim of this study was to perform a comprehensive analysis of the putative nonribosomal peptide synthetases from the *Sphagnum* moss metagenome found by Müller et al. (2015). Activity-based screenings were applied to examine the NRPS-containing fosmid clones regarding their role in the production of antibiotics, antifungals and siderophores. For further product detection and characterization, thin layer chromatography (TLC) and high performance liquid chromatography (HPLC) were applied. Furthermore, one NRPS-clone was chosen for whole-insert sequencing and sequence analysis, searching for putative open reading frames (ORFs) and gene clusters. Heterologous expression of the NRPS-assigned sequence of this clone using a suitable vector system in *E. coli* was evaluated to further characterize the synthetase gene. Figure 5 provides an overview of the different methods employed to examine and characterize the NRPS-containing fosmid clones.

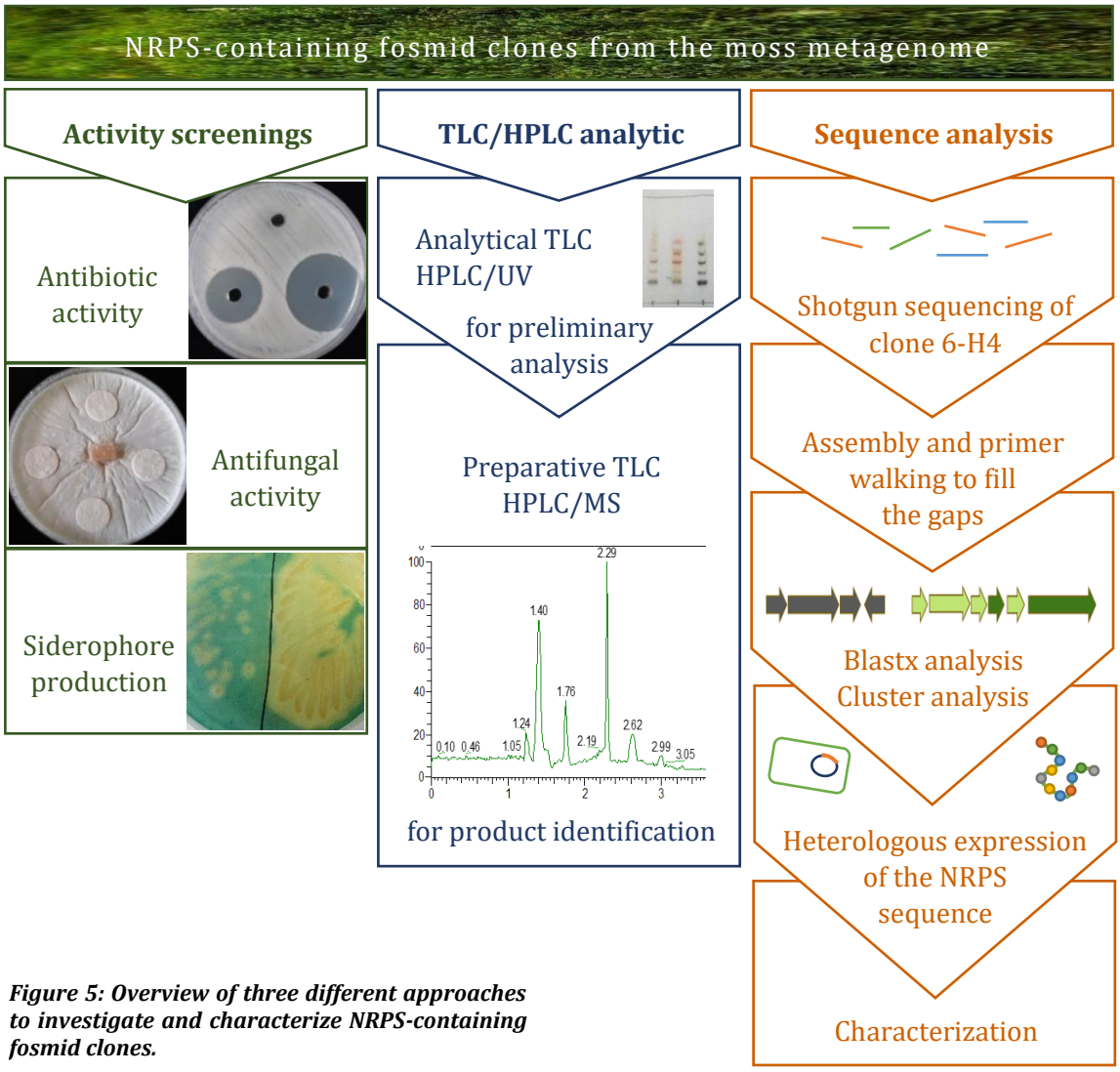


Figure 5: Overview of three different approaches to investigate and characterize NRPS-containing fosmid clones.

3. Material and Methods

All chemicals were of analytical grade and purchased from Sigma-Aldrich (Missouri, USA), Carl Roth (Karlsruhe, Germany), or Merck (Darmstadt, Germany) if not stated otherwise.

3.1 Bacterial strains, fungal strains and vector systems

3.1.1 *Escherichia coli* strains and vectors

E. coli strains (Table 1) were thawed from -70 °C stock cultures and spread on LB (lysogeny broth) agar plates (LB medium and agar-agar) or grown in LB medium (10 g/l peptone, 10 g/l NaCl, 5 g/l yeast extract), containing the appropriate antibiotic(s). Plates or liquid cultures (shaking, 120 rpm) were incubated overnight at 37 °C. After incubation, plates were stored at 4 °C until further usage. Table 1 also provides an overview of the vector systems used in this project.

Table 1: Escherichia coli strains and vector systems used in this project.

Escherichia coli	**Abbreviation**	**Antibiotics etc.**	**Reference**
EPI300-T1R	EPI300	/	Epicentre
K-12 *ΔentF*	*ΔentF*	Kan 25 µg/ml	Baba et al. 2006
K-12 *trfA* A250V *ΔentF*	*trfA* A250V *ΔentF*	Kan 25 µg/ml Tc 10 µg/ml	This work
S17.1(λpir) (containing pRS48)	S17.1(λpir)	Tc 10 µg/ml	Aavik et al. 2009
EC100D pir-116 (TransforMax)	pir-116	/	Epicentre
BAP1	BAP1	/	Pfeifer et al. 2001
Vector	**Information**		**Reference**
pCC2FOS	Cloning vector for Copy Control Fosmid library, 8181 bp Cm^R		Epicentre
pRS48	Suicide vector, encoding a mini-Tn 5 transposon and the *trfA* gene, 10.5 kb Tc^R, Amp^R		Aavik et al. 2009
pET-28a(+)	Bacterial expression vector, T7 promotor, *lacI* coding sequence, N- or C-terminal His-Tag, thrombin cleavage site 5369 bp, Kan^R		Novagen

Kan: Kanamycin; Tc: Tetracycline; Cm: Chloramphenicol; AmP: Ampicillin.

3.1.2 NRPS-containing fosmid clones (EPI300 pCC2FOS_NRPS)

E. coli EPI300 pCC2FOS_NRPS were provided by Christina Müller (Institute of Environmental Biotechnology, Technical University Graz). Clones were thawed from -70 °C stock cultures and

spread on LB agar plates or grown in LB medium, containing 12.5 µg/ml chloramphenicol (Cm) and fosmid auto-induction solution (AI 1x; Epicentre, Wisconsin, USA) or 0.01% (w/v) L-arabinose. Plates or liquid cultures (shaking, 120 rpm) were incubated overnight at 37 °C. After incubation, plates were stored at 4 °C until further use.

Clone-IDs were taken over from Müller et al. (2015) and are listed in Table 2, which also provides an overview of the closest hit from BlastX analysis and putative enzymes encoded by the NRPS sequences.

Table 2: NRPS fosmid clone IDs and the putative peptides encoded by the NRPS sequences. (Müller et al. 2015)

Clone-ID	Description and % identity (BlastX results)
2-C8	Non-ribosomal peptide synthetase module, partial (*Rubidibacter lacunae*) 86.2%
2-D4	Hypothetical protein, partial (*Kutzneria albida*) 47.9%
2-F4	Hypothetical protein, partial (*Pseudomonas syringae*) 64.2% Syringomycin synthetase E, partial (*Pseudomonas syringae*) 57.3%
3-F3	Peptide synthase (*Pseudomonas* sp. Ag1) 99.4% Pyoverdine sidechain peptide synthetase III, partial (*Pseudomonas amygdali*) 60.6%
3-G9	Non-ribosomal peptide synthetase (*Xenorhabdus doucetiae*) 74.4%
3-H3	Peptide synthetase (*Pseudomonas extremaustralis*) 91.1%
6-B1	Amino acid adenylation protein (*Dyella jiangningensis*) 72.8%
6-H4	Peptide synthetase (*Caulobacteraceae* bacterium PMMR1) 64.1%
7-B9	Hypothetical protein, partial (*Lysobacter antibioticus*) 62.9% Non-ribosomal peptide synthetase modules, partial (*Pseudomonas syringae*) 57.2%
7-C3	Hypothetical protein, partial (*Pseudomonas syringae*) 90.0%
7-D4	Thioester reductase (*Dyella jiangningensis*) 62.7% Gramicidin synthetase LgrC, partial (*Streptacidiphilus albus*) 50.0%
8-C8	Amino acid adenylation (*Rhodopseudomonas palustris*) 50.3%

3.1.3 Bacterial and fungal strains for antimicrobial activity assays

Fungal and bacterial strains from the culture collection (Technical University Graz) were initially thawed from -70 °C stock cultures and spread on appropriate agar plates. Bacterial strains were spread on nutrient agar plates (NB II and agar-agar) without antibiotics. *Staphylococcus epidermidis* and *Bacillus subtilis* were incubated overnight at 30 °C, while *Micrococcus luteus* was incubated for 2 to 3 days at the same temperature, due to the strain's slow growth. *Staphylococcus aureus* was grown at 37 °C. After incubation strains were kept at 4 °C until further use. Strains were regularly reactivated. *Rhizoctonia solani*, *Botrytis cinerea*, *Alternaria alternata*, *Fusarium oxysporum* and *Sclerotium rolfsii* were re-activated twice a month, placing an overgrown PDA agar-plug on a fresh PDA agar plate. Plates were incubated and stored at room temperature until use. *Candida albicans* was provided on a PDA agar plate and re-streaked

weekly on PDA agar (Potato Extract Glucose Bouillon and agar-agar). *C. albicans* was grown overnight at 30°C and then kept at 4 °C until further use.

3.2 Vector DNA isolation

Overnight cultures (ONCs) of *E. coli* strains harboring plasmids or fosmids were prepared in 5 ml LB medium containing the appropriate antibiotic(s) and incubated at 37 °C (shaking, 120 rpm) for 16 h. For fosmid isolation, it was necessary to add fosmid auto-induction solution (AI, 1x) or 0.01% (w/v) L-arabinose to the growth medium, to switch the fosmids from single-copy to high-copy number.

Cultures were harvested for 2 minutes at 6000 rpm and plasmid isolation was performed using the GeneJet Plasmid Miniprep Kit (Thermo Scientific, Massachusetts, USA), employing the manufacturer recommendations and minor modifications. Instead of elution buffer, 30 or 40 µl nuclease-free water (Thermo Scientific, Massachusetts, USA) was used for sample elution. For fosmid elution, the water was heated up to 70 °C and the column was incubated for 4 minutes before centrifugation. If necessary, clean up was performed using the Wizard SV Gel and PCR clean-up system (Promega, Wisconsin, USA). The concentration of isolated plasmids and fosmids was measured using spectrophotometry (NanoDrop 2000; Thermo Scientific, Massachusetts, USA). Isolated plasmids and fosmids were stored at -20 °C.

3.3 Restriction analysis of vectors

NRPS-containing vectors (pCC2FOS_NRPS) and the empty vector control (pCC2FOS_empty; harboring a small random pUC19 fragment) were analyzed using restriction digestion with FastDigest BamHI (Thermo Scientific, Massachusetts, USA) in FastDigest Green Buffer (1x, Thermo Scientific, Massachusetts, USA). Fosmid DNA (0.8 to 1 µg) was applied in a total reaction volume of 20 µl. The reaction mix was incubated for 30 minutes at 37 °C and the restriction enzyme was inactivated at 80 °C for 5 minutes. The size of pRS48 was controlled using FastDigest NdeI (Thermo Scientific, Massachusetts, USA) in 1x FastDigest Green Buffer. Fosmid DNA (0.8 to 1 µg) was used in a total reaction volume of 20 µl. The reaction mix was incubated for 15 minutes at 37 °C followed by inactivation at 80 °C for 5 minutes. Restriction analysis of pET-28a(+) was performed using EcoRV (New England Biolabs, Massachusetts, USA) in 1x NEBuffer 3.1 (New England Biolabs, Massachusetts, USA). Fosmid DNA (0.8 to 1 µg) was used in a total reaction volume of 50 µl. The reaction mix was incubated for 1 h at 37 °C and the enzyme was inactivated at 80 °C for 20 minutes.

Restriction fragments and uncut vectors were controlled via agarose gel electrophoresis.

3.4 Agarose gel electrophoresis

If not mentioned otherwise 1% (w/v) agarose (peqGOLD Universal-Agarose, VWR International Ltd., Erlangen, Germany) was melted in 1x TAE (TRIS acetate EDTA) electrophoresis buffer using a microwave and 1x TAE was used as running buffer as well. A 1 kb ladder (GeneRuler™, 0.1 µg/µl, 50 µg; Thermo Scientific, Massachusetts, USA) was used as a marker. Samples were mixed with loading dye (0.03% bromophenol blue, 0.03% xylene cyanol FF, 60% glycerol, 60 mM EDTA in 10 mM Tris-HCl, pH 7.6) before being loaded onto the gel. Time per run and voltage depended on the used chamber size, and were set between 45 to 75 minutes and 100 to 120 volt, respectively. Agarose gels were stained in ethidium bromide (0,0001%) for approximately 30 minutes. DNA fragments were visualized and photographed using UV transillumination.

3.5 Preparation of electrocompetent *E. coli* cells

Overnight cultures (ONCs) of *E. coli* were prepared in 5 ml LB medium containing the appropriate antibiotic(s) and incubated at 37 °C (shaking, 120 rpm) for 16 h. This ONC was used to inoculate (2% v/v) 400 ml LB medium containing the appropriate antibiotic(s) and incubated at 37 °C (shaking, 120 rpm) until an optical density at 600 nm (OD_{600}) of 0.5 to 0.7 was reached. Cells were cooled on ice for 20 minutes, and were then centrifuged at 4000 g for 15 minutes at 4 °C. The supernatant was discarded and cells were resuspended in 400 ml ice-cold 10% (v/v) glycerol, followed by a second centrifugation step. This process was repeated twice, reducing the amount of 10% glycerol first to 200 ml and then to 50 ml. After this washing process, cells were resuspended in 2 ml ice-cold 10% glycerol. 40 µl of the cells were transferred into Eppendorf tubes, frozen in liquid nitrogen and stored at -70 °C.

3.6 Siderophore production

3.6.1 Media and solutions

Iron-deficient MM9 liquid medium was prepared, according to Payne (1994). The basic medium contained 0.3 g/l KH_2PO_4, 0.5 g/l NaCl, 1.0 g/l NH_4Cl, 6 g/l NaOH, and 30.24 g/l PIPES. Upon autoclaving, 1 l of the medium was supplemented with the following filter sterilized (Minisart®,0.2 µM pore size; Göttingen, Germany) substances: 20 ml of a 20% (w/v) glucose solution, 1 ml of a 1 M $MgCl_2$ solution, 1 ml of a 0.1 M $CaCl_2$ solution and 30 ml of a 10% (w/v) deferrated casamino acid solution. Casamino acids were deferrated via extraction with an equal volume of 3% 8-hydroxyquinoline in chloroform. An additional extraction with chloroform was performed to remove traces of 8-hydroxyquinoline.

Chromazurol-S (CAS) assay and shuttle solution were prepared after Schwyn & Neilands (1987). A modified version was tested as well, where anhydrous piperazine was replaced with PIPES and the pH was raised to 5.6 with 50% (w/v) NaOH.

CAS agar plates were prepared after Louden et al. (2011), Fujita et al. (2011) and Rondon et al. (2000).

3.6.2 Generation of the strain *E. coli* K-12 *trfA* A250V *ΔentF* harboring pCC2FOS_NRPS

E. coli K-12 *ΔentF* is a strain from the Keio collection (Baba et al. 2006). Due to the knockout of the *entF* gene, this strain should not be able to produce the native *E. coli* siderophore enterobactin. In this study, in order to enhance the ability of the *trfA* gene to upregulate the fosmid copy number, a point mutation was inserted into the gene on the vector pRS48. After the mutated *trfA* A250V gene was inserted into *E. coli* K-12 *ΔentF*, the NRPS-containing fosmids were retransformed into the strain.

3.6.2.1 Verification that ΔentF lacks siderophore production

E. coli K-12 *ΔentF* and EPI300 were grown in 5 ml MM9 medium, containing the appropriate antibiotics for 16 h at 37 °C (shaking, 120 rpm). Those overnight cultures were used to inoculate 20 ml MM9 medium in a 100 ml Erlenmeyer flask (1%, v/v). After incubation at 37 °C for 16 h (shaking, 120 rpm), the liquid CAS assay was performed as described in chapter 3.6.4 to verify siderophore production.

3.6.2.2 Polymerase chain reaction to control the trfA gene on pRS48

A polymerase chain reaction (PCR) method was designed to verify the existence of the *trfA* gene on pRS48 and to control the insertion of the gene into the chromosome of *ΔentF* later on. pRS48 was isolated from S17.1(λpir) as described in chapter 3.2. The following primers were designed using CLC Main Workbench (QIAGEN Bioinformatics): trfA-F (5'-GGAAGGCATACAGGCAAG-3') and trfA-R (5'-GGTGCGAGCTGAAATAGT-3'). Primers were ordered from Sigma-Aldrich (Missouri, USA). The PCR reaction contained 12.5 µl Taq 2X Master Mix (New England Biolabs, Massachusetts, USA), 1.0 µl trfA-F (10 µM), 1.0 µl trfA-R (10 µM), 4.0 µl vector DNA (10 ng/µl) and 6.5 µl nuclease-free water. The PCR cycling conditions were as follows: Initial denaturation at 95 °C for 2 min, 30 cycles of denaturation at 95 °C for 20 s, annealing at 52 °c for 20 s, elongation at 68 °C for 1 min, followed by a final elongation at 68 °C for 5 min.

As a negative control, nuclease-free water was used instead of vector DNA. PCR results were evaluated via agarose gel electrophoresis (chapter 3.4).

3.6.2.3 Site-directed mutagenesis of the trfA gene on pRS48

The aim was to exchange adenine by valine at the amino acid residue 250 in the TrfA gene (Durland et al. 1990) using site-directed mutagenesis. Primers were designed using CLC Main Workbench (QIAGEN Bioinformatics) and denominated TrfA-A250V-F (5'-CTACAGGTGAC-GGCGATGGGCTTCAC-3') and TrfA-A250V-R (5'-GTGAAGCCCATCGCCGTCACCTGTAG-3'). The exchanged DNA base is underlined. Primers were ordered from Sigma-Aldrich (Missouri, USA). Different PCR mixtures and conditions were tested to achieve finally the desired mutation. The conditions that were tested and varied to achieve best results included the combination of a two-step PCR using different polymerases (Q5 HF DNA polymerase, Phusion DNA polymerase, Taq-&Go™ PCR Mix), varying annealing time (7, 8, 9 and 10 min), and gradient PCR (annealing temperatures from 48 to 68 °C). The two-step PCR consisted of a first step, where forward and reverse primer were pipetted in separate reactions for linear amplification over 3 to 5 PCR cycles, followed by combination of both reactions for exponential amplification over 15 to 30 PRC cycles. The final PCR reactions contained 5 µl Phusion HF Buffer (5x), 0.25 µl Phusion DNA Polymerase, 0.5 µl dNTPs (10 mM), 5 µl pRS48 (10 ng/µl), 0.75 µl DMSO, 13 µl nuclease-free water and 0.5 µl TrfA-A250V-F (10 µM) or TrfA-A250V-R (10 µM), respectively. Reactions without primers were used as negative control. The cycling conditions of the two-step PCR were as follows: Initial denaturation at 98 °C for 2 min, 5 cycles (first step) or 30 cycles (second step) of denaturation at 98 °C for 30 s, annealing at 65 °C for 1 min, elongation at 72 °C for 10 min, followed by a final elongation at 72 °C for 20 min (second step).

PCR reactions were analyzed with agarose gel electrophoresis as described in chapter 3.4. Phusion and Q5 polymerase were purchased from New England Biolabs (Massachusetts, USA), while Taq-&Go™ Ready-to-use PCR Mix was purchased from MP Biomedicals (California, USA).

3.6.2.4 Transformation of pRS48 trfA A250V into pir-116

After site-directed mutagenesis the mutated pRS48 plasmid was termed pRS48-*trfA* A250V. Prior to transformation, the reactions containing the amplified and mutated *trfA* A250V were combined. Amplicon and negative control were cleaned-up using the Wizard SV Gel and PCR clean-up system (Promega, Wisconsin, USA) and controlled by agarose gel electrophoresis (chapter 3.4). DpnI (10 U/l; Thermo Scientific, Massachusetts, USA) digestion was performed, to remove the unmutated pRS48 used as template in the PCR reaction. The reaction mix was prepared as recommended by the manufacturer and incubated for 16 h at 37 °C in a PCR cycler (lid heated to 40°C). Re-ligation of pRS48 *trfA* A250V was performed, using the protocol and the Fast-Link™ DNA Ligase from Epicentre. Subsequent, a clean-up was performed and controlled by agarose gel electrophoresis.

E. coli EC100D pir-116 (TransforMax™ electrocompetent *E. coli* cells; Epicentre, Wisconsin, USA) cells were thawed on ice for 10 minutes. 40 µl of the cells were mixed with 1 µl pRS48 *trfA* A250V or control reaction, respectively. After incubation for 1 minute on ice, the mix was transferred into a pre-cooled electroporation cuvette (Micropulser™ electroporation cuvettes, 0.1 cm gap; Bio-Rad, California, USA) and electroporation was performed using MicroPulser™ Electroporation Apparatus (Bio-Rad, California, USA). Program Ec1 (1.8 kV, 1 pulse) was used and the time constant after each electroporation was noted. Immediately following the pulse, 960 µl TB medium (24 g/l yeast extract, 12 g/l tryptone, 0.4% (v/v) glycerol, 17 mM KH_2PO_4, 72 mM K_2HPO_4) was added to the cuvette and the transformation reaction was transferred to a sterile Eppendorf tube. Reactions were incubated for 1 h at 37 °C (shaking, 120 rpm). A volume of 25 µl, 50 µl, 100 µl and the residual cell pellet (spinned down for 1 min at 4000 rpm) of the reaction were plated onto LB agar plates containing Tc (10 µg/ml) and Amp (100 µg/ml). Plates were incubated at 37 °C for 16 h.

3.6.2.5 Transformation of pRS48 trfA A250V into E. coli ΔentF

Before pRS48 *trfA* A250V was transformed into *ΔentF*, it was necessary to confirm the point mutation of the *trfA* gene. Five colonies from the pRS48 *trfA* A250V transformation plates were picked and the plasmids were isolated as described in chapter 3.2. The plasmids were send for sequencing at Microsynth (Balgach, Switzerland), using the trfA-R primer (chapter 3.6.2.2). To confirm the point mutation of the gene, alignments of the *trfA* A250V and wild type *trfA* DNA sequences were performed using CLC Main Workbench.

Upon confirmation of the mutation, pRS48 *trfA* A250V was transformed into *E. coli ΔentF* via electroporation. Electrocompetent *ΔentF* cells were prepared as described in chapter 3.5. Cells were thawed on ice for 10 minutes. Up to 900 ng pRS48 *trfA* A250V were mixed with the cells. Nuclease-free water was used as a negative control. A volume of 1 µl pUC19 (50 pg/µl) was used to determine transformation efficiency. The following steps were performed as described in the chapter 3.6.2.4 with minor modifications Reactions were incubated for 4 h at 37 °C (shaking, 120 rpm). A volume of 50 µl, 100 µl, 200 µl and residual cell pellet (spinned down for 1 min at 4000 rpm) of the reaction were plated onto LB agar plates containing Tc (10 µg/ml) and Kan (25 µg/ml). Plates were incubated at 30 °C for 16 h.

3.6.2.6 Verification of the transposon integration into ΔentF

The integration of the transposon containing the *trfA* A250V gene into the *ΔentF* chromosome was controlled with two different methods. ONCs of transformants were prepared in 5 ml LB Medium containing Tc (10 µg/ml) and Kan (25 µg/ml). Plasmid preparation and agarose gel

electrophoresis were performed as described in chapter 3.2 and 3.4, respectively. If the transposon was integrated into the chromosome, no plasmid DNA should be seen on the agarose gel. The same transformants were used for colony PCR, to verify the integration of the transposon into the *ΔentF* chromosome. Transformants were picked, resuspended in 15 µl nuclease-free water and incubated at 98 °C for 15 minutes. After a short centrifugation step (4000 rpm, 2 min), the supernatant containing the DNA was directly used for the PCR reaction. The cycling conditions were the same as described in chapter 3.6.2.2. The PCR reaction contained 12.5 µl Taq 2X Master Mix (New England Biolabs, Massachusetts, USA), 1.0 µl trfA-F (10 µM), 1.0 µl trfA-R (10 µM), 5.0 µl denaturated DNA and 5.5 µl nuclease-free water.

Upon verification of successful transposon integration, transformants were termed *E. coli* K-12 *trfA* A250V *ΔentF* and stock cultures were prepared and frozen at -70 °C.

3.6.2.7 Transformation of pCC2FOS_NRPS into E. coli K 12 trfA A250V ΔentF

Electrocompetent *trfA* A250V *ΔentF* cells were prepared as described in chapter 3.5 and thawed on ice for 10 minutes. Fosmid DNA (200-300 ng) was mixed with the cells. Nuclease-free water was used as a negative control and 1 µl pUC19 (50 pg/µl) was used to determine transformation efficiency. After incubation for 1 minute on ice, the mix was transferred into a pre-cooled electroporation cuvette (0.1 cm gap) and electroporation was performed using MicroPulser™ Electroporation Apparatus (voltage set manually at 1.3 kV). Immediately following the pulse, 960 µl TB medium was added to the cuvette and the transformation reaction was transferred to a sterile Eppendorf tube. Reactions were incubated for 1 h at 37 °C (shaking, 120 rpm). A volume of 50 µl, 100 µl, 200 µl and residual cell pellet (spinned down for 1 min at 4000 rpm) of the reaction were plated onto LB agar plates containing Kan (25 µg/ml) and Cm (12.5 µg/ml). Plates were incubated at 30 °C for 16 h.

Transformants were streaked on LB agar plates containing either Kan (25 µg/ml) and Cm (12.5 µg/ml) or Kan (25 µg/ml) and Tc (10 µg/ml) to confirm that the transposon was still integrated.

3.6.2.8 Enhancing the copy number of pCC2FOS_NRPS in trfA A250V ΔentF

trfA A250V expression from the integrated transposon was induced with *m*-toluate. Final *m*-toluate concentrations of 0.25 mM, 0.5 mM, 1 mM and 2 mM were tested. Clones were grown in 5 ml LB medium with Kan (25 µg/ml), Cm (12.5 µg/ml) and the respective *m*-toluate concentration. After incubation at 37 °C (shaking, 120 rpm) for 16 h, fosmid preparation, restriction analysis and agarose gel electrophoresis were performed as described in the chapters 3.2, 3.3 and 3.4.

3.6.3 Growth of EPI300 pCC2FOS_empty and *trfA* A250V *ΔentF* pCC2FOS_empty in MM9

ONCs of EPI300 pCC2FOS_epmty and *trfA* A250V *ΔentF* pCC2FOS_empty were grown in MM9 basic medium containing the appropriate antibiotics and additives (AI or 0.5 mM *m*-toluate, respectively). Clones were incubated for 16 h at 37 °C (shaking, 120 rpm). Those ONCs were employed on the one hand as sample for the 16 h measurement and on the other hand to inoculate (1% v/v) 20 ml of MM9 containing the appropriate antibiotics. The medium was either supplemented with 2 µM $FeCl_3$ or 2 µM $FeSO_4$ or lacked iron supplementation. Cultures were incubated at 37 °C (shaking, 120 rpm). After 2 h, 4 h, 6 h, 8 h and 24 h, samples were taken and OD_{600} measurements were performed. To the same time points samples were taken to perform the liquid CAS-assay as described in chapter 3.6.4.

3.6.4 Siderophore production in liquid culture

The MM9 medium was supplemented with 0.1 mM $FeSO_4$ or lacked iron supplementation. ONCs of EPI300 pCC2FOS_empty, EPI300 pCC2FOS_NRPS, *trfA* A250V *ΔentF* pCC2FOS_empty and *trfA* A250V *ΔentF* pCC2FOS_NRPS were grown in MM9 medium containing the appropriate antibiotics and additives (AI or 0.5 mM *m*-toluate, respectively). As a control strain, *Pseudomonas sp.* KA RM 1-1-4 (Institute of Environmental Biotechnology, Graz, Austria) was grown in MM9 without antibiotics. After incubation at 37 °C (shaking, 120 rpm), 30 °C (shaking, 110 rpm) or 25 °C (shaking, 200 rpm) for 16 to 24 h, 20 ml of MM9 (containing the appropriate antibiotics and additions) were inoculated with the ONCs (1%). Cultures were grown at 37 °C (shaking, 120 rpm), 30 °C (shaking, 110 rpm) or 25 °C (shaking, 200 rpm). Samples were taken after 2 h, 4 h, 6 h, 8 h, 16 h, 24 h and if not stated otherwise after 48 h and 72 h.

Samples were centrifuged (4500 rpm, 15 min) and the photometric CAS assay was performed using cell supernatants. A volume of 0.5 ml of supernatant was mixed with 0.5 ml CAS assay solution or CAS shuttle solution. Samples were incubated for 20 minutes at room temperature before the absorbance was measured at 630 nm with the U-2001 Spectrophotometer (Hitachi, Tokyo, Japan). As reference, a mixture of MM9 (with or without $FeSO_4$, respectively) and CAS assay or shuttle solution was prepared. MM9 without $FeSO_4$ was used as blank reaction. Samples were stored at room temperature and visually examined again after 24 h.

The absorbance of the references was set to 100%. Siderophore production was counted positive if there was a drop in absorbance, due to a color change from blue to orange/red.

3.6.5 Siderophore production on CAS agar plates

Strains were streaked on CAS agar plates in triplicates and incubated at room temperature up to two weeks. Plates were examined daily for orange halo formation around the colonies. Table 3 summarizes the different CAS agar plates and provides an overview of the used strains.

Table 3: Different agar plates recipes, supplements and strains used for the CAS agar plate assay

CAS plates	Supplements	Strains
after Louden et al. (2011)	Kan (25 µg/ml), Cm (12.5 µg/ml), *m*-toluate (0.5 mM)	*trfA* A250V *ΔentF* pCC2FOS_empty *trfA* A250V *ΔentF* pCC2FOS_NRPS
	without	KA RM 1-1-4
after Fujita et al. (2011)	Cm (12.5 µg/ml), AI (1x)	EPI300 pCC2FOS_NRPS
	without	EPI300, KA RM 1-1-4
after Rondon et al. (2000)	Cm (12.5 µg/ml), L-arabinose (0.01 %), $FeSO_4$ (0.1 mM)	EPI300 pCC2FOS_NRPS EPI300 pCC2FOS_empty
	$FeSO_4$ (0.1 mM)	KA RM 1-1-4

3.7 Antimicrobial activity assays

NRPS-containing fosmid clones were examined for their ability to produce antimicrobial substances. As test organisms, bacterial and fungal strains mentioned in chapter 3.1.3 were used.

3.7.1 Preparation of the test samples

The test samples for the antimicrobial activity assays, containing possible NRPS products, were prepared in different ways as described below.

Culture supernatants of EPI300 pCC2FOS_NRPS

EPI300 pCC2FOS_NRPS and the negative control EPI300 pCC2FOS_empty were grown in 5 ml LB medium containing Cm (34 µg/ml) and L-arabinose at 37 °C (shaking, 120 rpm) or 25 °C (shaking, 200 rpm) for 16 h. Those ONCs were used to inoculate (1%) 20 ml of TB medium, containing Cm (34 µg/ml) and L-arabinose. Cultures were incubated at 37 °C (shaking, 120 rpm) or 25 °C (shaking, 200 rpm) up to 72 h. Samples were taken after 24 h, 48 h and 72 h. Culture supernatants were obtained after centrifugation at 4500 rpm for 15 minutes at 4 °C. Supernatants were sterile filtrated (Minisart®, 0.2 µM pore size; Göttingen, Germany) and kept at 4 °C until further use.

Ethyl acetate extracts of EPI300 pCC2FOS_NRPS supernatants

Supernatants were obtained as described above. The incubation temperature for the ONCs and main cultures was 25 °C. Samples were only taken after 72 h of incubation. Supernatants were

acidified to pH 2 using 37% (v/v) HCl and extracted twice with the same volume of ethyl acetate (Roth, Karlsruhe, Germany). The ethyl acetate fraction was collected and vaporized using Rotovapor R 110 (water bath at 40 °C; Büchi, Switzerland). The dried extracts were resuspended again in 2 ml ethyl acetate and kept at 4 °C until further use.

Lyophilizates of EPI300 pCC2FOS_NRPS supernatants

EPI300 pCC2FOS_NRPS and the negative control EPI300 pCC2FOS_empty were grown in 5 ml M9 medium (12.8 g/l Na_2HPO_4-$7H_2O$, 3 g/l KH_2PO_4, 0.5 g/l NaCl, 1 g/l NH_4Cl, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.4% w/v glucose) containing 34 µg/ml Cm and L-arabinose at 37 °C (shaking, 120 rpm) for 16 h. Those ONCs were used to inoculate (1 %) 50 ml M9 medium, containing Cm and L-arabinose. Cultures were incubated at 37 °C (shaking, 120 rpm) for 16 h. Culture supernatants were obtained after centrifugation at 4500 rpm for 15 minutes at 4 °C. Supernatants were frozen in liquid nitrogen and lyophilized in a FreeZone 4.5 freeze dryer (Labconco, USA) for two days. The lyophilizates were resuspended in 2 ml M9 medium, filter sterilized (Minisart®,0.2 µM pore size) and stored at 4 °C until usage.

3.7.2 Well-diffusion assay using soft agar

Staphylococcus epidermidis and *Candida albicans* were grown in 5 ml LB medium or PDA medium, respectively. Cultures were incubated for 16 h at 37 °C (*S. epidermidis*, shaking, 120 rpm) or 30 °C (*C. albicans*; shaking, 110 rpm). For the assay, NB II or PDA agar was used, respectively. The soft agar had an agar concentration of only 8 g/l, instead of the regular 15 g/l. The agar was cooled down to 50 °C and 100 µl of bacterial or fungal suspension (ONCs) were added to 20 ml of agar. The suspension was quickly mixed and poured into petri dishes. For *C. albicans* some plates also contained 0.1 mM resazurin as redox indicator. The antimicrobial activity of EPI300 pCC2FOS_NRPS lyophilizates was investigated via agar-well test. Holes with a diameter of 6 mm were punched into the plates and filled twice with 50 µl of the test samples. The lyophilisate of EPI300 pCC2FOS_empty was used as a negative control. Ethanol (≥99,5%) and 50 µg Kan were used as a positive control. Plates were incubated at 37 °C for *S. epidermidis* or 30 °C for *C. albicans* for 16 h and then examined for growth inhibition around the wells.

3.7.2 Agar disc-diffusion and agar well-diffusion assay

Bacterial strains and *C. albicans* were freshly streaked on NB II agar plates or on a PDA agar plate, respectively. A bacterial or fungal suspension with a McFarland of 0.5 was prepared in 0.85 % (w/v) sterile NaCl. McFarland was adjusted by optical comparison with a 0.5 McFarland standard (bioMérieux, Marcy-l'Étoile, France) placed in front of a Wickerham card. This sus-

pension was used to inoculate Mueller-Hinton agar (2 g/l beef extract, 17.5 g/l casein hydrolysate, 1.5 g/l starch from potatoes, 17 g/l agar, pH 7.0). Mueller-Hinton agar plates were standardized by pouring exactly 25 ml of agar, leading to plates with an agar thickness of 4 mm. For *C. albicans*, Mueller-Hinton agar was supplemented with 2% glucose and in some cases also with 0.5 µg/ml methylene blue (NCCLS 2004). The plates were inoculated using sterile non-toxic cotton swaps on a wooden applicator (Raucotupf®, 15 cm, small; Lohman & Rauscher, Vienna, Austria). Cotton swabs were dipped into the inoculum and the excess fluid was pressed out by rotating the soaked swabs firmly against the inside wall of the inoculum tube. The entire agar surface was streaked with the same swab three times, turning the plate at an approximately 60 °C angle between each streaking. The inoculum was dried for 5 minutes.

For the agar well-diffusion assay, 6 mm holes were punched into the inoculated Mueller-Hinton plates and filled with 100 µl culture supernatants (chapter 3.7.1). Kanamycin was used as a positive control (100 µl of a 0.5 mg/ml stock solution). Plates were incubated at 25 °C until the growth of the strains was strong enough to determine the inhibition zone of the positive control. Inhibition zones were measured in mm.

The blotting paper discs for the agar disc-diffusion assays were prepared after Laireiter et al. (2014). For testing the culture supernatants, round discs with a diameter of 1.6 cm were perforated out of blotting paper and then sterilized by autoclaving. Culture supernatants (150 µl) were pipetted onto the disc and then dried in an open sterile petri-dish for 16 h at 25 °C or 30 °C. As a positive control blotting paper discs with 50 µg Kan were used. Filter discs with a diameter of 6 mm were used for the ethyl acetate extracts (chapter 3.7.1). Extracts (50 µl) were pipetted onto the discs and then dried in an open sterile glass petri-dish for 16 h at 25 °C. The dried paper discs were placed onto the inoculated Mueller-Hinton agar and the plates were incubated at 25 °C, 30°C or 37 °C (depending on strain and experiment), until the growth of the strains was strong enough to determine the inhibition zone of the positive control. Inhibition zones were measured in mm. If not used immediately, the dried paper discs were stored in plastic bags at 4 °C.

PDA agar plates (17 g/l agar; 4 mm thickness) were used for testing the filamentous fungi. The agar well- and disc-diffusion assays were performed as described above, but without addition of Kan as a control. For the agar well-diffusion assay, also the lyophilizates (100 µl) were employed. Overgrown agar-plugs of the different fungi were placed on the plates after the test samples were applied. The plates were incubated at room temperature until the growth of fungi reached the rim of the petri-dishes.

3.8 Thin-Layer Chromatography

Samples for thin-layer chromatography (TLC) were prepared as the ethyl acetate extract samples described in chapter 3.7.1, with minor modifications. ONCs and main cultures were grown in MM9 medium. After 16 h of incubation, 50 ml of the cultures were centrifuged and the supernatants were used for ethyl acetate extraction as described previously. Dried extracts were resuspended in 0.5 to 1 ml ethyl acetate.

TLC Silica gel 60 F_{254} and TLC Silica gel 60 RP-18 F_{254}s (both Merck, Darmstadt, Germany) were used for analytical and preparative TLC. Various solvent systems were tested. Ethyl acetate (100%) and ethyl acetate:acetic acid:water (14:3:3) for silica plates and acetonitrile:water:formic acid (40:60:0.1) for reverse phase (RP) plates displayed good separation and were used for the analysis of the samples. Upon separation, plates were examined under UV light or stained with either ninhydrin stain (0.1 g ninhydrin, 0.5 ml acetic acid, 100 ml acetone) or iron-stain (0.1 M $FeCl_3$ in 0.1 M HCl).

Preparative TLC was performed with the extracts of clone 6-H4 and the control EPI300 pCC2FOS_empty. Upon separation, those bands visible under UV light were marked and extracted from the TLC plates (with methanol and 0.1% formic acid). The extracted bands were analyzed using high performance liquid chromatography/mass spectrometry as described in chapter 3.9.2.

3.9 High Performance Liquid Chromatography

Samples for high performance liquid chromatography (HPLC) were prepared in the same way as described for TLC in chapter 3.8, with only one modification: the dried extracts were resuspended in methanol instead of ethyl acetate.

3.9.1 HPLC coupled with UV-detection

Samples were analyzed using a C-18 silica column in a Dionex UltiMate 3000 HPLC System (Thermo Scientific, Massachusetts, USA). The mobile phase consisted of aqueous formic acid (0.1%, v/v, solvent A) and acetonitrile (100%, solvent B). Starting conditions for the gradient elution were 10% A and 90% B. The conditions were gradually changed to 80% A and 20% B within 30 minutes. This process was followed by 5 minutes at 10% A and 90% B (readjustment to initial conditions). The eluent flow rate was set at 0.7 ml/min and the column temperature was maintained at 25 °C. Samples were analyzed using UV-detection and the results were compared manually and examined for differences.

3.9.2 HPLC coupled with mass spectrometry (MS)

Samples were analyzed with a combined HPLC-hybrid quadrupole-orbitrap mass spectrometer (Q Exactive; Thermo Scientific, Massachusetts, USA). An Atlantis dC18 3 µM 2.1×100 mm column (with 1 cm precolumn) was used to separate the different metabolites from the culture supernatant extracts. Acetonitrile was used as solvent A and aqueous formic acid (0.1%, v/v) as solvent B. Starting conditions for the gradient elution were 10% A and 90% B. The conditions were gradually changed to 80% A and 20% B within 20, 30 or 40 minutes. This process was followed by 5 minutes at 10% A and 90% B (readjustment to initial conditions). The eluent flow rate was set at 0.3 ml/min and the column temperature was maintained at 25 °C. Samples were analyzed with positive and negative ion HESI detection. HESI conditions were set to 3 kV spray voltage and 330 °C capillary temperature for negative mode and 3.5 kV spray voltage and 250 °C capillary temperature for positive mode. The scans for both modi were recorded in the range 100.0-1500.0 m/z, an orbitrap resolution of 70,000, AGC values of 1×10^6 and maximal accumulation time of 200 ms. ddMS2 cycles (resolution 17,500; AGC values of 1×10^5; max. accumulation time 100 ms; NCE stepped 15, 20, 25) followed the full MS.

HPLC/MS data were analyzed with Xcalibur™ and Compound Discoverer 2.0 (Predicted formula, SFit; Thermo Scientific, Massachusetts, USA). Suggestions from ChemSpider and KEGG were further examined by manual comparison with mzCloud.

3.10 Sequence analysis of fosmid clone 6-H4

Clone 6-H4 was chosen for whole-insert sequencing due to preliminary results from a liquid siderophore assay in a previous study.

3.10.1 Shotgun-sequencing, primer walking and sequence analysis

In a previous project, fosmid clone 6-H4 was shared using ultrasound to create fragments with a size between 1 to 2 kb. A clone library was generated and sent for sequencing at Eurofins (Luxemburg). Sequences were analyzed using CLC Main Workbench (QIAGEN Bioinformatics) and primers were designed with the same program to fill the gaps between the assembled contigs. Those primers are listed in Attachment, Table A1.

In this study the assembled insert sequence of clone 6-H4 was analyzed for potential open reading frames (ORFs) using NCBI ORFfinder. Blast2Go Pro (Götz et al. 2008) and WebMGA (Wu et al. 2011) were used for annotation of the putative genes. For the Blast2Go Pro NCBI blast, the settings were adopted from blastp at NCBI, using a Blast expectation value threshold of 10 and a word size of 6, while the number of Blast hits, the HPS length cutoff and the HPS-

hit coverage remained at the default values of 20, 33 and 0, respectively. Blastp was performed against the non-redundant database from GenBank. The Blast description annotator was turned on, while the low complexity filter was turned off. For the Cloudblast analysis (blastp), the putative ORFs of clone 6-H4 were blasted against the non-redundant database, applying the default settings (Blast expectation value threshold of 1.0E-3 and a word size of 3). Blast description annotator was turned on, as was the low-complexity filter. Upon performing an InterPro Scan with all options available, the identified GO terms of this scan were merged with the GO terms detected by Blast2Go Pro Mapping. Annotation was performed using the default settings. ANNEX (Annotation Expander) was applied subsequently and the final step of this analysis process was the execution of GO-Slim, using the Metagenomics Slim option. To visualize the summarized annotations after GO-Slim, combined GO graphs were made, with the GO level set to 1 and intermediate GO terms filtered.

The domain structure of potential proteins was analyzed with NCBI Conserved Domain Search. The software antiSMASH (Medema et al. 2011) was used for cluster identification and SOSUI (Hirokawa et al. 1998) , TmPred (Hofmann & Stoffel 1993), THMM (Krogh et al. 2001) and Kyte & Doolittle Hydrophobicity plot (ProtScale, Expasy) were applied for the identification of transmembrane segments. MEGA7 was used for phylogenetic analysis (Hall 2013). A phylogenetic tree was constructed with this software, using ClustalW for sequence alignment (Multiple Alignment Gap Opening penalty 3 and Multiple Alignment Gap Extension penalty 1.8 [Hall 2013]) and Neighbor-Joining algorithm for tree generation (bootstrap of 1000 replicates, number of differences method and complete deletion of gaps/missing data).

For the identification of possible substrates for the A-domain of the putative NRPS of clone 6-H4, NRPSpredictor2 (Röttig et al. 2011) was used.

3.10.2 Generation of pET28a(+) 6-H4_Synth expression vector and transformation

The 6-H4_Synthetase gene was cloned into pET28a(+) via PLICing (Blanusa et al. 2010). The used phosphorothioated oligonucleotide (PTO) primers (Sigma-Aldrich, Missouri, USA) were 6-H4_Synth_PLIC_F (5'-GGACAGCAAATGGCGCGGCGCTCAGAGCTG-3'), 6-H4_Synth_PLIC_R (5'-GGGCTTTGTTAGTCACATCGATTCCCCATGCAAACC-3'), pET28a(+)_PLIC_F (5'-CTAACAAAGCCCCGAAAGGAAGCTGAGTTGG-3') and pET28a(+)_PLIC_R (5'-CATTTGCTGTCCACCAGTCATGCTAGCCATATG-3'). The complementary phosphorothioated nucleotides are underlined.

The PCR reactions contained 5 µl Phusion HF Buffer (5x), 0.25 µl Phusion DNA Polymerase, 0.5 µl dNTPs (10 mM), 0.5 µl pET28a(+) or fosmid 6-H4 (10 ng/µl), 1.5 µl DMSO, 16.75 µl nuclease-free water and 0.5 µl of one of the above-mentioned primers (10 µM). A two-step PCR was

performed (chapter 3.6.2.3) and consisted of a linear amplification step of 5 PCR cycles, followed by the combination of the separate PCR reactions for exponential amplification over 25 cycles. Reactions without primers were used as negative control. The cycling conditions were as follows: Initial denaturation at 98 °C for 30 s, 5 cycles (first step) or 25 cycles (second step) of denaturation at 98 °C for 20 s, annealing at 58 °C for 30 s, elongation at 72 °C for 3 min, followed by a final elongation at 72 °C for 10 min (second step). A gradient two-step PCR with annealing temperatures ranging from 58 °C to 72 °C and elongation at 72 °C for 2 min was additionally performed for fosmid 6-H4. PCR reactions were cleaned-up using the Wizard SV Gel and PCR clean-up system (Promega, Wisconsin, USA) and controlled by agarose gel electrophoresis (chapter 3.4). Prior to the clean-up, DpnI (10 U/l; Thermo Scientific, Massachusetts, USA) digestion was performed, to remove the wild type template in the pET28a(+) PCR reaction. The reaction mix was prepared as recommended by the manufacturer and incubated for 16 h at 37 °C in a PCR cycler.

The phosphorothioate-based ligase-independent cloning (PLICing) was performed as described by Blanusa et al. 2010. Vector to insert ratios were 1:3, 1:6, and 1:10, with vector concentrations between 0.01 and 0.02 pmol/µl and insert concentrations between 0.06 and 0.2 pmol/µl. For the iodine treatment, 1 µl of cleavage buffer (0.5 M Tris–HCl, pH 9.0), 0.6 µl of iodine stock solution (100 mM iodine in 99% ethanol), and 0.4 µl of nuclease-free water were mixed with 4 µl vector or insert DNA, respectively, and incubated for 5 minutes at 70 °C in a PCR cycler. After this step, the same volume of vector and insert DNA were mixed and incubated for 5 minutes at room temperature, before 1 µl of the mixture was transformed into electrocompetent *E. coli* BAP1 cells. As a control, 1 µl of iodine-treated vector (0.02 pmol/µl) was transformed. After incubation for 1 h at 37 °C, 25 µl, 50 µl, 100 µl of the reactions and the residual cell pellet (spinned down for 1 min at 4000 rpm) were plated onto LB agar plates containing Kan (50 µg/ml). Plates were incubated at 37 °C for 16 h.

Several colonies were picked and restriction analysis was performed as described in chapter 3.3.

3.10.3 Expression of 6-H4_Synth in BAP1

For the expression of the 6-H4 synthetase, 20 ml TB medium containing 50 µg/ml Kan were inoculated with 2% (v/v) of an overnight culture (30 °C) of *E. coli* BAP1 6-H4_Synth or *E. coli* BAP1 pET28a(+)_empty, respectively. Cultures were incubated at 37 °C until an OD_{600} of 0.5 or 1 was reached. Expression was then induced with 0.4 or 1 mM isopropyl-β-D-thiogalactopyranosid (IPTG) and cultures were further incubated at 20, 25 or 30 °C. Samples (2 ml) were collected after 2 h, 4 h and 20 h upon induction.

After centrifugation (4500 rpm, 10 min, 4 °C), the cell pellet was resuspended in 200 µl PBS buffer (137 mM, NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4) and transferred into tubes with 20% (v/v) lysis matrix (0.09-0.15 mm). Cells were lysed using a FastPrep®-24 Instrument (MP Biomedicals, California, USA), applying 6.0 m/s for 60 seconds. Lysates were centrifuged at maximum speed and 4 °C for 15 minutes and the supernatant was transferred into a fresh tube. The remaining cell pellet was resuspended in 100 µl PBS buffer. Cell free lysate and resuspended cell pellet were stored at 4 °C.

3.10.4 SDS-PAGE analysis

Samples were denatured in 4x Laemmli buffer for 5 minutes at 98 °C. SDS-polyacrylamide gels were loaded with 5 to 10 µl of the denatured samples per slot and compared to a molecular weight marker (PageRuler Prestained Protein Ladder; Fermentas, St. Leon-Rot, Germany). Samples were separated for 75 minutes at 30 mA in a Mini-Protean Tetra Cell system from Bio-Rad (München, Germany). Gels were stained in a Coomassie Brilliant Blue solution for 20 minutes, followed by a destaining step in a 10% (v/v) acetic acid solution overnight.

The compositions of the used buffers, the SDS-polyacrylamide gels, as well as the Coomassie Brilliant Blue stain are listed in Attachment, Table A2.

4. Results

4.1 Siderophore production of the NRPS fosmid clones

In a previous study thirteen fosmid clones containing NRPS-related sequences were identified in the *S. magellanicum* metagenome. The possible production of siderophores by the putative synthetase genes contained in the fosmid clones should be analyzed employing the CAS-assay. Since the host strain *E. coli* EPI300 produces the siderophore enterobactin, an *E. coli* deletion mutant (*ΔentF*) should be employed instead. But, unlike EPI300, K-12 *ΔentF* does not produce the protein TrfA, which is necessary to enhance the copy number of the cloning vector pCC2FOS. An enhanced copy number may lead to more heterologous product production, which may be easier to detect. In order to enhance the ability of the *trfA* gene to upregulate the fosmid copy number, a point mutation at the amino acid position 250 was inserted into the gene on the vector pRS48. After the mutated *trfA* A250V gene was inserted into *E. coli* K-12 *ΔentF*, the NRPS-containing fosmids could be retransformed into the strain.

4.1.1 Successful generation of *E. coli* K-12 *trfA* A250V *ΔentF* pCC2FOS_NRPS

Plasmid pRS48 harbors a wild type *trfA* gene under a PmG5 promotor. This promotor is activated by XylS in the presence of *m*-toluate (Aakvik et al. 2009). The problem is, that wild type TrfA protein is not capable to raise plasmid copy numbers in *E. coli* to a high extent. Durland et al. (1990) identified several point mutations in the *trfA* gene sequence which led to an enhancement of the copy number of the RK2 plasmid – one of those mutations was a replacement of adenine with valine at the 250th position of the TrfA amino acid sequence. This amino acid substitution in pRS48 TrfA should be achieved by site-directed mutagenesis – a difficult process due to the large size of pRS48 (10 kb).

Upon verifying that *ΔentF* lacked siderophore production (Fig. 6), and that pRS48 harbored the *trfA* gene (data not shown), site-directed mutagenesis of the *trfA* gene was performed as described in chapter 3.6.2.3. After transformation of pRS48 *trfA* A250V into *E. coli* pir-116, five colonies were picked, the mutated plasmid was re-isolated and sent for sequencing to confirm the point mutation.

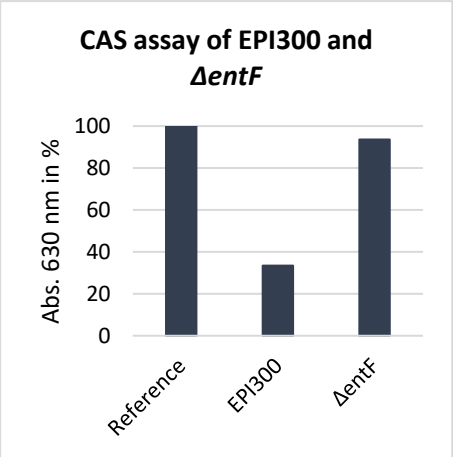


Figure 6: CAS assay of ΔentF and EPI300. *The absorbance of the reference was set to 100%. Siderophore production was counted positive if there was a strong drop in absorbance.*

The sequences of the five pRS48 *trfA* A250V were aligned with the wild type *trfA* sequence, which originated from the plasmid pJB658 (Aakvik et al. 2009). All five plasmids showed the point mutation from a cytosine at the 749th position of the *trfA* gene sequence to a tyrosine, resulting in the substitution of adenine in 250th position of TrfA to valine, which was the desired point mutation (Fig. 7).

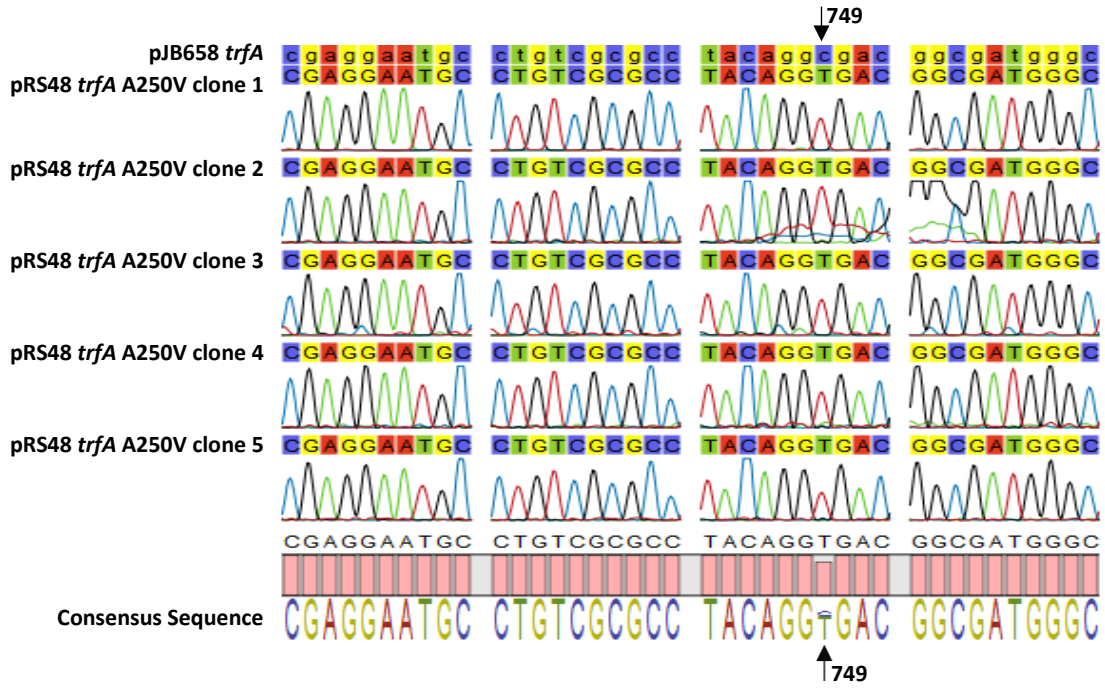


Figure 7: Site-directed mutagenesis of the trfA gene on pRS48. *Mutagenesis was controlled by sequencing and subsequent alignment of the sequences to the wild-type trfA gene on pJB658. An exchange of cytosine to thymine at the 749th position of the trfA gene sequence was desired.*

The plasmid pRS48 *trfA* A250V of clone 1 was chosen for transformation into *E. coli ΔentF*. Since no pir-protein for pRS48 replication is produced in *ΔentF,* the transposon with the mutated *trfA* sequence should be integrated into the genome of *ΔentF.* This was controlled, on one hand, with vector isolation and agarose gel electrophoresis. If the transposon was successfully integrated into the chromosome, no vector DNA should be re-isolated and be visible on the gel. This was true for all six tested transformants (data not shown). Additionally, a colony PCR using the *trfA* primers (chapter 3.6.2.2) was performed to further verify the integration of the *trfA* A250V gene. A PCR product of the size of 960 bp was expected and was detected in all examined transformants (Fig. 8).

The NRPS-containing fosmids listed in Table 2 were isolated and transformed into electrocompetent *trfA* A250V *ΔentF* 1. All NRPS-fosmids, as well as pCC2FOS_empty, were successfully transformed by electroporation and the transposon was stably integrated into the chromosome, without antibiotic pressure (data not shown; see chapter 3.6.2.7 for the method).

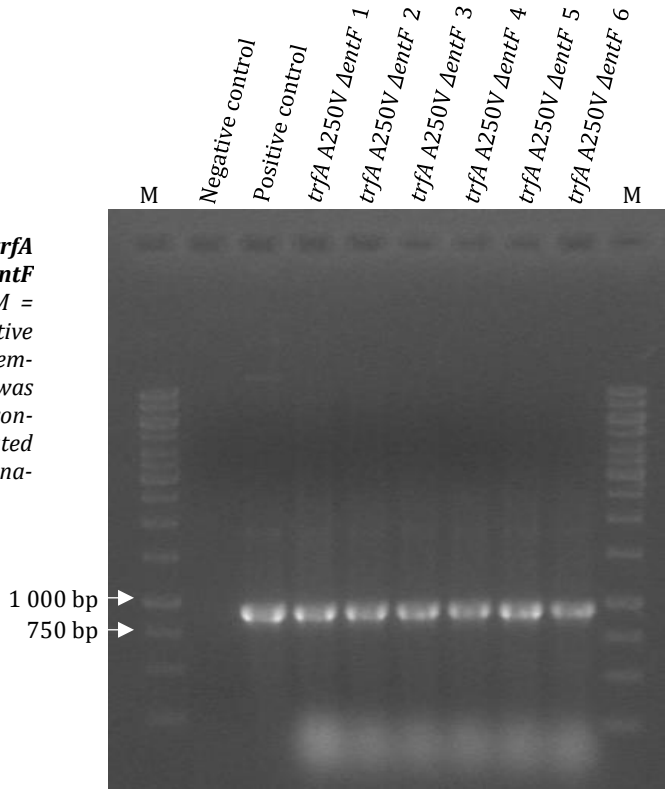


Figure 8: Verification of trfA A250V integration into the ΔentF chromosome by colony PCR. *M = GeneRuler™ 1 kb ladder. As negative control a PCR reaction without template was used. Isolated pRS48 was used as template for the positive control. Six transformants, designated trfA A250 ΔentF 1 to 6, were analyzed.*

The new, successfully generated *E. coli* K-12 clones containing the NRPS-fosmids were designated *trfA* A250V *ΔentF* pCC2FOS_NRPS and *trfA* A250V *ΔentF* pCC2FOS_empty.

4.1.2 Copy number enhancement in *trfA* A250V *ΔentF* pCC2FOS is not detectable

Like in Aakvik et al. (2009), *m*-toluate was used to activate the transcription of the mutated *trfA* gene. Different concentrations of the inducer were tested, but none showed a visible enhancement of the pCC2FOS_NRPS copy number in *trfA* A250V *ΔentF*, at least not with NanoDrop measurements (not shown) and agarose gel electrophoresis (Fig. 9).

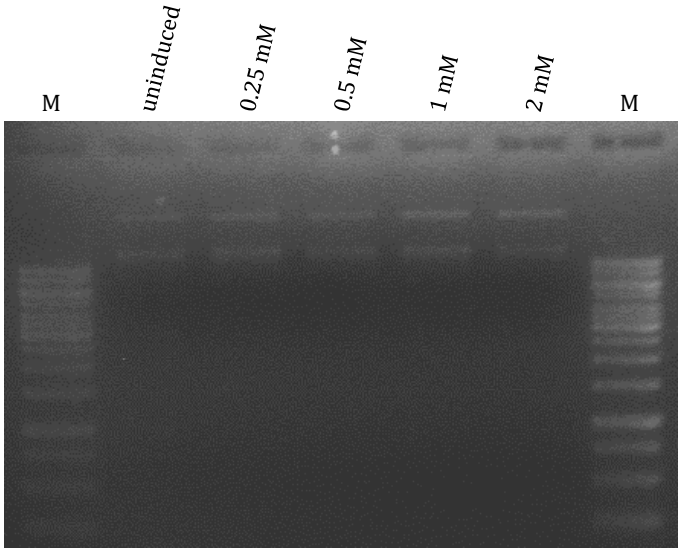


Figure 9: Fosmid 2-C8 isolated from trfA A250V ΔentF after induction with m-toluate. *M = GeneRuler™ 1 kb ladder. Cultures were incubated for 16 h prior to fosmid isolation. An uninduced culture served as a reference value. None of the tested final m-toluate concentrations considerably enhanced the fosmid copy number.*

4.1.3 Reduced growth of EPI300 and *ΔentF* when grown in iron-deficient medium

Although no enhancement of the fosmid copy number in *trfA* A250V *ΔentF* upon induction was detectable, 0.5 mM m-toluate were added to the growth medium of this strain. With the experimental set up described in chapter 3.6.3, the growth rate of both strains when grown in iron-deficient MM9 was about 50% reduced in comparison to their growth in the same medium supplemented with iron (Fig. 10). This growth inhibition is due to the iron starvation when grown in iron-deficient MM9 (Schwyn & Neilands 1987) and was the desired effect. Therefore, the iron-deficient MM9 medium described in chapter 3.6.1 was suitable for siderophore production analysis using CAS assay.

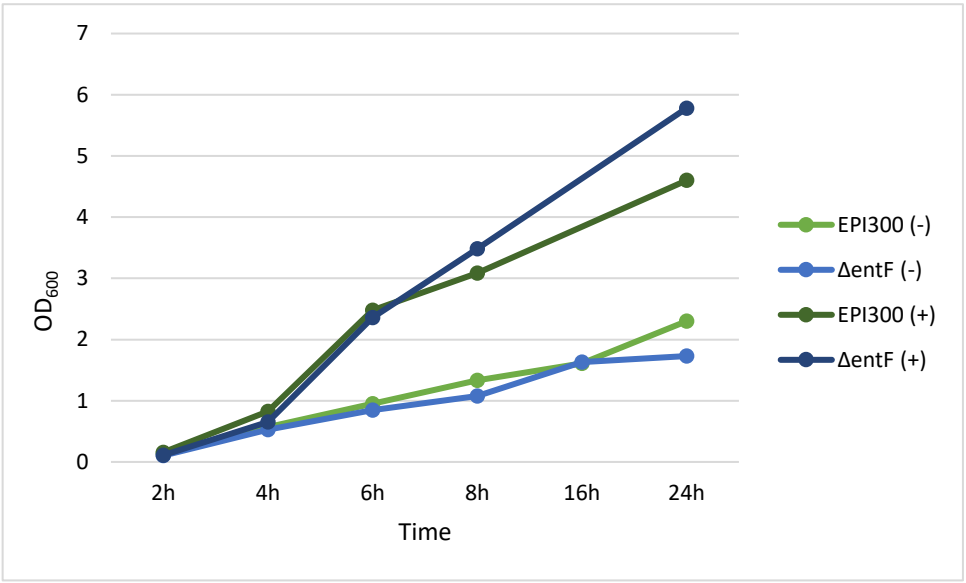


Figure 10: Growth of E. coli EPI300 and ΔentF measured as OD_{600} using iron-deficient MM9 with or without 2 µM $FeSO_4$ supplementation. *(+) = with iron, (-) = without iron.*

4.1.4 Investigation of fosmid clone siderophore production using liquid CAS assay

The liquid CAS assay was performed with both type of clones, *trfA* A250V *ΔentF* pCC2FOS_NRPS and EPI300 pCC2FOS_NRPS as described in chapter 3.6.4. Since the copy number enhancement was not visible in *trfA* A250V *ΔentF*, leading to a possibly low siderophore production that may not be detectable, it was decided to perform the CAS assay with both strains. To overcome the problem of native enterobactin production, EPI300 pCC2FOS_NRPS were grown in MM9 supplemented with 0.1 mM $FeSO_4$ (Rondon et al. 2000). EPI300 pCC2FOS_empty and *Pseudomonas sp.* KA RM 1-1-4 in iron-deficient MM9 were used as positive controls for siderophore production. Figure 11 shows the results of the CAS assay after 48 h of growth in MM9. Both, *trfA* A250V *ΔentF* pCC2FOS_NRPS (Fig. 11A) and EPI300 pCC2FOS_NRPS (Fig. 11B) showed no siderophore production.

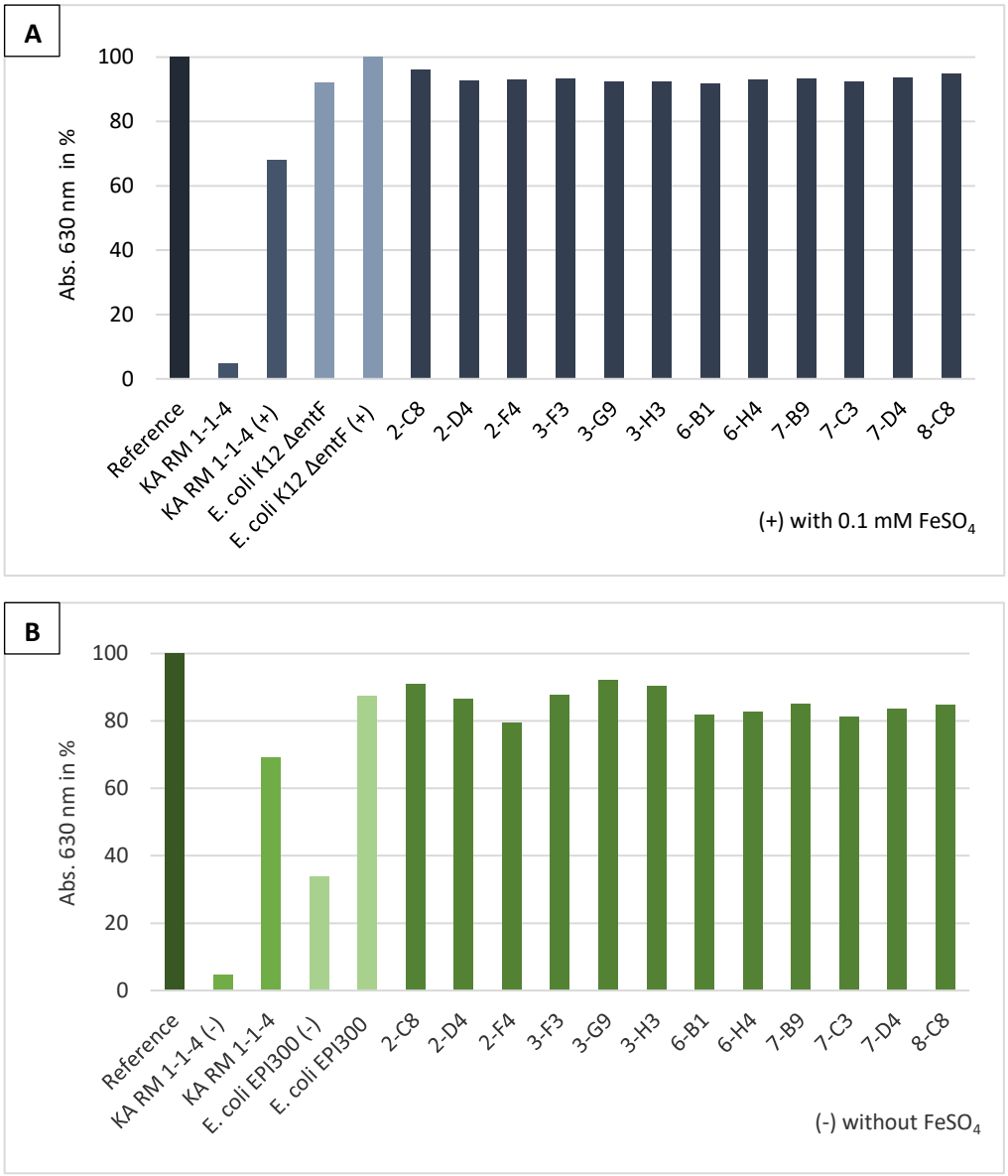


Figure 11: Siderophore detection assay (CAS assay) of trfA A250V ΔentF pCC2FOS_NRPS (A) and EPI300 pCC2FOS_NRPS (B) after 48 h of growth. *The absorbance of the reference was set to 100%. Siderophore production was counted positive if there was a strong drop in absorbance. Results are shown as means of triplicate measurements.*

Positive control KA RM 1-1-4 showed a strong drop in absorbance when grown in iron-deficient MM9 only after 48 h and 72 h of incubation. The reason therefore was the slow growth of the *Pseudomonas* sp. in this medium. The addition of 0.1 mM $FeSO_4$ to the growth medium of EPI300 pCC2FOS_NRPS and pCC2FOS_empty successfully suppressed the enterobactin production of the strain (Fig. 11B), even after 72 h of incubation (data not shown). As already mentioned, none of the NRPS fosmid clones exhibited siderophore production at none of the time points CAS assay was performed (data not shown for 16 h, 24 h and 72 h measurements).

4.1.5 Siderophore production of the NRPS fosmid clones on CAS agar plates

In addition to the liquid CAS assay, different CAS agar plate assays were performed, where the plates were incubated at room temperature for two weeks, to detect a possible late siderophore production. The plates prepared after Fujita et al. (2011) were LB agar plates supplemented with CAS assay solution. Since those plates were not iron-deficient minimal media plates, enterobactin production of EPI300 clones was slightly suppressed, but a small orange halo was still visible. None of the EPI300 pCC2FOS_NRPS clones showed an enhanced halo formation, indicating heterologous siderophore production (data not shown). The other used CAS agar plates were based after Louden et al. (2011). For EPI300 fosmid clones, the agar medium was again supplemented with 0.1 mM $FeSO_4$ to suppress native siderophore production (Rondon et al. 2000). KA RM 1-1-4 was streaked on CAS plates with and without iron supplementation. As visible in Figure 12A, the *Pseudomonas* sp. exhibits strong orange halo formation when grown on agar without iron, while $FeSO_4$ suppresses the siderophore production of the strain completely (Fig. 12B). Like in the liquid CAS assay, none of the NRPS fosmid clones showed signs of siderophore production (Fig. 12C and D), even after 14 days of incubation.

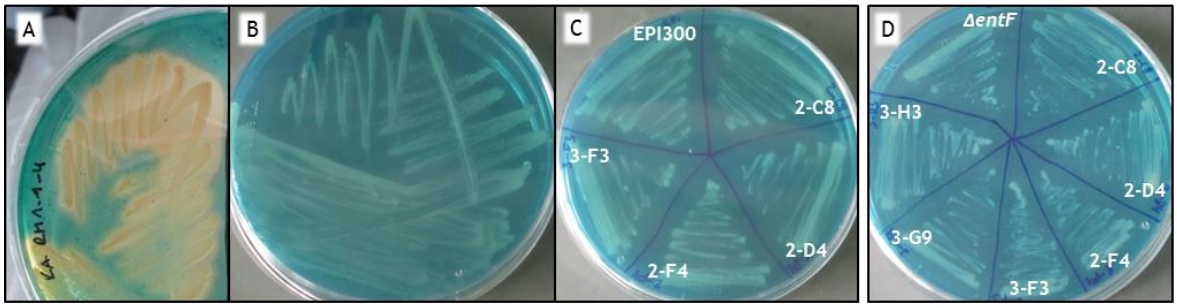


Figure 12: Siderohpore detection on CAS agar plates of clones trfA A250V ΔentF pCC2FOS_NRPS and EPI300 pCC2FOS_NRPS. *Pictures taken after 3 days of incubation. Pseudomonas sp. KA RM 1-1-4 on a CAS plate without iron supplementation (A) and with 0.1 mM $FeSO_4$ (B). EPI300 pCC2FOS_empty and NRPS clones 2-C8, 2-D4, 2-F4, 3-F3 on a CAS plate with 0.1 mM $FeSO_4$ (C). trfA A250V ΔentF pCC2FOS_empty and NRPS clones 2-C8, 2-D4, 2-F4, 3-F3, 3-G9, 3-H3 on a CAS plate without $FeSO_4$.*

4.2 Analyzing the NRPS clones for production of antimicrobials

Since nonribosomal peptide synthetases are known to produce antibiotics, like the in chapter 1.4.2 mentioned vancomycin, another goal of this project was to examine the NRPS clones for the production of antimicrobials. Different bacterial and fungal strains (see chapter 3.1.3) were used as test organisms. Methods, like cross-streaking the bacterial test organisms and the NRPS clones, or agar-overlay assays could not be employed, since the metagenome clones need chloramphenicol (Cm) to maintain the fosmids, and all the bacterial test strains were Cm sensitive. Although the fungal strains could have been placed on Cm plates to perform dual culture assays, it was decided to examine possible production of antifungals of the NRPS clones with the same methods as for the bacterial strains – using agar disc- and agar well-diffusion assays.

4.2.1 Clone 2-F4 exhibited antimicrobial activity in one round of testing

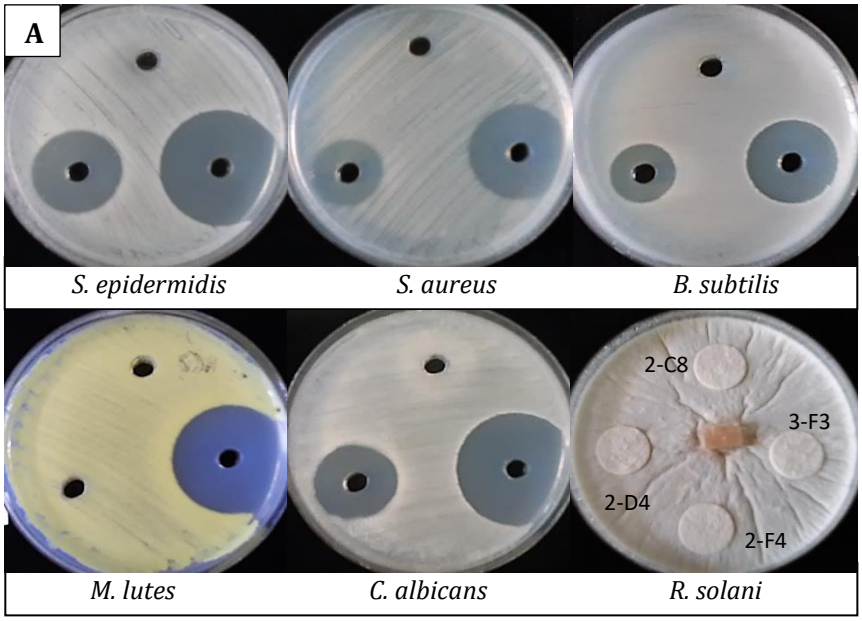
Different test samples from NRPS clones were prepared to enhance the possibility to detect antimicrobial producing fosmid clones. Main testing material was the native culture supernatant of the NRPS clones, but also the ethyl acetate extracts and the lyophilisates of culture supernatants were examined. The aim of analyzing extracts and lyophilisates was to concentrate possible antimicrobials, so that an antimicrobial effect might be easier to detect.

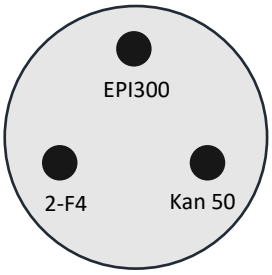
The lyophilisates of the NRPS clones were examined using the soft agar well-diffusion assay and *C. albicans* and *S. epidermidis* as test organisms. Ethanol and kanamycin were used as a positive control. While kanamycin led to growth inhibition of both test strains, ethanol was only able to inhibit the growth of *S. epidermidis* (data not shown). None of the NRPS fosmid clone lyophilisates inhibited the growth of *S. epidermidis* and *C. albicans*, but rather enhanced their growth around the wells (data not shown). This effect might be due to the enhanced concentration of nutrients in the lyophilisates.

While the concentrated supernatant extracts showed no growth inhibition against the tested bacterial and fungal strains as well, the native supernatant of clone 2-F4 exhibited growth inhibition of *S. epidermidis, S. aureus*, *B. subtilis* and the yeast *C. albicans*, but not against *M. luteus* and the other tested fungal strains (Fig. 13 A and B). This effect was technically reproducible, but not biologically. Therefore, measures were taken to control if the active supernatant really was from the fosmid-containing *E. coli* culture and not from an unwanted bacterial contaminant. Fosmid isolation, restriction analysis, sequencing as well as HPLC/MS analysis confirmed that the antimicrobial activity was originated from clone 2-F4 (data not shown).

Subsequently, a more detailed look was taken at the insert sequence of clone 2-F4 and its putative NRPS sequence, the results thereof are described in the following chapter.

A





B

	B. subtilis	*S. epidermidis*	*S. aureus*	*C. albicans*
Kanamycin 50 µg	2.6 ± 0.08 cm	3.6 ± 0.08 cm	3.1 ± 0,1 cm *	3.5 ± 0.08 cm
Clone 2-F4	1.77 ± 0,05 cm	2.57 ± 0.05 cm	2.05 ± 0.05 cm *	1.97 ± 0.05 cm

Figure 13: Antimicrobial effect of clone 2-F4. *(A) Agar well-diffusion assays. The scheme on the right shows the positions of the applied native supernatants of clone 2-F4 and the negative control EPI300, as well as the positive control kanamycin (Kan, 50 µg). Positive antimicrobial effects of 2-F4 were detected against S. epidermis, S. aureus, B. subtilis and C. albicans. (B) Diameter of the inhibition zones of Kan and clone 2-F4 as average of technical replicates; n = 3 (*n = 2).*

4.2.2 The 2-F4 synthetase shows similarity to antimicrobial synthesizing NRPSs

Blastx analysis of clone 2-F4 by Müller et al. (2015) revealed a 57.3% identity to the Syringomycin synthetase E of *Pseudomonas syringae*. Via primer walking (Fig. 14) two longer sequences (5173 bp, 2697 bp) were obtained and blasted again (blastx) against the non-redundant protein database.

The blast results revealed identities with non-ribosomal peptide synthetases *of Lysobacter gummosus* (WP_057943825.1, 60% identity) and *Pseudomonas syringae* (WP_046719117.1, 58% identity) for the 5173 bp sequence and identities with a hypothetical protein of *Andrepre-votia chitinilytica* (WP_051711486.1, 66% identity) and non-ribosomal peptide synthetases of various *Pseudomonas* species (e.g. WP_085656799.1) for the 2697 bp sequence.

Sequence identities were not only found with uncharacterized synthetases, but also with several synthetases responsible for the production of antimicrobials and toxins (Table 4).

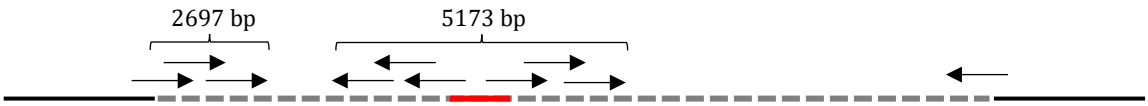


Figure 14: Primer walking of clone 2-F4.

vector (pCC2FOS) — insert sequence — primary NRPS sequence → sequences obtained via primer walking

Txo1 is a NRPS that together with Tox2 synthesizes Teixobactin, a new antibiotic so far without detectable resistances (Ling et al. 2015). Syringopeptins are a group of antimicrobial cyclic lipodepsipeptides and syringomycin is a member of a class of lipodepsinonapeptide molecules, which also shows antimicrobial effects. Both groups of peptides also exhibit phytotoxicity (Bender et al. 1999). NunD and ThaA are parts of the machineries that produce nunamycin (Michelsen et al. 2015) and thanamycin (Johnston et al. 2015), respectively – both are peptides with potent antifungal activity. Meanwhile, tyrocidine and gramicidin are both well-studied peptide antibiotics synthesized by linear non-ribosomal peptide synthetases (Felnagle et al. 2008)

Table 4: Identities of 2-F4 sequences with NRPSs playing a role in antimicrobial and toxin production. *Blastx analysis was performed against the non-redundant sequences database.*

Sequence	Accession No.	Description [source]	Query cover [%]	Identity [%]
5173 bp	AJF34463.1	Txo1 [*Eleftheria terrae*]	99	53
	KPY67055.1	Syringopeptin synthetase C [*Pseudomonas syringae pv. syringae*]	99	53
	AHL29289.1	NunD [*Pseudomonas fluorescens*]	99	52
	ALG65294.1	ThaA [*Pseudomonas fluorescens*]	99	52
	SDT16599.1	Syringomycin synthetase protein SyrE [*Pseudomonas asplenii*]	99	50
2697 bp	OFA03998.1	Tyrocidine synthase 3 [*Duganella sp. HH101*]	94	55
	AJF34463.1	Txo1 [*Eleftheria terrae*]	92	50
	ADO72770.1	LgrB-like linear gramicidin synthetase subunit B protein [*Stigmatella aurantiaca DW4/3-1*]	94	39

Taking a further look on the domain structure of the two 2-F4 sequences revealed that the adenylation domain present at the N-terminus of the shorter sequence has similarities with an acyl-CoA synthetase (reading frame +3, see Attachment, Fig. A1). The lipopeptide antibiotic daptomycin is known to initiate with an N-terminal lipid and the daptomycin NRPS itself initiates with a C domain rather than an A domain (Felnagle et al. 2008). Interestingly the blastx

result of the reading frame +1 revealed parts of a condensation domain at the C-terminal end of the 2697 bp sequence, following the initial acyl-CoA synthetase like A-domain (Attachment, Fig. A1). Alignment of the longer 2-F4 sequence with a partial daptomycin biosynthetic protein subunit of *Streptomyces filamentosus* (accession no. AAB96629.1) led to 39% identity with a query cover of 99%.

Taking the results of the partial sequence analysis of 2-F4 together, this metagenomic fosmid clone might harbor a NRPS system responsible for the synthesis of a lipopeptide-type antibiotic. Further sequence analysis of this clone is advised.

4.3 TLC and HPLC analysis for further product detection

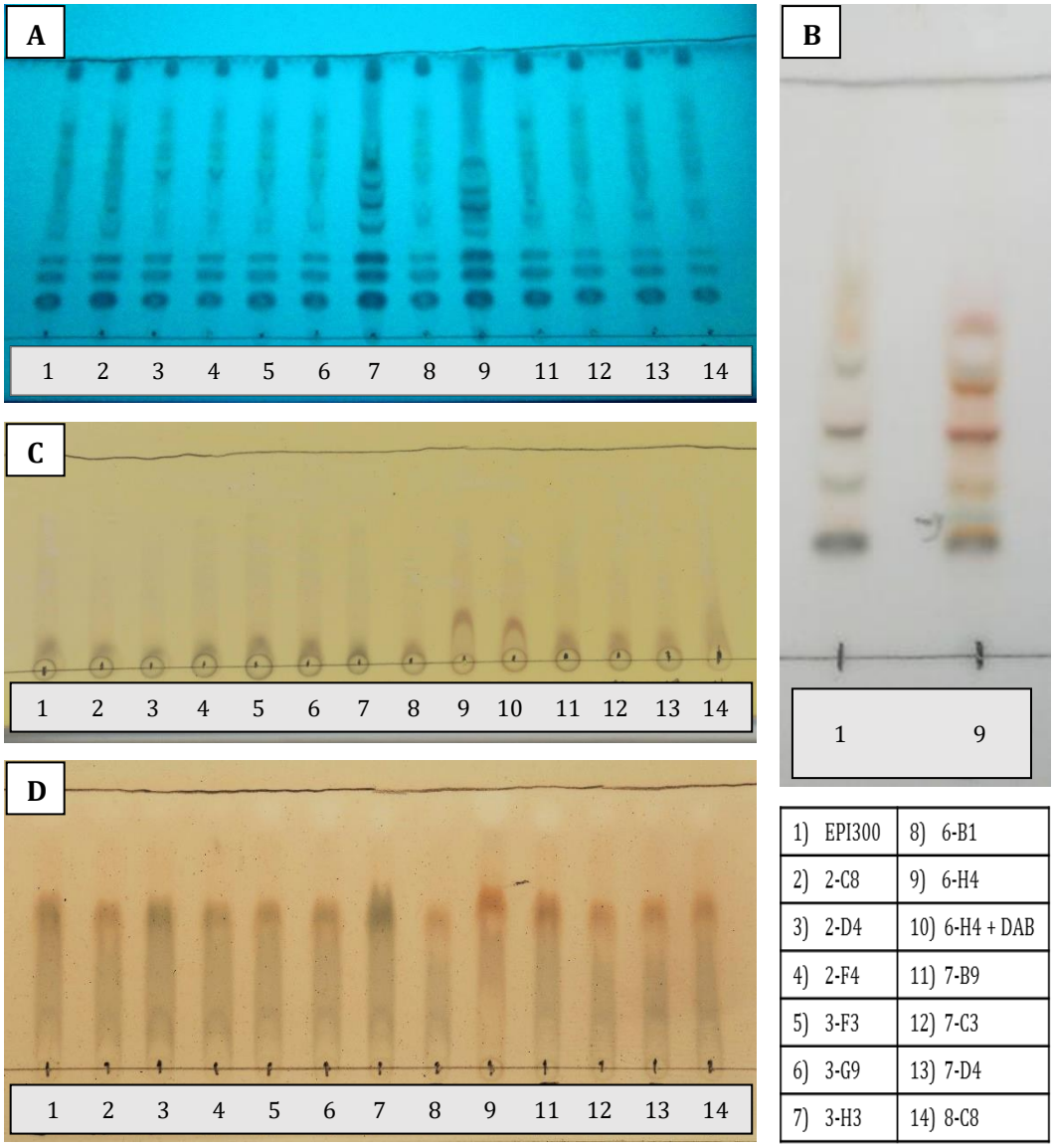
Thin-layer chromatography and high performance liquid chromatography were applied to detect other possible products of the NPRS fosmid clones of the moss metagenome, besides siderophores and antimicrobials. Analytical TLC and HPLC coupled with an UV detector were used to identify differences between the fosmid clones and the host strain EPI300. Afterwards preparative TLC and HPLC/MS were applied to determine possible unique products of the fosmid clones.

4.3.1 Clone 6-H4 showed differences on TLC plates and unique peaks in HPLC/UV

Like in chapter 3.9 mentioned, the clones were grown in MM9 – the same medium that was used for the liquid siderophore assay. This medium was chosen on one hand because it is a minimal medium, allowing lower extraction of medium components than with a rich medium, which gives a lower medium background in the analyses. On the other hand, due to the growth of the clones in this medium, *E. coli* would produce the siderophore enterobactin, which can be detected with TLC and HPLC/MS and served as a control.

After determining the solvent systems that worked best for the analysis of the fosmid clone extracts, analytical TLC with normal and reverse phase silica plates was performed. With all three separation setups mentioned in chapter 3.8, the differences between the separation patterns and colorings of clone 6-H4 and the negative control EPI300 were the most prominent (Fig. 15).

The TLC plates were analyzed using the Fe-stain, a stain usually used to identify catechol- and hydroxamate-type siderophores (Clark 2004). Several spots of clone 6-H4 displayed a reddish color, indicating that hydroxamic acids are present in the clone's supernatant extract. Meanwhile the spots of EPI300 were mainly grey, indicating catechol-type iron binding compounds.



1) EPI300	8) 6-B1
2) 2-C8	9) 6-H4
3) 2-D4	10) 6-H4 + DAB
4) 2-F4	11) 7-B9
5) 3-F3	12) 7-C3
6) 3-G9	13) 7-D4
7) 3-H3	14) 8-C8

Figure 15: TLC analysis of the NRPS-containing fosmid clones. *The table in the bottom right corner lists the positions of the applied NRPS clone supernatant extracts (DAB = diaminobutyric acid). (A, B) Separation pattern of the extracts on reverse phase silica plates using a mixture of acetonitrile:water:formic acid as solvent system and UV-detection (A) and Fe-staining (B) as visualization methods. (C, D) Normal phase silica plates using ethyl acetate (C) or ethyl acetate:acetic acid:water (D) as solvent system and Fe-staining for visualization.*

The second staining that was used for visualization was ninhydrin – a staining used to identify amino acids. With this staining, no differences were detected between the NRPS clones and the control EPI300 (data not shown).

Besides TLC analysis, HPLC coupled with UV-detection was performed to detect if the NRPS clones displayed unique peaks when compared with the negative control. Like with the TLC analyses, clone 6-H4 exhibited the most prominent differences when compared to EPI300 – 6-H4 not only had peaks with enhanced peak height, but also displayed some seemingly unique peaks (Fig. 16).

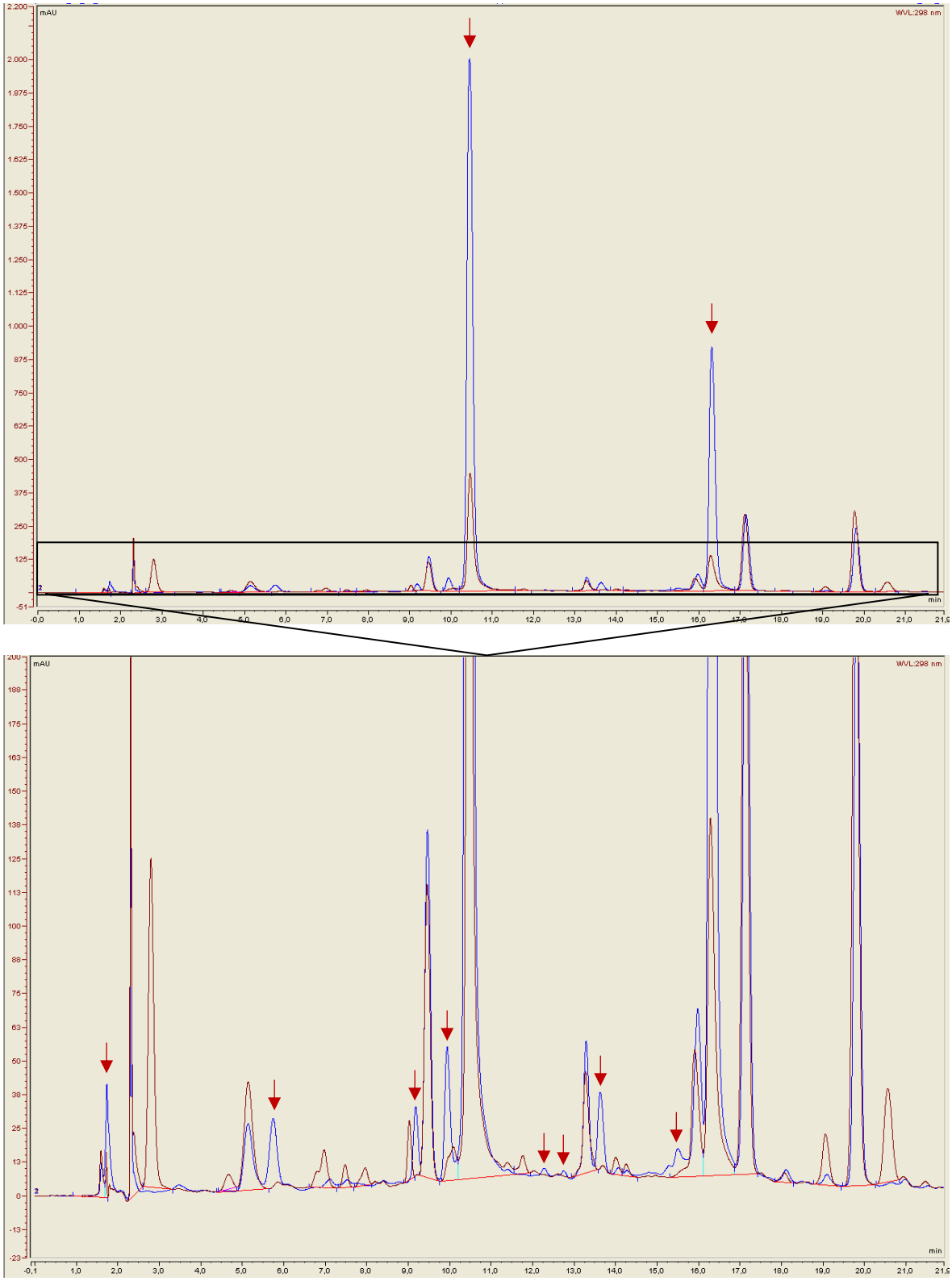


Figure 16: Overlay of the HPLC chromatograms of clone 6-H4 and EPI300. *The red chromatogram belongs to the negative control EPI300, while the blue one is of clone 6-H4. The red arrows indicate enhanced peak height or unique peaks of 6-H4. UV measurements were performed at 298 nm. x-axis: time (in min), y-axis: milli absorption units (mAU).*

The chromatograms of the other NRPS clones showed only peaks with enhanced height, but lacking unique peaks when compared with EPI300 (data not shown). Since clone 6-H4 showed

the most prominent differences on TLC plates as well as with HPLC/UV when compared with the negative control EPI300, this clone was chosen for further examination via preparative TLC and subsequent HPLC/MS analysis.

4.3.2 Preparative TLC and HPLC/MS revealed unique products of clone 6-H4

Plates and solvent systems for preparative TLC were the same as for analytical TLC (see chapter 3.8). Upon separation of the extracts, spots of EPI300 and clone 6-H4 visible under UV light (Fig. 17) and spots visible with Fe-stain were marked, their retention values were calculated (Table 5) and then the spots were extracted from the TLC plates and analyzed with HPLC/MS (see chapter 3.9.2 for the settings).

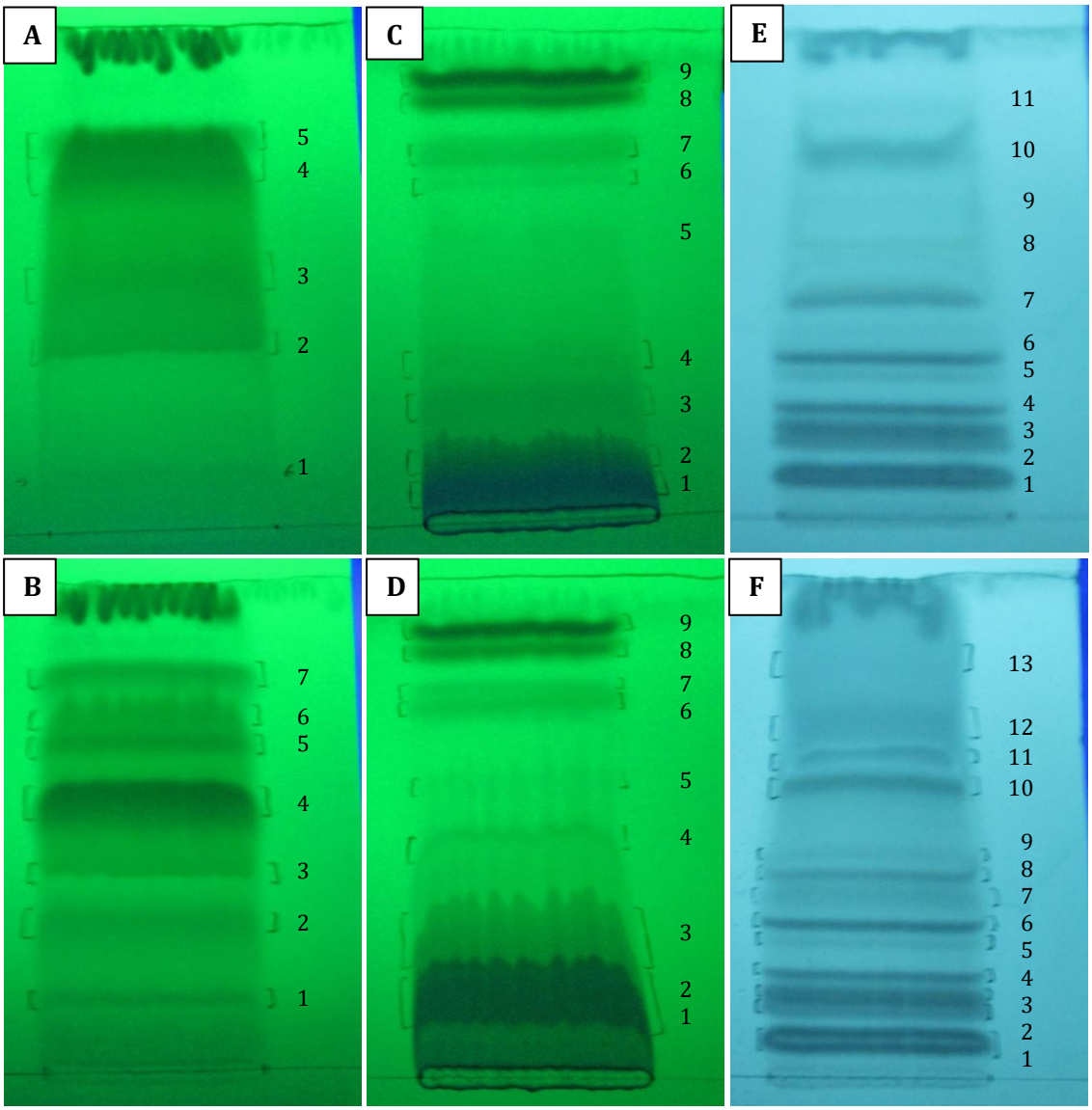


Figure 17: Preparative TLC of EPI300 and clone 6-H4. *The upper row (A, C, E) shows the separation patterns of EPI300, the lower row (B, D, F) the separation patterns of clone 6-H4. (A, B) Normal phase silica plates with ethyl acetate:acetic acid:water as solvent system. (C, D) Normal phase plates with ethyl acetate as solvent system. (E, F) Reverse phase silica plates with acetonitrile:water:formic acid as solvent system. Spots were marked under UV light.*

Fe-staining revealed that some spots which seemed to be a single analyte under UV light had differences in coloring and therefore were treated as separate analytes. As can be seen in Table 5, many spots of clones 6-H4 and EPI300 had similar retention values. Spots of clone 6-H4 with unique retention values are underlined.

Table 5: Retention values (R_F values) of EPI300 and clone 6-H4. *Retention values were determined as follows: distance the spot traveled divided by the distance the mobile phase (solvent) traveled. Unique R_F values of clone 6-H4 are underlined and colored.*

Spot Number (Fig. 17)	Ethyl acetate:acetic acid:water (normal phase silica)		Ethyl acetate (normal phase silica)		Acetonitrile:water:formic acid (reverse phase silica)	
	EPI300 (A)	6-H4 (B)	EPI300 (C)	6-H4 (D)	EPI300 (E)	6-H4 (F)
1	0.115	0.123	0.063	0.119	0.074	0.051
2	0.282	0.272	0.127	0.202	0.136	0.114
3	0.500	0.370	0.228	0.333	0.173	0.152
4	0.744	0.519	0.329	0.452	0.210	0.190
5	0.795	0.642	0.570	0.571	0.272	0.253
6		0.704	0.684	0.714	0.309	0.278
7		0.790	0.747	0.762	0.432	0.329
8			0.848	0.833	0.531	0.380
9			0.911	0.893	0.642	0.418
10					0.741	0.557
11					0.827	0.608
12						0.684
13						0.823

Each of those spots with determined retention value was extracted from the TLC plates. The spots A4 and A5 were extracted as one spot, since separation was not possible. Upon extraction, the various spots were analyzed via HPLC coupled to mass spectrometry (MS).

Figure 18 shows the chromatograms of clone 6-H4 spots extracted from the normal phase silica plate with ethyl acetate:acetic acid:water as solvent system (Spots B1-B6), measured with positive ion HESI detection, since those chromatograms displayed the most unique peaks in comparison with the negative control EPI300.

The chromatograms of the other extracted spots, as well as of the whole extracts of EPI300 and 6-H4 can be found in the Attachment, Fig. A2-A7.

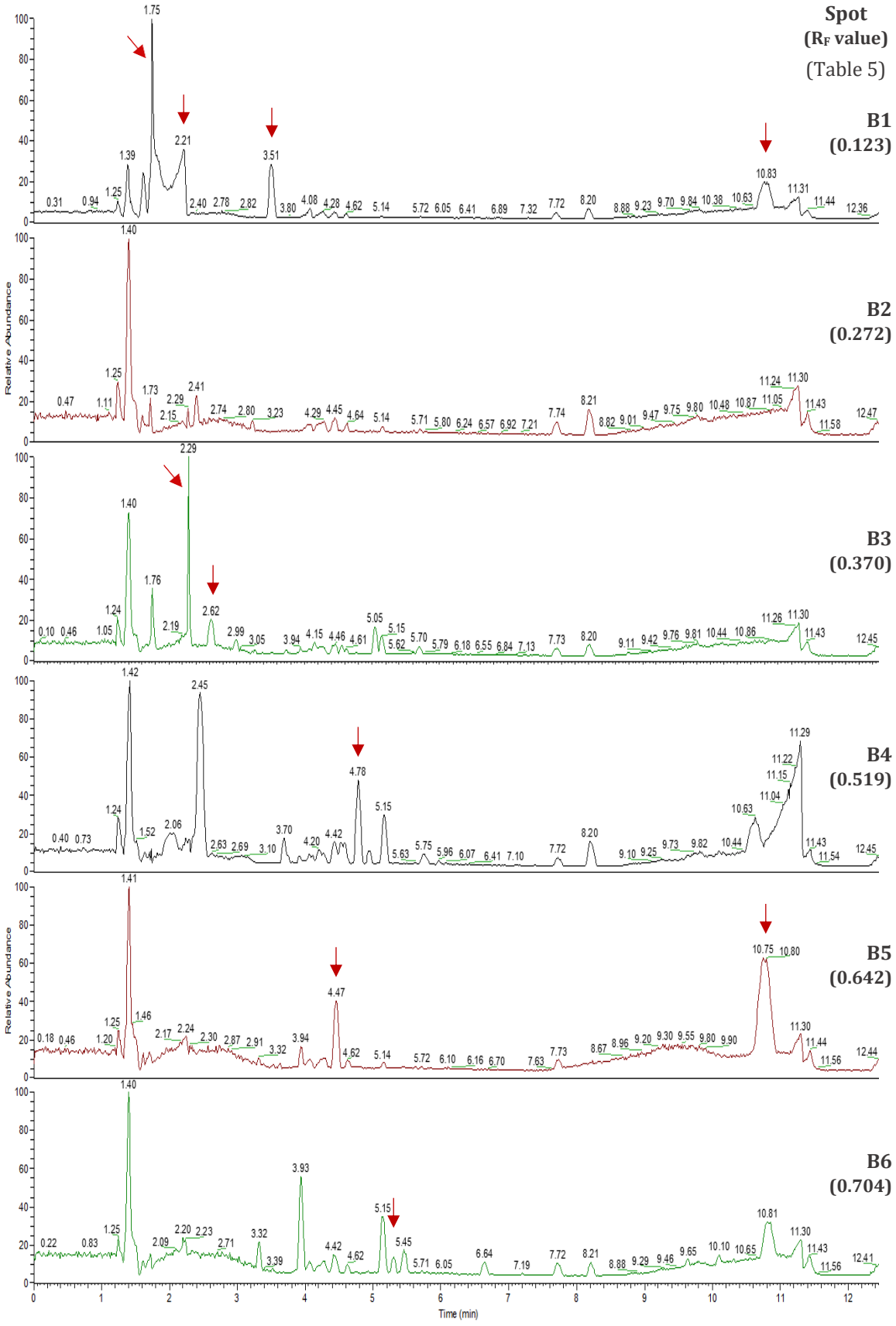


Figure 18: HPLC chromatograms (positive mode) of clone 6-H4 extracted spots. *Spots were extracted from normal phase silica plates with ethyl acetate:acetic acid:water as solvent system. The spot numbers are shown on the right side, the corresponding R_F values can be looked up in Table 5. Samples were analyzed with positive ion HESI detection. The red arrows indicate peaks unique or strongly upregulated in comparison to EPI300. X-axis: time (min), y-axis: relative abundance.*

With the help of Compound Discoverer 2.0 software, the mass spectra of 6-H4 products were then manually compared with the mass spectra of the suggested products. This led to the conclusive identification of some products of clone 6-H4, but many masses and mass spectra could not be assigned to known compounds, mainly because many predicted chemical formulas were assigned to more than one compound and manual comparison with the spectra of all possible substances was not feasible. Table 6 lists the conclusively identified compounds produced by clone 6-H4, as well as one partially identified product. Table 7 lists the predicted chemical formulas of some of the unassigned or novel substances produced by this NRPS clone.

Table 6: Identified substances produced by clone 6-H4. *Substances were identified by manual comparison of the mass spectra with the spectra of suggested substances (suggestions by Compound Discoverer 2.0, ChemSpider, KEGG).*

Conclusively identified substances					
Component	**Spot No. (R_F Value)**	**Retention Time (min)**	**Molecular Formula**	**Monoisotopic Mass**	**Details**
L-pyroglutamic acid	B3 (0.370)	2.29 (Fig. 18)	$C_5H_7NO_3$	129.0429 Da	mzCloud match 86%, SFit 63%, Delta Mass 2.93 ppm
Hydroxy-isophthalic acid	B1 (0.123)	3.51 (Fig. 18)	$C_8H_6O_5$	182.02153 Da	SFit: 56%, Delta Mass 1.81 ppm it cannot be differentiated between 2-, 3-, 4-, and 5-hydroxy-isophthalic acid
Phenylalanine	F7 (0.608)	5.63/5.79 (Fig. A6)	$C_9H_{11}NO_2$	165.07935 Da	Consistent with fragment list; mzCloud match 71%
Partial identified substances					
Leucine fragments	B4 (0.519)	4.78 (Fig. 18)	$C_6H_{13}NO_2$	131.09499 Da	Part of the mass 242.12713 Da, while the leucine fragments could be clearly identified but the parent ion could not

L-Pyroglutamic acid, or 5-oxoproline, is the only slightly studied cyclic lactam of glutamic acid or glutamine (Kumar & Bachhawat 2012). Hydroxyisophthalic acids have structural similarities with salicylic acid. 4-hydroxyisophthalic acid, as well as 2-hydroxyisophthalic acid are by-products in the production process of salicylic acid (Collier & Chesher 1956). Phenylalanine and leucine are both proteinogenic amino acids (Ambrogelly et al. 2007).

Table 7: Unassigned or novel substances produced by 6-H4.

Unassigned or novel substances			
Spot No./ Retention Time/ Figure No.	**Monoiso-topic Mass (Da)**	**Predicted Formula**	**Details**
B5/4.85 min/Fig. A3	130.06154	no formula	ChemSpider suggestions ranging from C_4 to C_8
B1/2.21 min/Fig. 18	143.05824	$C_6H_9NO_3$	SFit: 65%, Delta Mass: 2,44 ppm
B2/3.27 min/Fig. A3	146.03678	$C_9H_6O_2$	SFit 73%, Delta Mass 0.95 ppm
B1/1.75 min/Fig. 18	148.01982	$C_5H_8O_3S$	SFit 54%, Delta Mass 2.75; similarities with the spectrum of Tetra-hydro-4H-thiopyran-4-one, 1.1-dioxide
D3/4.04 min/Fig. A5	182.05685	no formula	ChemSpider suggestions ranging from C_4 to C_7
B4/4.77 min/Fig. A3	252.06229	$C_{12}H_{12}O_6$	SFit 85%
B6/5.31 min/Fig. 18	256.14285	C12H20N2O4	SFit 54%
B1/3.55 min/Fig. A3	270.03773	$C_{11}H_{10}O_8$	SFit 86%, Delta Mass 0.59 ppm
D9/7.41 min/Fig. A4	274.22616	$C_{14}H_{30}N_2O_3$	SFit 52%, Delta Mass 1.88 ppm
B2/4.99 min/Fig. A3	286.04985	$C_{16}H_6N_4O_2$	SFit 32% some similarities with Hexadecanoic acid, 15-hydroxy-, methyl ester
F2/8.25 min/Fig. A6	333.17899	$C_{10}H_{24}N_9O_2P$	SFit 93%
D6/7.30 min/Fig. A4	342.12977	$C_{13}H_{22}N_6OS_2$	SFit 53%
D9/10.14 min/Fig. A4	374.19217	$C_{22}H_{30}O_3S$	only few fragments: 17.002: OH; 32.075: methyl ester; 87.04541: methyl ester or ethyl ester; 55 ChemSpider suggestions: e.g. 7α-thiospironolactone
B1, B5/10.83 min/ Fig. 18	514.35155	$C_{23}H_{47}N_8O_3P$	SFit 48%
	531.37852	C30H45N9	SFit 82% ChemSpider: $C_{29}H_{49}N_5O_4$ (59%)
F7/5.63 min/Fig. A6	552.31468	$C_{21}H_{45}N_8O_7P$	SFit 54%, ChemSpider suggestions from C_{25} bis C_{34} – e.g. 2-{(E)-[2-{Bis[3-(dimethylamino)propyl]amino}-1-phenyl-4(1H)-quinolinylidene]methyl}-3-methyl-1,3-benzothiazol-3-ium
F7/5.60 min/Fig. A6	566.33044	$C_{32}H_{48}N_4OP_2$ $C_{28}H_{46}N_4O_8$	both formulas with SFit 70%
F7/5.79 min/Fig. A6	667.32085	$C_{33}H_{49}NO_{13}$ $C_{38}H_{47}N_5O_2P_2$	both with SFit 84% ChemSpider suggestions of amino acid chains (threonine, ornithine)
D6/10.38 min/ Fig. A4	884.36280	$C_{26}H_{66}N_{10}O_{15}P_2S_2$	SFit 42%, ChemSpider suggestions ranging from C_{34} to C_{51}

Further analysis of the unassigned substances is necessary, to see if they are already known or if they are completely novel substances produced by this NRPS-harboring fosmid clone.

4.4 Sequence analysis of fosmid clone 6-H4

Besides the activity-based screenings and the TLC/HPLC assays, analysis of one selected NRPS fosmid clone DNA-sequence was performed to complete the comprehensive analysis of NRPSs from the moss metagenome. Clone 6-H4 was chosen for whole-insert sequencing and sequence analysis, due to preliminary results from a liquid siderophore assay performed in a previous study. The main goal of this analysis was to retrieve the whole nucleotide sequence of the non-ribosomal peptide synthetase of clone 6-H4 for subsequent heterologous expression and further characterization of the synthetase gene. In addition, the insert sequence of that clone was analyzed for other putative open reading frames (ORFs) and gene clusters.

4.4.1 Clone 6-H4 harbors 84 putative open reading frames coding for various enzymes

Shotgun-sequencing and subsequent primer walking revealed that clone 6-H4 harbors a metagenomic insert sequence of 44.7 kilobases (kb). Using NCBI ORFfinder (settings: minimal ORF length of 300 nucleotides, standard genetic code and "ATG" only as start codon) 84 putative open reading frames (ORFs) were discovered (Attachment, Fig. A8). The amino acid sequences of all 84 ORFs were subsequently analyzed with Blast2Go Pro, using NCBI blast and Cloudblast functions to find homologous sequences and then proceed with gene ontology (GO) mapping and annotation of the sequences. Figure 19A gives an overview of the analysis process of the 84 putative open reading frames of clone 6-H4 using Blast2Go Pro.

Sequence analysis using the Blast2Go Pro NCBI blast and the described settings (chapter 3.10.1) led to blastp results for 62 of the 84 sequences, and of those 62 sequences 20 could be mapped. In the end, using the Blast2Go Pro default settings for annotation (cutoff of 55, GO weight of 5, E-value hit filter of 1.0E-6), 18 sequences could be annotated. A table of those blastp results can be found in the Attachment, Table A3.

Since many of those blastp results had a low score or showed low complexity upon manual inspection, an additional Blast search was performed, using the Cloudblast option of Blast2Go Pro. Cloudblast is a function that combines standard NCBI blast with searches directly from the Blast2Go Pro computing cloud. For the Cloudblast analysis (blastp) of the 84 putative ORFs of clone 6-H4 against the non-redundant database, the default settings were retained (chapter 3.10.1). A lower E-value threshold and the enabled low-complexity filter led to more stringent, but overall to less blastp results compared to the NCBI blast search – only 45 of 84 ORFs were successfully blasted (Figure 19A). Upon merging the GO terms identified by the InterPro Scan

with the GO terms detected by Blast2Go Pro Mapping, annotation was performed and subsequently ANNEX (Annotation Expander) was applied, which uses a second gene ontology layer to suggest new biological process annotations based on the existing molecular function annotations (Fig. 19B). To give a broad classification of the molecular functions of the genes, as well as the biological processes they are involved, GO-Slim was carried out.

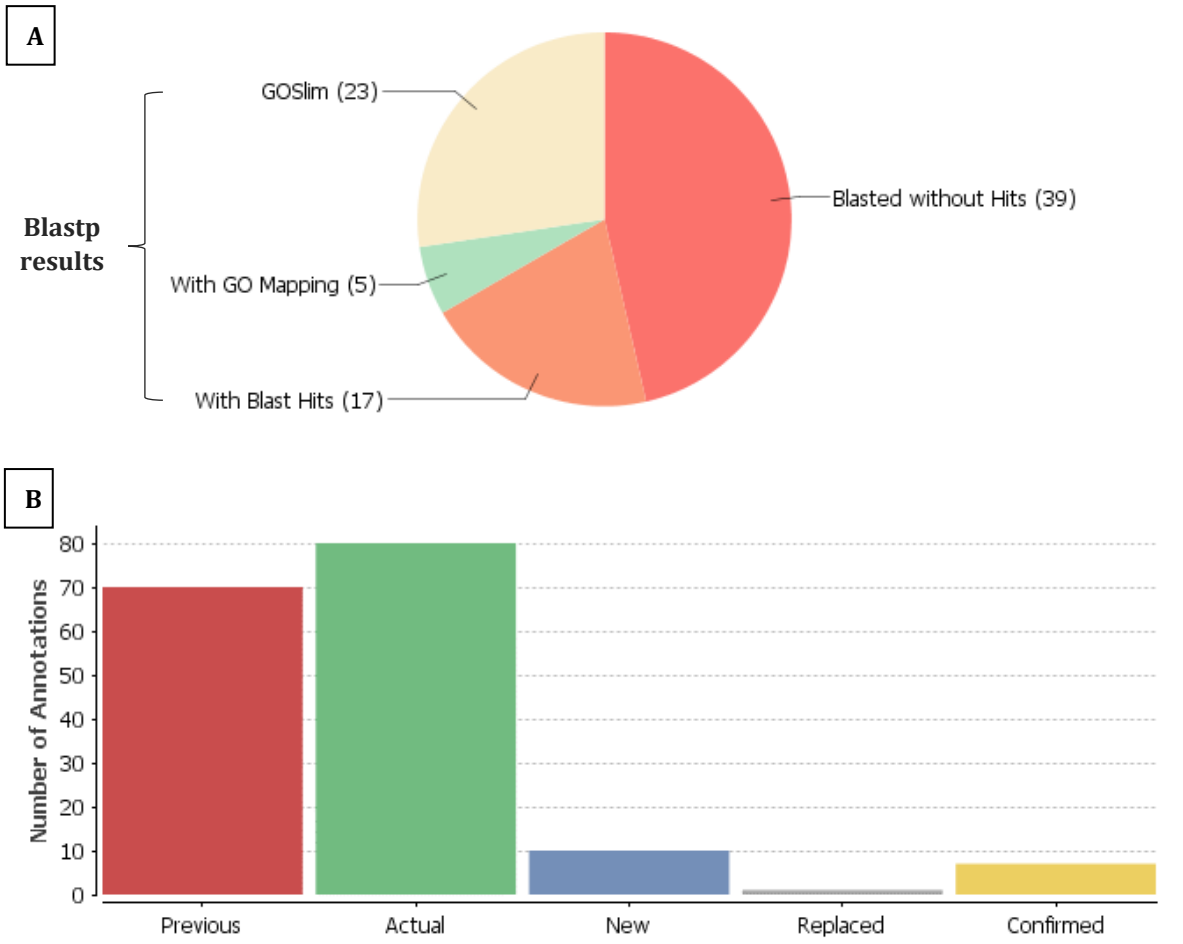


Figure 19: Blast2Go Pro analysis process. *The 84 putative open reading frames were analyzed using Blast2Go Pro. (A) Cloudblast was performed and led to blastp results for 45 sequences, while 39 were without blast hits. Of those 45 sequences with blast results, 28 could be mapped (merged with InterPro Scan mapping) and a total of 23 could be annotated (Go-Slim annotation). (B) Before performing Go-Slim, ANNEX was used to suggest new biological process annotations based on the molecular function annotations of the 6-H4 putative genes. 10 new GO terms were added to the already existing ones.*

To visualize the summarized molecular function and biological process annotations of the 6-H4 genes after GO-Slim, GO graphs were made with the 'Combined GO Graph' function of Blast2Go Pro and the results are summarized in Fig. 20. The applied settings (chapter 3.10.1) allowed categorizing the general functions of the genes encoded by the fosmid clone 6-H4.

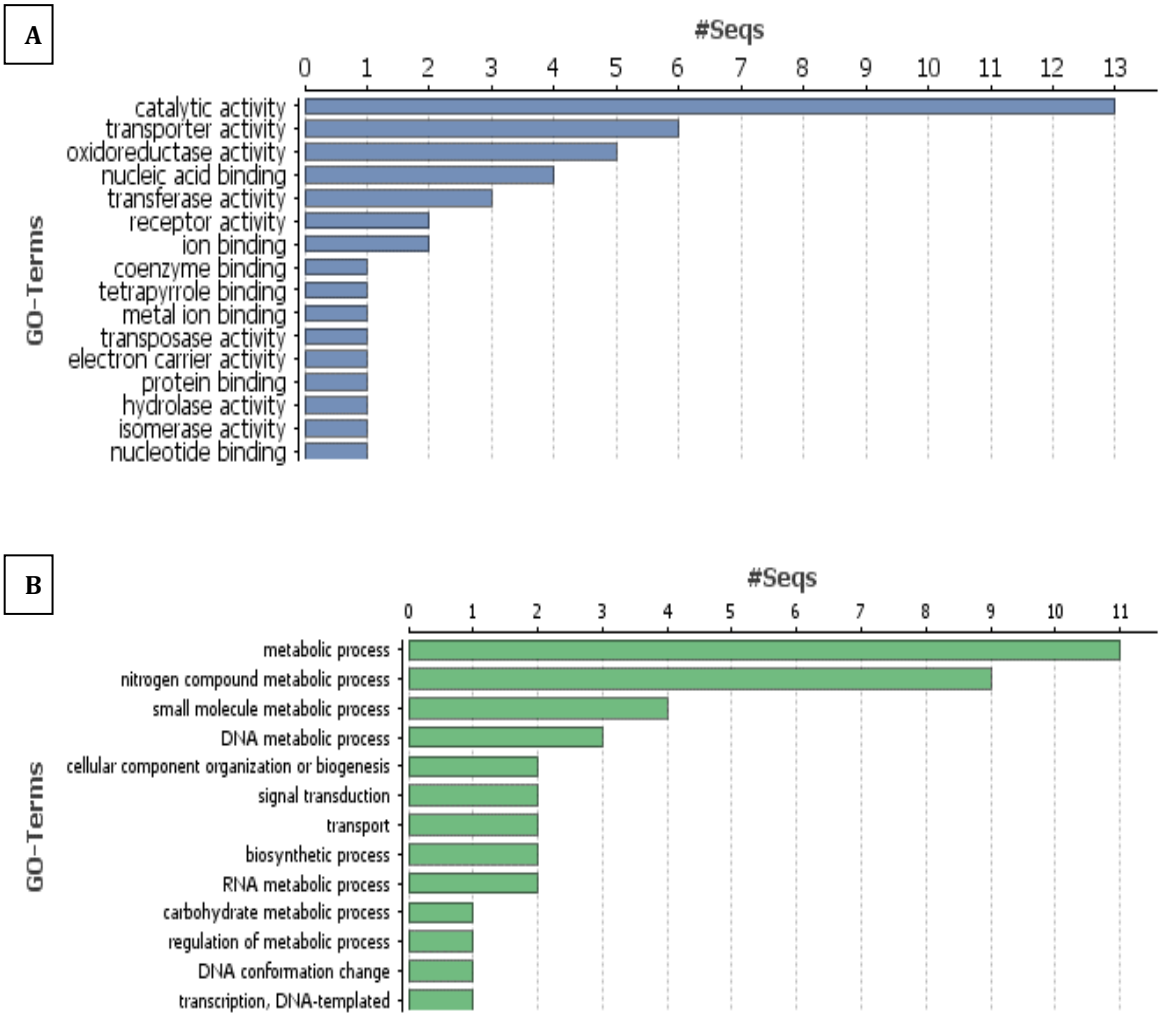


Figure 20: General molecular function and biological process annotations of 6-H4 genes. *Graphs were generated using the Combined GO Graph function of Blast2Go Pro, setting the GO level to 1 and filtering intermediate GO terms. (A) Distribution of the molecular functions of the genes. (B) Distribution of the biological processes they could be involved.*

Major molecular functions were categorized as catalytic activity (13 genes), transport activity (6 genes) and oxidoreductase activity (5 genes). In terms of biological processes to which those genes belong to, the most representatives are: metabolic processes (11 genes), nitrogen compound metabolic processes (9 genes), and small molecule metabolic processes (4 genes).

In addition to the analysis and annotation with Blast2Go Pro, annotation with WebMGA was performed. While Blast2Go Pro uses GO annotation, for the analysis with WebMGA the COG (Clusters of Orthologous Groups) function annotation for prokaryotic proteins was employed. In comparison with Blast2Go Pro, WebMGA was able to annotate 25 of the 84 genes of 6-H4. Some genes were assigned to multiple COG classes, while other were only given a general prediction. Figure 21 gives an overview of the COG annotation results of clone 6-H4 putative genes. A table showing the WebMGA RSPBLAST results against the COG database can be found in the Attachment (Table A4)

Main COG functions were “cell wall/membrane/envelope biogenesis” (6 genes), “amino acid transport and metabolism” (4 genes) and “intercellular trafficking, secretion, and vesicular transport” as well as “inorganic ion transport and metabolism” (3 genes each). A large group of genes was categorized as “general function prediction only” (5 genes), as they could not be assigned to a specific COG class.

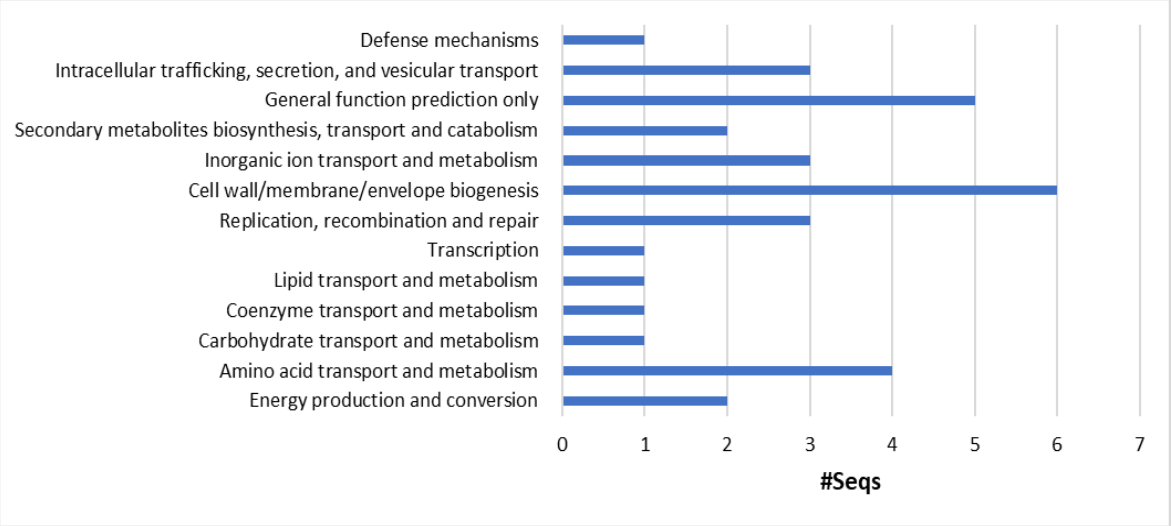


Figure 21: Annotation of 6-H4 genes using WebMGA/COG database.

Comparing the “biological process” GO and the COG annotations (see Table 8), overlapping annotations existed for some of the 6-H4 genes. ORFs 9, 48 and 79 were assigned to DNA metabolic processes (replication, recombination, repair), while ORFs 66 and 74 were assigned to transport, ORF 83 to transcription and ORF 21 to carbohydrate metabolism. However, other comparable annotation categories, like “biosynthetic process” (GO) and “secondary metabolites biosynthesis, transport and catabolism” (COG) were assigned to completely different ORFs.

The nonribosomal peptide synthetase-like enzyme encoded on the 6-H4 insert (ORF 27) was also successfully blasted and annotated. The putative synthetase was assigned to “metabolic process” (GO, biological process) and “catalytic activity” (GO, molecular Function, Attachment Table A5), as well as to “secondary metabolites biosynthesis, transport and catabolism” (COG).

Taking together, the overall result of the 6-H4 insert sequence analysis shows that the *Sphagnum* moss metagenome is a rich source for novel enzymes, since only a 44.7 kb fragment of metagenomic DNA harbors 84 putative ORFs coding for enzymes with various activities (catalytic activity, oxidoreductase activity, isomerase/transferase activity etc., see Attachment, Tables A3 and A4) and diverse biological functions.

Table 8: GO and COG annotations of the 6-H4 genes. *The table shows the two different annotations and their assigned ORFs. Groups or classes that were similar for both annotations are highlighted.*

Blast2Go Pro (GO annotation)		**WebMGA (COG annotation)**	
Function	**ORFs**	**Function**	**ORFs**
Metabolic process	2, 9, 13, 14, 21, 27, 39, 41, 48, 79, 83	Energy production and conversion	8, 10
Nitrogen compound metabolic process	2, 9, 13, 14, 39, 41, 48, 79, 83	Amino acid transport and metabolism	2, 8, 13, 14
Small molecule metabolic process	2, 13, 14, 31	Coenzyme transport and metabolism	39
DNA metabolic process	9, 48, 79	Replication, recombination and repair	9, 48, 79
Cellular component organization or biogenesis	9, 41	Cell wall/membrane/envelope biogenesis	21, 22, 23, 51, 81, 82
Signal transduction	66, 74	Intracellular trafficking, secretion, and vesicular transport	23, 51, 81
Transport	66, 74	Inorganic ion transport and metabolism	35, 66, 74
Biosynthetic process	39, 83	Secondary metabolites biosynthesis, transport and catabolism	27, 68
Transcription, DNA-templated	83	Transcription	83
Carbohydrate metabolic process	21	Carbohydrate transport and metabolism	21
Regulation of metabolic process	83	Defense mechanisms	67
DNA conformation change	9	Lipid transport and metabolism	68
RNA metabolic process	41, 83	General function prediction only	1, 27, 41, 57, 68

4.4.2. The 6-H4 NRPS (ORF 27) has an unusual structure and is part of a gene cluster

To further analyze the 6-H4 insert sequence for possible biosynthetic gene clusters, cluster analysis with antiSMASH 3.0 was performed, applying the sites default settings for bacterial sequences. A gene cluster spanning about half of the 44.7 kb insert was identified (Fig. 22), consisting of several transport related genes (ORFs 22, 23, 35, 66), as well as two genes assigned to biosynthesis (ORFs 10 and 27) and four other genes (ORFs 8, 9, 25, 39). Blast results of those ORFs can be found in the Attachment (Table A3).

As mentioned in the previous chapter, ORF 27 encodes for the putative NRPS of clone 6-H4. Taking a further look at the domain architecture, the 6-H4 synthetase has an adenylation domain (A-domain) and a PCP domain typical for nonribosomal peptide synthetases, but it lacks a condensation domain for peptide bond formation, as well as a thioesterase domain for peptide release. Instead, this putative peptide synthetase has an NRPS terminal domain of unknown function (Fig. 22). The described domain architecture of the 6-H4 synthetase is quite unusual when compared to the standard architecture of several NPRSs, like those described in chapter 1.4.2.

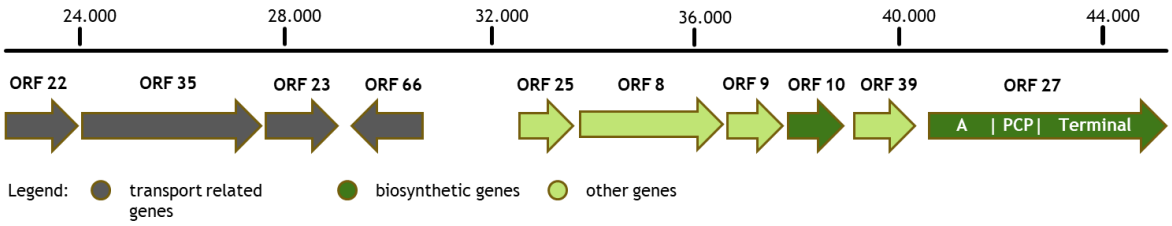


Figure 22: Cluster analysis of the 6-H4 insert. *A gene cluster, spanning about 22 kb was detected with antiSMASH 3.0. The biological functions are related to transport (gray), biosynthesis (dark green) and unknown functions (light green). The domains of the putative peptide synthetase (ORF 27), are shown in the graph: adenylation domain (A), peptide carrier protein/phosphopantetheine attachment side (PCP) and terminal domain of unknown function (Terminal).*

The neighboring genes of the NRPS are possibly associated to transport or yet unknown functions (Fig. 22).

Using NRPSpredictor2, possible substrate specificity of the adenylation domain of the synthetase for hydrophobic-aliphatic amino acids was predicted, with 2,4-diaminobutyric acid as the nearest neighbor (score of 60%). The by NRPSpredictor2 extracted specificity-conferring code of the A-domain is "DIECNGTVTK". With the general domain architecture and possible substrate specificity at hand, cloning and subsequent expression of the 6-H4 unusual peptide synthetase was performed as described in the following chapter.

4.4.3. Heterologous expression of the 6-H4 synthetase led to insoluble protein

For heterologous expression of the 6-H4 synthetase, BAP1 was used as host strain. This *E. coli* strain was specifically engineered for NRPS expression, as it co-expresses the *Bacillus subtilis* phosphopantetheinyl transferase *sfp* to assure the phosphopantetheinylation – a step necessary to convert the PCP domains from their inactive apo form to their active holo form (Pfeifer et al. 2001). The expression vector was generated via PLICing. With the PTO primers and PCR cycling conditions described in chapter 3.10.2, the correct PCR product of the pET28a(+) vector backbone was obtained. The PCR reaction for the 6-H4 synthetase gene was more unspecific, leading to several additional PCR products besides of the correct one (Fig. 23A). Subsequently, gradient PCR was performed with the 6-H4 synthetase gene in an attempt to reduce the generation of unspecific products. In the end, the correct PCR product was excised from the agarose gel and cleaned up (Fig. 23B). The ligation via PLICing was performed as described in chapter 3.10.2 and after transformation of the 6-H4_Synth expression vector in BAP1 several clones were picked for analysis by restriction. Out of all analyzed clones, only one exhibited the correct restriction pattern (Fig 23C). After the correct "integration" of the synthetase gene into the pET28a(+) vector was confirmed via sanger sequencing (data not shown), the positive clone was named BAP1 6-H4_Synth.

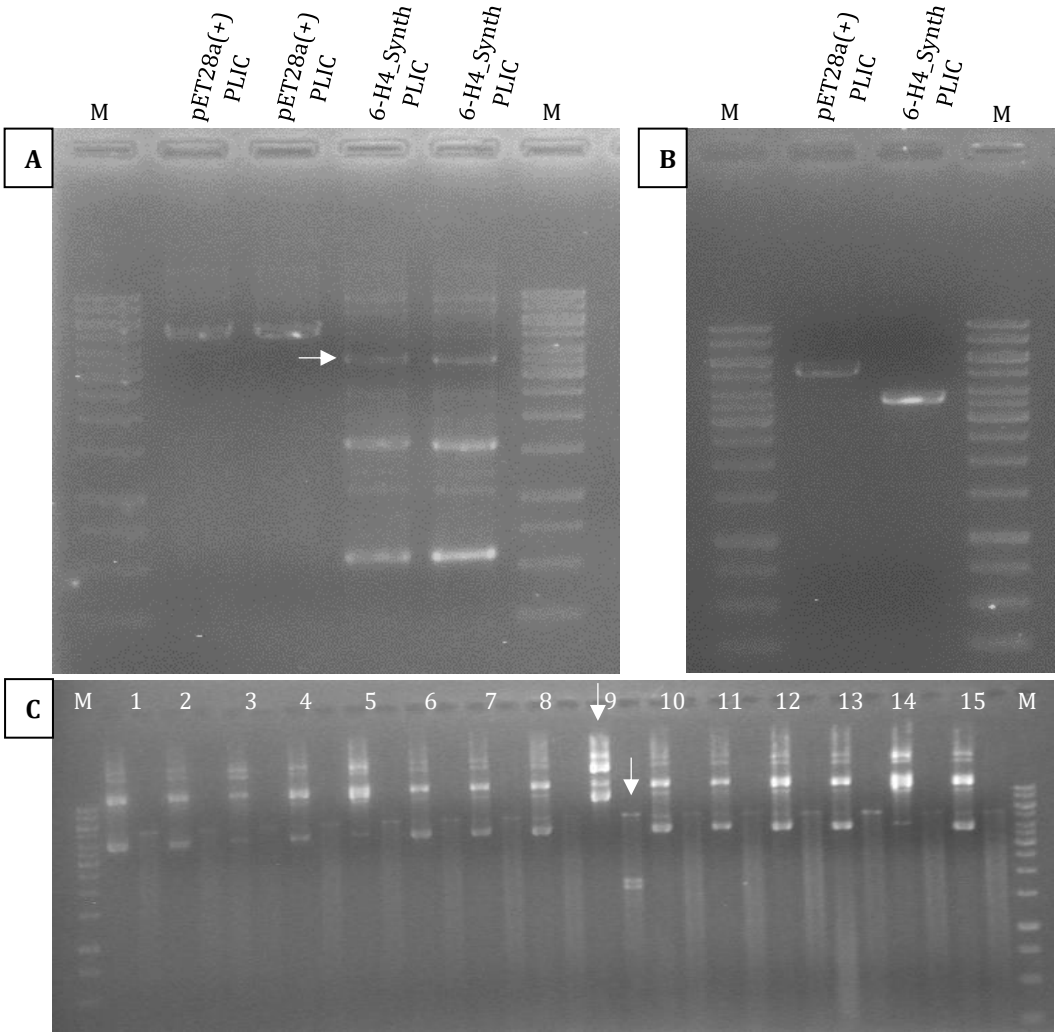


Figure 23: Generation of 6-H4_Synth expression vector. *(A) PCR products of 6-H4 and pET28a(+) after cycling reactions with the corresponding PTO primers. The white arrow indicates the correct product of the 6-H4 synthetase gene. (B) pET28a(+) and 6-H4 synthetase gene after clean-up. (C) Restriction analysis of vectors isolated from 15 BAP1 transformants. Uncut vectors are followed by the cut vectors. Vector no. 9 with the correct restriction pattern is indicated by the white arrows. M = GeneRuler™ 1 kb ladder.*

Following the successful generation of *E. coli* BAP1 6-H4_Synth, heterologous expression of the synthetase genes was tested as described in chapter 3.10.3. The weight of the synthetase was calculated with CLC Main Workbench and was expected to be 131 kDa. All tested conditions led to expression of the synthetase as an insoluble protein, since a protein band of the estimated size was only detected in the cell pellet fraction (whole cells, including insoluble membrane bound proteins, inclusion bodies, and soluble proteins) but not in the soluble cell free lysate fraction. Figure 24 shows the SDS PAGE results of the expression cultures incubated at 30°C, with different inducer concentrations and induction times. The best expression result in the insoluble fraction was achieved with an IPTG concentration of 0.4 and 1 mM and induction at an OD of 1 after 4 h of incubation. After 20 h the protein band is not detectable anymore.

Heterologous expression of foreign proteins in *E. coli* often leads to insoluble protein. Membrane proteins are especially prone to aggregate due to their hydrophobic transmembrane segments. Therefore, the 6-H4 synthetase gene was further analyzed for transmembrane domains, to be able to optimize the heterologous expression later on, should the synthetase prove to be a membrane protein.

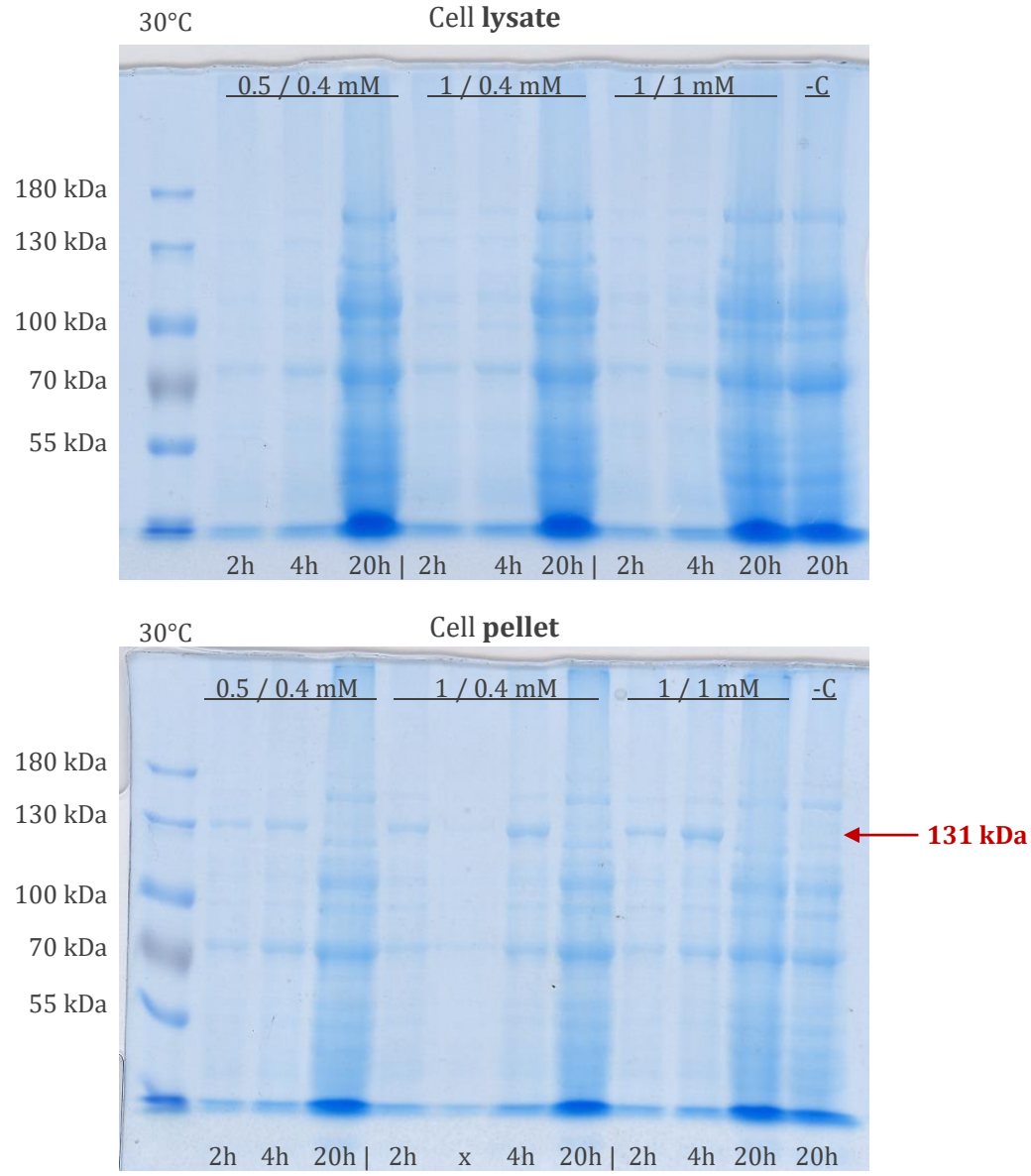


Figure 24: SDS PAGE analysis of the heterologously expressed 6-H4 synthetase. *Cultures were incubated at 37 °C until an optical density of 0.5 or 1 was reached. Expression was then induced with 0.4 or 1 mM IPTG and cultures were further incubated at 30 °C. Samples were collected after 2 h, 4 h and 20 h upon induction and SDS PAGE analysis was performed. As negative control (-C) a BAP1 clone harboring an empty pET28a(+) vector, induced with 1 mM IPTG at an optical density of 1, was used. The expected weight of the 6-H4 synthetase was 131 kDa.*

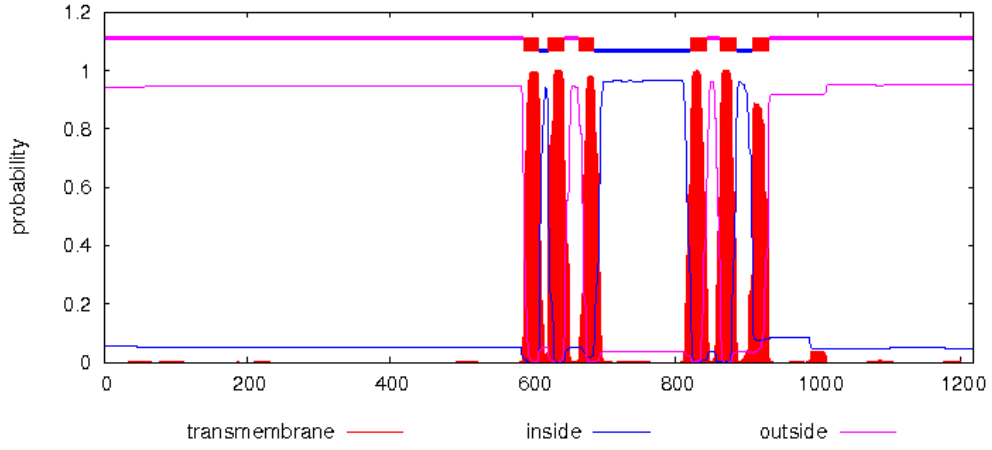
4.4.4. The 6-H4 synthetase is a membrane-related protein

Different programs were used to analyze the 6-H4 synthetase gene for transmembrane domains, as mentioned in chapter 3.10.1. While SOSUI detected five transmembrane helices (four primary, one secondary, Fig. 25A), THMM predicted six (Fig. 25B) and TmPred predicted 29, but only five of the predicted transmembrane helices had a high score over 1000.

A

No.	N terminal	transmembrane region	C terminal	type	length
1	588	VAAQTIALVPIFATAGIQWIFPY	610	PRIMARY	23
2	628	IAGAAFVVMPPLVLLLSIAVKWI	650	PRIMARY	23
3	676	FLAIIPIQFLAGTPIMRVYYRLL	698	SECONDARY	23
4	820	VNLGLLVAAPLLPLAAVIPIAPG	842	PRIMARY	23
5	861	LSPVLALVYVVLMCLLTVVVKWG	883	PRIMARY	23

B



C

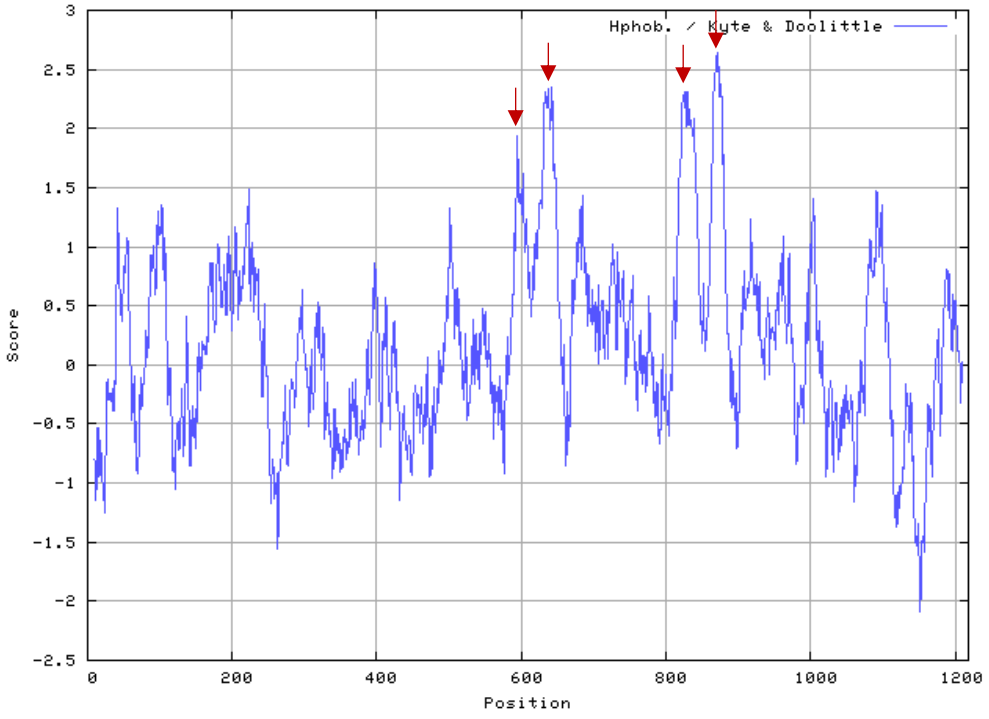


Figure 25: Transmembrane domain analysis of the 6-H4 synthetase. *Transmembrane helices were predicted with SOSUI (A) and THMM (B). Kyte & Doolittle Hydrophobicity plot (C) revealed four strong transmembrane helices indicated by the red arrows.*

Since all three employed programs predicted a different number of transmembrane helices, the synthetase was analyzed with the Kyte & Doolittle Hydrophobicity plot to get a better overview of the hydrophobic regions of the NPRS. Windows size 19 was used for the analysis, as this yields a hydrophobicity plot where transmembrane regions stand out sharply. The resulting plot of the 6-H4 synthetase indicated four strong transmembrane helices in the enzyme's terminal domain of unknown function (Fig. 25C).

Comparing the Kyte & Doolittle Hydrophobicity plot with the results of SOSUI and THMM led to the conclusion that the 6-H4 synthetase contains four strong transmembrane helices. The positions of those hydrophobic segments as predicted by SOSUI are shown in Fig. 25A (No. 1-2, 4-5). The fifth helix (No. 3) predicted by the program is not indicated to be a transmembrane domain when considering the Kyte & Doolittle plot, as is the case of the sixth domain predicted by THMM.

Based on this transmembrane domain analysis and undermined by the expression results, it can be hypothesized that the 6-H4 synthetase is a membrane protein. This is further affirmed by the Blast2Go Pro mapping and annotation results, as the synthetase was mapped and annotated as part of the cell membrane (gene ontology ID C: GO:0016020).

With the newly acquired knowledge that the NRPS-related synthetase of clone 6-H4 is presumably a membrane protein, it was attempted to discover if other membrane-related nonribosomal peptide synthetase were reported before.

4.4.5. The 6-H4 synthetase has similarities with homopoly(amino acid) synthetases

The effort of searching other membrane-related NRPSs led to the discovery of unusual peptide synthetases that synthesize so called homopoly(amino acids), biopolymers consisting only of one type of amino acid (Hamano et al. 2013). So far, only two of those homopoly(amino acid) synthesizing NRPSs have been characterized: the ε-poly-L-lysine synthetase (Yamanaka et al. 2008) and the poly(L-diaminoproprionic acid) synthetase (Xu et al. 2015). Both synthetases are membrane proteins with an A-domain and a PCP domain, typical for NRPSs, but instead of a condensation and a thioesterase domain, they have a NRPS terminal domain of unknown function. Upon further investigation by Yamanaka et al. (2008) and Xu et al. (2015), this domain of unknown function contains six transmembrane domains. Concurrently, the transmembrane domains separate three tandem soluble domains which are responsible of catalyzing the bond formation between the amino acids in the biopolymer chain.

The described domain architectures of those homopoly(amino acid) synthesizing NRPS are very similar to the domain structure of the 6-H4 synthetase mentioned in the previous chapter.

Figure 26: Domain structure comparison of the 6-H4 synthetase with homopoly(amino acid) synthesizing NRPSs. *The 6-H4 synthetase, the ε-poly-L-lysine synthetase (BAG68864.1) and the poly(L-diaminoproprionic acid) synthetase (EXU85975.1) have similar domain architecture. All three harbor an adenylation domain, a PCP domain and a NRPS terminal domain of unknown function, which contains left-handed beta helix structures.*

To better understand the similarities between the homopoly(amino acid) synthesizing NRPS and the 6-H4 synthetase, the amino acid sequences of the ε-poly-L-lysine synthetase and the poly(L-diaminoproprionic acid) synthetase were retrieved (NCBI accession numbers BAG68864.1 and EXU85975.1, respectively), their domain architecture was analyzed with NCBI Conserved Domain Search and compared with the domain architecture of the 6-H4 synthetase (Fig. 26). All three synthetases have an adenylation domain, a PCP domain and a terminal domain of unknown function with transmembrane domains separating soluble tandem domains. Those soluble domains contain left-handed beta helix structures which are often found in enzymes with acetyltransferase activity. Interestingly, it is said that acetyltransferases show structural similarities to NRPSs condensation domains (Hamano et al. 2013). Therefore, Yamanaka et al. (2008) and Xu et al. (2015) named those soluble domains C-domains as well.

Phylogenetic analysis was conducted, to further investigate the similarities between the 6-H4 synthetase and the homopoly(amino acid) synthetases. Several synthetase sequences that had similarities with the 6-H4 synthetase, the ε-poly- L-lysine synthetase of *Streptomyces albulus* and the poly (L-diaminoproprionic acid) synthetase of *Streptomyces albulus* PD-1 and were obtained from NCBI following a blastp analysis of those the three synthetases at hand. In addition, the amino acid sequence of the enterobactin synthetase EntF was obtained and used for the analysis. A phylogenetic tree was constructed using MEGA7 (Fig. 27).

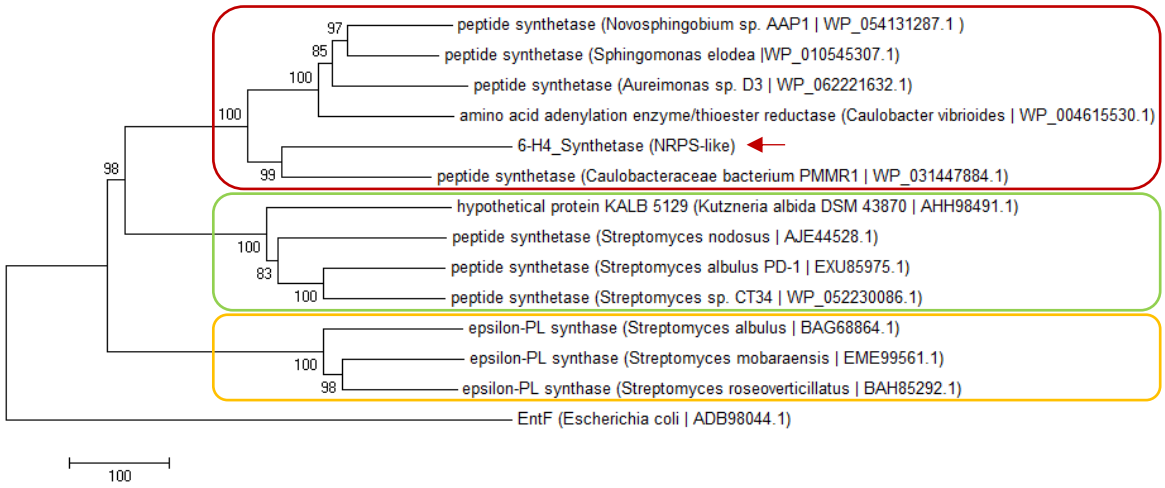


Figure 27: Phylogenetic analysis of the 6-H4 synthetase. *Tree constructed with MEGA7. Sequences were obtained from the NCBI protein database. ClustalW was used for sequence alignment and Neighbor-Joining algorithm for tree generation. The red arrow indicates the position of the 6-H4 synthetase.*

The synthetase sequences used for the phylogenetic analysis clustered in three different groups, one of ε-poly-L-lysine synthetases, one with the poly(L-diaminoproprionic acid) synthetase of *Streptomyces albulus* PD-1 and one containing the 6-H4 synthetase. The tree was rooted by the enterobactin synthetase EntF. The phylogenetic analysis revealed that the 6-H4

synthetase has more phylogenetic similarities with the known as well as putative homopoly(amino acid) synthesizing NRPSs, as with a typical NRPS like EntF, which presented itself as an outgroup.

Based on the results of the domain structure comparison and the transmembrane domain analysis, a model of the possible organization of the 6-H4 synthetase was generated (Fig. 28). The two soluble domains of the synthetase were named "C1" and "C2", applying the style of Yamanaka et al. (2008) and Xu et al. (2015).

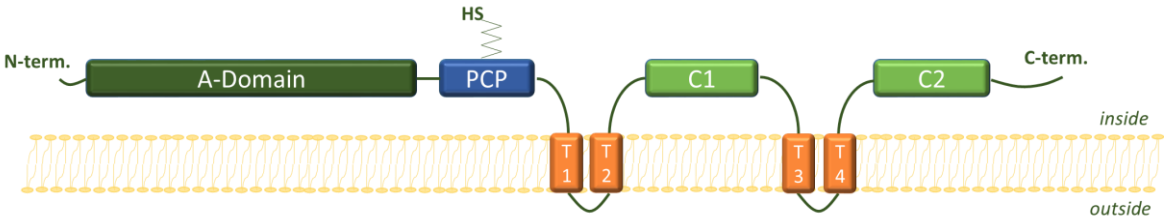


Figure 28: Model of the 6-H4 synthetase. *The adenylation domain, the thiolation domain (PCP), the four transmembrane domains and the two soluble C domains are shown schematically.*

This model has strong similarities with the models generated by Yamanaka et al. (2008) and Xu et al. (2015). The main difference between the two characterized homopoly(amino acid) synthetases and the 6-H4 synthetase is the number of transmembrane domains and tandem soluble domains. Both characterized synthetases have three condensation-like domains, however the 6-H4 synthetase has only two. Pairwise comparison of those soluble domains was performed, not only of the 6-H4 synthetase but also taking into account the soluble domains of the ε-poly-L-lysine synthetases and the poly(L-diaminoproprionic acid) synthetase (Fig 29). This analysis was performed, since both, Yamanaka et al. (2008) and Xu et al. (2015) detected sequence similarities between the soluble domains.

		1	2	3	4	5	6	7	8
6-H4_Synth_C1	1		23,40	31,72	22,05	21,35	18,62	22,22	26,40
6-H4_Synth_C2	2	23,40		17,28	37,08	17,22	19,79	39,78	24,04
BAG68864_S. albulus_C1	3	31,72	17,28		21,76	22,68	20,00	18,00	20,81
BAG68864_S. albulus_C2	4	22,05	37,08	21,76		20,86	17,55	36,90	22,34
BAG68864_S. albulus_C3	5	21,35	17,22	22,68	20,86		17,80	21,05	37,99
EXU85975_S. albulus PD-1_C1	6	18,62	19,79	20,00	17,55	17,80		20,73	20,62
EXU85975_S. albulus PD-1_C2	7	22,22	39,78	18,00	36,90	21,05	20,73		19,37
EXU85975_S. albulus PD-1_C3	8	26,40	24,04	20,81	22,34	37,99	20,62	19,37	

Figure 29: Pairwise comparison of the homopoly(amino acid) NRPSs and 6-H4 synthetase soluble domains. *The pairwise comparison table was generated using CLC Main Workbench applying the same settings as for the phylogenetic analysis. BAG68864 = ε-poly-L-lysine synthetase. EXU85975 = poly(L-diaminoproprionic acid) synthetase.*

The two soluble domains of the 6-H4 synthetase display 23.4% similarity, which is in a similar range as the comparison of the C-like domains of the homopoly(amino acid) NRPSs. Interestingly, the C1 domain of the 6-H4 synthetase and the C1 domain of the ε-poly-L-lysine synthetase have 31.72% similarity, and the C2 domain has 37.08% and 39.78% similarity with the C2 domains of the ε-poly-L-lysine synthetase and the poly(L-diaminoproprionic acid) synthetase, respectively.

Summarizing, the results from this chapter concerning domain structure and phylogenetic analysis, as well as pairwise comparison of the tandem soluble domains strongly suggest that the 6-H4 synthetase might be a novel homopoly(amino acid) synthesizing nonribosomal peptide synthetase.

5. Discussion

The great structural diversity of NRPs makes those bioactive compounds promising targets for biotechnological and pharmaceutical applications, as they include siderophores, antibiotics and surfactants (Martínez-Núñez & López 2016). In 2015, Müller et al. identified partial sequences of NRPS-related genes in a *Sphagnum magellanicum* metagenomic fosmid library, but no expression of the identified genes was attempted. It was unknown if the NRPS genes or gene clusters were complete and the NRPS-clones were able to synthetize functional peptides that could be further characterized. Therefore, the aim of this study was to perform a comprehensive analysis of the putative NRPSs from the Sphagnum moss metagenome found by Müller et al. (2015).

Since some of the NRPSs from the moss metagenome showed similarity to genes from biosynthetic pathways involved in the production of toxins (e.g. syringomycin), siderophores (e.g. pyoverdine) or antibiotics (e.g. gramicidin) (Müller et al. 2015), activity-based screenings were performed to examine the NRPS-containing fosmid clones for their role in the production of siderophores, antibiotics and antifungals. TLC and HPLC were applied for further product detection and NRPS-clone 6-H4 was chosen for whole-insert sequencing and sequence analysis, focusing on the recovery of the complete nonribosomal peptide synthetase gene for subsequent heterologous expression and further characterization. In the following chapters, the results of each analysis will be discussed separately.

5.1 Analysis of siderophore production

Due to the important role of siderophores in the environment and the wide field of possible applications, siderophores – small iron chelating compounds often synthesized by NPRSs – are gaining more and more attention in environmental research. Ahmed & Holmström (2014) reviewed the role of siderophores in the environment, like soil mineral weathering and biogeochemical cycling of iron in the ocean, and looked at biotechnological applications of siderophores. Applications that were discussed are the biological control of plant pathogens and fish pathogens, bioremediation of metals and petroleum hydrocarbons, nuclear fuel reprocessing, and the usage as optical biosensor or for bio-bleaching of pulps. Siderophores can also be used for microbial ecology and taxonomy studies by characterizing the microbes according to the siderophore types they produce – a process called siderotyping (Meyer et al. 2002). Metagenomic studies of microbial communities of various environments will not only help to answer important remaining questions regarding, for example, the importance of different siderophore structures in relation to microbial communities, but will also lead to the detection of

(novel) siderophore biosynthetic genes that could be used for the improvement or development of applications for those iron chelating compounds (Ahmed & Holmström 2014).

Some of the NRPSs from the moss metagenome showed similarities to biosynthetic pathway genes involved in the production of siderophores. Additionally, the DNA sequence of the 6-H4 synthetase gene harbored an A-domain with similarities to siderophore synthesizing NRPSs (SidN3-like and EntF-like). Therefore, the NRPS-harboring fosmid clones were analyzed for siderophore production using different approaches. One was the application of chromazurol S (CAS) agar plates. For this mean, the CAS plates had to be optimized first for the screening of metagenomic clones. Both, the plate assay, as well as the liquid photometric assay were developed by Schwyn & Neilands (1987). In this study, the CAS plates were prepared after Louden et al. (2011), Fujita et al. (2011) and Rondon et al. (2000). While the plates after Louden et al. were very similar to the plates by Schwyn & Neilands (both used deferrated casamino acids), Fujita et al. (2011) used LB agar mixed with a chromazurol S (CAS) solution. The used positive control, a *Pseudomonas* species from the institute's strain collection, showed on both types of plates a distinct orange halo formation. This indicates that the LB/CAS plates could also work well for siderophore screening of cultivable environmental strains, with the main advantage of being less time consuming in preparation. But the plates from Louden et al. and Fujita et al. showed the same disadvantage when it comes to the screening of metagenomic libraries in *E. coli*, as *E. coli* has native siderophore production. Fujita et al. reported that the LB/CAS plates are effective and sensitive enough for the screening of metagenomic libraries, and while this might be true when the heterologous expression of the siderophore synthetizing genes is strong, a slight heterologous siderophore production could be overshadowed by *E. coli's* native siderophore production. To overcome this problem, different methods were examined. One was the addition of Fe to the media to suppress the native siderophore production of *E. coli* (Rondon et al. 2000), and another was the employment of an *E. coli* deletion mutant (K-12 *ΔentF*), where a gene responsible for enterobactin synthesis is knocked out (Baba et al. 2006). In addition to the plate assay, the liquid CAS assay was employed, as this assay was thought to be more sensitive than the plate assay.

The re-transformation of the NRPS-harboring fosmids into the K-12 *ΔentF* strain was met with another problem, as, due to the lack of the protein TrfA, the copy number of the fosmids cannot be enhanced and a heterologous product production might be harder to detect. Therefore, a mutated version of the *trfA* gene was inserted into the genome of the knock-out strain, in order to enhance the ability of TrfA to upregulate the fosmid copy number. Unluckily, the chosen point mutation did not lead to a "copy-up" of the fosmids in the *trfA* A250V *ΔentF*

pCC2FOS_NRPS clones. An explanation for the lacking induction of the copy number could possibly be found in the used expression system for the *trfA* gene – the gene is under regulation of a PmG5 promotor which is activated by XylS in the presence of *m*-toluate (Aakvik et al. 2009). Mermod et al. (1986) wrote that "the degree of inducibility of the Pm promoter in *E. coli* is dependent on the level of *xylS* gene expression." It is possible that the *xylS* gene on the pRS48 transposon is transcribed from its native promoter, which is seemingly not efficient in *E. coli*. Additionally, the *trfA* gene is under a PmG5 promoter, a mutated derivative of Pm, which is considered a weaker promoter that shows lower background expression in the absence of an inducer (Bakkevig et al. 2005). Maybe the interplay of a weaker, less leaky promoter and an inefficient XylS production could have been the cause for the lack of fosmid copy number enhancement. Other factors influencing the copy-up efficacy need to be considered as well. Those factors could be the selected point mutation and as well as the employed host strain. In the Epicentre manual for EPI300 and pCC2FOS it is written, that a mutant *trfA* is used for the copy-up of the fosmid vector (Copy-Control™ HTP Fosmid Library Production Kit with pCC2FOS™ Vector; Epicentre, Wisconsin, USA), but there is no mention of how that mutation looks like. Perhaps, one of the other point mutations described by Durland et al. (1990) could be more effective, especially if the host strain *E. coli* K12 *ΔentF* is considered as well – the host range of copy-up mutants is limited not only by species-specific differences (Haugan et al. 1995), but seemingly also by strain-specific differences (Blatny et al. 1997).

Although a copy-up in the *entF* knock-out strain was not detectable with the methods applied, the production of the mutated TrfA was still induced. However, neither with the knock-out strain, nor with the EPI300 strain (where the native siderophore production was suppressed), it was possible to detect heterologous siderophore production of the NRPS-harboring fosmids– not with the plate assay, nor with the liquid photometric assay. Reasons therefore could be the bottlenecks of heterologous expression of foreign genes in *E. coli*. As mentioned in the introduction chapter 1.3., some of those hurdles could be partially overcome by using different host expression systems (Culligan et al. 2014; Uchiyama & Miyazaki 2009). Another reason could be that, although siderophore synthesizing NPRSs may be present on the fosmids, the gene clusters for the synthesis of a functional peptide are not complete, as those clusters frequently exceed the capacity of fosmid vectors (Nováková & Farkašovský 2013). At last, there is also the possibility that the NRPS genes on the moss metagenome containing fosmids do not play a role in siderophore production at all, although siderophore-like motives (like on clone 6-H4) or sequence similarities with siderophore synthesizing NRPSs (like clone 3-F3) exist. For example, the ε-poly-L-lysine synthetase characterized by Yamanaka et al. (2008) has also an A-domain with SidN3- and EntF-like motives, despite not acting as an iron chelating compound.

Although none of the investigated NRPS-containing fosmid clones exhibited siderophore production, the moss metagenome might still harbor various novel genes or gene cluster for siderophore synthesis. In the future, a siderophore screening of the whole metagenomic library generated by Müller et al. (2015) on CAS plates could be performed, to better assess the role of the moss metagenome in iron acquisition.

5.2 The antimicrobial activity of clone 2-F4

The steady increase of antibiotic resistances in bacteria is the driving force behind the search for novel antibiotics. NRPSs are known to produce several antibiotics, like vancomycin (described in chapter 1.4.2), but also daptomycin, teichoplanin, capreomycin and β-lactams (Felnagle et al. 2008). The NRPS-like sequences of the fosmid clones 2-F4 and 7-D4 detected by Müller et al. (2015) have sequence similarities with syringomycin and gramicidin synthetases. Syringomycin and gramicidin are both peptides with antimicrobial activity (Bender et al. 1999; Felnagle et al. 2008). Therefore, the NRPS-harboring fosmid clones were investigated for the production of antimicrobials.

Brady & Clardy (2000) as well as Rondon et al. (2000), Lim et al. (2005) and MacNeil et. al (2001) used the agar-overlay method to investigate their metagenomic libraries: the library was plated on agar plates containing the corresponding antibiotic needed to sustain the metagenomic DNA-containing vectors and overlayed with a *Bacillus subtilis* strain resistant to the same antibiotic. Agar-overlay is also a possible method for antifungal screening, as Chung et al. (2008) successfully screened a metagenomic library for antifungal production using a *Saccharomyces cerevisiae* overlay. Since a chloramphenicol-resistant *Bacillus subtilis* strain was not available, different methods were tried to investigate the NRPS-containing fosmid clones of Müller et al (2015), namely agar-disc and agar-well diffusion assays using culture supernatants, extracts and lyophilisates. Regarding the possible production of antifungals by the NRPS-clones, agar-overlay with *Candida albicans* could have been used for the screening, as this yeast, like *S. cerevisiae*, exhibited chloramphenicol resistance, but it was decided to investigate putative antifungal production with the same methods as applied for the bacterial strains.

Using the screening methods described in chapter 3.7, the native supernatant of clone 2-F4 exhibited growth inhibition of *S. epidermidis, S. aureus*, *B. subtilis* and the yeast *C. albicans*, but not of *M. luteus* and the other tested filamentous fungi. The antimicrobial effect was technically but not biologically reproducible. Nonetheless, control experiments confirmed that the antimicrobial activity was originated from clone 2-F4.

The most likely explanation for the non-reproducibility lies in the heterologous expression in *E. coli*. A successful heterologous expression depends on various factors: the ability to recognize foreign promoters, codon usage, and the ability to secret the products on the one hand, as well as incubation temperature and time, media composition, etc., on the other hand (Rosano & Ceccarelli 2014). It is possible that there were minimal, inconspicuous differences in incubation temperature or media composition, that led to a successful production and secretion of the antimicrobial once. In a future study, it would be advisable to vary and optimize the expression parameters in order to reproduce the production of the antimicrobial. Another factor is that *E. coli* might not be the most suitable host for a successful production and characterization of this putative antimicrobial. Iqbal et al. (2014) screened a soil metagenomic library in *Ralstonia metallidurans* and identified six clones exhibiting antimicrobial activity. Wondering why enzymes with antimicrobial activity are not reported more frequently in metagenome screenings, they investigated if those antimicrobial activities identified in the *R. metallidurans* library could have been detected with *E. coli* as a host, since *E. coli* is still the most used host for functional metagenomics. Interestingly, none of those six identified clones exhibited a detectable antimicrobial phenotype in *E. coli*, underlining the importance of different hosts in functional screenings of metagenomic libraries. Using a different expression host for the 2-F4 biosynthetic gene or gene cluster would be a promising strategy to obtain a more stable production of the antimicrobial.

To get an idea of the type of the antimicrobial produced by 2-F4, a further look was taken at the partial sequences of the NRPS of this clone. This analysis revealed similarities of the 2-F4 NRPS to several synthetases responsible for the production of antimicrobials and toxins (Table 4). Most interestingly, the domain structure of one of the two partial 2-F4 sequences has similarities with NRPS systems responsible for the synthesis of a lipopeptide-type antibiotic, like daptomycin (see chapter 4.2.2). Daptomycin has proven to be very amenable to engineering (Felnagle et al. 2008) – a point that also makes the 2-F4 synthetase itself very interesting besides of the putative synthesis of the product itself. It would be highly interesting to investigate if the modules of the 2-F4 synthetase could also be used for combinatorial biosynthesis.

Despite of the hurdles in producing an active antimicrobial, clone 2-F4 should be further analyzed. Even if the re-production of the antimicrobial proves not to be possible, the NPRS sequence of 2-F4 is a promising target for combinatorial biosynthesis.

5.3 TLC and HPLC analysis for metagenomic library screening

Compared with the aforementioned assays used for the detection of siderophores and antimicrobials, screenings of metagenomic libraries using thin-layer chromatography (TLC) and

high-performance liquid chromatography (HPLC) are sparse. Rabausch et al. (2013) developed a so-called "metagenome extract thin-layer chromatography analysis", short META, for rapid detection of glycosyltransferases and Wang et al. (2000) and Courtois et al. (2003) screened metagenomic libraries using HPLC, detecting novel chemical entities. Interestingly, both Wang et al. and Courtois et al. detected the novel compounds using *Streptomyces lividans* as host, further emphasizing the previously mentioned importance of using different host organisms. Besides the host strain, other bottlenecks may exist for detecting novel chemical compounds using TLC and HPLC, for example choosing the appropriate growth medium, extraction method and measuring parameters (HPLC) or solvent systems and staining (TLC).

In this study, TLC was performed to investigate the NRPS fosmid clones from the moss metagenome for the production of novel compounds. Several parameters were tested and optimized, such as the use of normal and reverse phase silica plates, three different solvent systems and three different visualization methods Additionally, HPLC coupled with UV detection was performed with the same goal. Those analyses showed that clone 6-H4 showed the most prominent differences in the product profile compared with the host strain, and was therefore chosen for preparative TLC and HPLC coupled with mass spectrometry for product identification. Out of 38 unique peaks of clone 6-H4 only three products could be conclusively identified: L-pyroglutamic acid, hydroxyisophthalic acid and phenylalanine.

L-Pyroglutamic acid, or 5-oxoproline, is the only slightly studied cyclic lactam of glutamic acid or glutamine and is present in almost all forms of life, from archaebacteria to humans (Kumar & Bachhawat 2012). Pyroglutamic acid was discovered in 1882 – when heated to 180 °C, glutamate is converted into its cyclized form. 50 years later the enzymatic formation of this metabolite by a γ-glutamyl cyclotransferase was demonstrated. Kumar & Bachhawat (2012) also mentioned that pyroglutamic acid can be formed from incomplete reactions following glutamate activation, as glutamate is highly unstable and prone to spontaneous cyclization when in an activated state – an interesting point, as glutamic acid, as an amino acid, could be the substrate of the 6-H4 NRPS-like synthetase. It could be possible that glutamic acid or glutamine was activated by the 6-H4 synthetase, but the following condensation process did not occur, leading to a high amount of activated glutamate/glutamine that spontaneously cyclized. Another amino acid was detected in elevated numbers in 6-H4 as well: phenylalanine. These two amino acids could be substrates for the 6-H4 synthetase. However, it cannot be excluded that they might be substrates or products of other enzymes present in the culture.

Although the abovementioned products of 6-H4 could be conclusively identified, many masses and mass spectra could not be assigned to known compounds. On the one hand, many predicted chemical formulas were assigned to several compounds and manual comparison with

the spectra of all possible molecules was not feasible. On the other hand, the metagenomic insert of this clone harbors various genes coding for novel enzymes (see chapter 4.4), therefore, there is a high possibility that some of the products found in the HPLC analysis could be completely novel substances with unknown mass spectra. In both cases, further analysis of the compounds, e.g. using NMR spectroscopy (Wüthrich 1986), would be necessary.

In conclusion, TLC and HPLC analytics are promising tools for the primary detection of novel chemicals in metagenomic libraries, although the experimental design used in this study would be too demanding for a complete library screening. Still, preparative TLC and HPLC/MS analytic with subsequent NMR spectroscopy could lead to the identification of several novel compounds of pre-selected clones – as Courtois et al. (2003) said, "[…] the strategic use of prescreening should greatly enhance the ability to detect novel and useful secondary metabolites."

5.4 Sequence analysis of the NRPS-containing fosmid clone 6-H4

While activity-based screenings and TLC/HPLC assays of metagenomic libraries can reveal novel compounds of biotechnological or medical interest, the actual frequency of identified compounds is quite low (Culligan et al. 2014). Incompatibility with transcription, absence of building blocks, different codon usage, lack of transport systems etc. are some of the expression-related limitations scientist have to face when it comes to heterologous expression of genes or gene clusters from metagenomes (Culligan et al. 2014; Uchiyama & Miyazaki 2009). This also means that a lot of potentially interesting enzymes would be lost, when only functional screenings of metagenomic libraries would be employed. Sequence-based screenings allow the identification of genes coding for target proteins, but those screenings still have one important limitation, as only variants of known protein families can be detected (Simon & Daniel 2011). Nonetheless, sequence-based screenings were capable of identifying novel enzymes interesting for industrial application, like [NiFe] hydrogenases (Maróti et al. 2009), nitrile hydratases (Liebeton & Eck 2004) or glycosyl hydrolases (Breves et al. 2007 [Patent]).

The *S. magellanicum* metagenomic library of Müller et al. (2015) was screened for NRPS and polyketide synthase genes using a sequence-based approach. The mentioned study used only partial sequences of the identified NRPS-related genes from the metagenome for identification and alignment to available databases. To complete a comprehensive analysis of the NPRSs from the moss metagenome, one NRPS fosmid clone was selected in this study for DNA-sequence analysis. Clone 6-H4 was chosen, due to preliminary results from a liquid siderophore assay. After whole-insert sequencing, the sequence of the NRPS-like synthetase was retrieved for

analysis. The insert sequence of this clones was additionally analyzed for other putative open reading frames (ORFs).

Using NBCI ORFfinder, 84 putative open reading frames could be identified from an insert sequence of 44.7 kb, leading to a first conclusion that the *Sphagnum* moss metagenome is a rich source for novel enzymes. Those 84 putative ORFs were analyzed with two different annotation methods: gene ontology (GO) annotation and clusters of orthologous groups of proteins (COG) annotation. Using GO annotation, 23 of the 84 genes of 6-H4 could be annotated, while 25 could be annotated using the COG database. A major bottleneck to the progress of metagenomics is the large number of not-validated annotations in current databases (Ferrer et al. 2009). Therefore, using different annotation methods can help to identify the correct function of a novel protein. On the one hand, when comparing similar categories of the GO and COG annotations of the 6-H4 ORFs, overlapping annotation existed for some genes, enhancing the possibility that the function of the gene was assigned correctly. On the other hand, other comparable annotation categories of COG and GO were assigned to completely different genes.

Some of the annotated genes had similarities with glycine dehydrogenases (ORF 2 and 14), dTDP-4-dehydrorhamnose reductases (ORF 21), peptidases (ORF 1), short-chain dehydrogenases (ORFs 57 and 68), oxidoreductases (ORFs 8 and 10) or β-galactosidases (ORF 21) – enzymes that are interesting for industrial applications. For example, β-galactosidases are widely used in the food industry and have been exploited in various other applications as well (Husain 2010). Further evaluation of the putative enzymes of 6-H4, including heterologous expression and characterization of selected genes is advisable. This could lead to discovery of novel enzymes for industrial applications.

Besides functional annotations, cluster analysis with antiSMASH was performed to identify putative gene clusters in the 6-H4 insert sequence. This analysis revealed a gene cluster spanning about 22 kb of the insert, consisting of four transport related genes, two biosynthetic genes and four other genes. The transport related genes showed sequence similarities with, for example, an efflux RND transporter periplasmic adaptor subunit (ORF 22) or a TonB-dependent receptor (ORF 66). The other genes showed similarities to an aldo-keto-reductase (ORF 10), a GMC family oxidoreductase (ORF 8), a DNA topoisomerase (ORF 9), a twin-arginine translocation pathway signal protein (ORF 25) and a 3-methyl-2-oxobutanoate hydroxymethyltransferase (ORF 39). These five genes are in close vicinity to the putative NRPS-like synthetase (ORF 27) of clone 6-H4 and could be involved in the biosynthesis of the nonribosomal peptide. Unbeknown for now is if the 6-H4 NRPS is really the last gene in the biosynthetic cluster or if there are still genes downstream of the synthetase, as the insert of 6-H4 ends shortly after the NRPS-

like gene. To investigate if this cluster is complete or not, a further screening would be necessary, for example a second PCR screening of the metagenomic fosmid library using the sequence of the synthetase for primer generation.

Moving forward with the characterization of the 6-H4 synthetase, a closer look was taken at the domain architecture of the enzyme. This analysis revealed the unusual structure of the putative synthetase, as it has an A-domain and a PCP domain typical for NRPSs, but it lacks a C-domain and a TE-domain. Instead, this novel synthetase has an NRPS terminal domain of unknown function. Further investigation revealed that this terminal domain consisted actually of two tandem soluble domains divided by four transmembrane helices. Searching for other membrane-related NRPSs led to the discovery of so called homopoly(amino acid) synthetases. The 6-H4 synthetase showed structural similarities with both characterized homopoly(amino acid) synthesizing NRPSs, the ε-poly-L-lysine synthetase (Yamanaka et al. 2008) and the poly(L-diaminoproprionic acid) synthetase (Xu et al. 2015). The most prominent difference between those homopoly(amino acid) synthetases and the 6-H4 synthetase was the number of tandem soluble domains, as both, ε-poly-L-lysine synthetase and the poly(L-diaminoproprionic acid) synthetase, have three of those condensation-like domains while the 6-H4 synthetase has only two. Interestingly, all putative homopoly(amino acid) synthetase sequences used for the phylogenetic study, as well as additionally analyzed similar sequences seem to have three tandem domains. It can be concluded that, concerning the number of tandem soluble domains, the 6-H4 synthetase is seemingly unique. Nonetheless, the similar domain architecture, the phylogenetic analysis and the pairwise comparison of the tandem domains strongly suggest that the 6-H4 synthetase might be a novel homopoly(amino acid) synthesizing nonribosomal peptide synthetase.

Substrate specificity analysis using NRPSpredictor2 was performed to get an idea of the possible substrate and therefore the putative homopoly(amino acid) the 6-H4 synthetase might produce. Possible substrates for the A-domain of the synthetase could be hydrophobic-aliphatic amino acids, although the prediction score was quite low (0.140733). 2,4-diaminobutyric acid was predicted as the nearest neighbor based on the Stachelhaus code with a score of 60%. This unnatural amino acid is a precursor of siderophores (Vandenende et al. 2004) and can be formed from L-glutamate and L-aspartate-β-semialdehyde via the action of an aminotransferase (Ikai & Yamamoto 1997). Since it is formed from L-glutamate, 2,4-diaminobutyric acid has structural similarities to both, glutamic acid and glutamine (Attachment, Fig. A9). Interestingly the cyclized form of those two amino acids, L-pyroglutamic acid, was found in elevated numbers in the HPLC/MS analysis of 6-H4 (see chapters 4.3.2 and 5.3). L-glutamate is formed from α-ketoglutarate, an intermediate product of the citric acid cycle (Takehara et al. 2008). The

oxidoreductase glutamate dehydrogenase is one of the enzymes that catalyze the reversible reaction of α-ketoglutarate to L-glutamate (Epstein & Grossowicz 1975). Two genes in the identified gene cluster of 6-H4 show similarities to oxidoreductases and one gene has strong similarities to a 3-methyl-2-oxobutanoate hydroxymethyltransferase (ketopantoate hydroxymethyltransferase) – an enzyme that plays a role in the biosynthesis of coenzyme A (Teller et al. 1976). Coenzyme A plays an important role in both, the citric acid cycle (Müller-Esterl et al. 2011) and the activation of the PCP domain from its apo to its active holo form (Hopwood 2009). But this is not the only reason this putative ketopantoate hydroxymethyltransferase of 6-H4 is interesting. One of the two products of this enzyme is tetrahydrofolate, which is involved in the conversion of formiminoglutamic acid to glutamic acid during histidine degradation (Müller-Esterl et al. 2011). These findings enhance the possibility that either L-glutamic acid, L-glutamine (synthetized from glutamate via the action of the glutamine synthetase, EC 6.3.1.2), or 2,4-diaminobutyric acid could be substrates for the putative homopoly(amino acid) synthetase of 6-H4.

Both diaminobutyric acid and glutamic acid are known to occur in homopoly(amino acids). While poly(γ-L-diaminobutanoic acid) is produced by *Streptomyces* species (Takehara et al. 2008), poly(γ-glutamic acid) is known to be mostly produced by *Bacillus* species (Hamano et al. 2013), although it has been shown that gram-negative bacteria are also capable to produce this homopolymer (Candela et al. 2009). Two mechanisms of biosynthesis for poly(γ-glutamic acid) have been postulated: the thiotemplate mechanism catalyzed by NRPSs (Gardner & Troy 1979) and the amide-ligation mechanism catalyzed by amide ligases (Hamano et al. 2013), with the latter being more probable so far. Still, the existence of a NRPS-like system for poly(γ-glutamic acid) is not ruled out, although corresponding synthetase genes remain to be discovered (Hamano et al. 2013). Regarding poly(γ-L-diaminobutanoic acid), genetic and biosynthesis studies still need to be performed to identify responsible genes or operons for the production of this homopolymer (Takehara et al. 2008). Although poly-L-glutamine is known to be part of cell walls of several bacterial species (Kandler et al. 1983; Tripathi et al. 2013), it was not yet detected as a secretion product. If the 6-H4 NRPS-like synthetase truly has either glutamic acid, glutamine or diaminobutyric acid as substrate and is responsible for homopoly(amino acid) synthesis it would be the first described unusual NRPS to synthetize homopolymers from the before mentioned amino acids.

Substrate specificity assays need to be performed to clearly determine the substrate of the 6-H4 synthetase. Various assays to determine the substrate of A-domains exist: PP_i-ATP exchange assay (radioactively labeled PP_i), malachite green assay, 2-amino-6-mercapto-7-methylpurine (MesG)-based assay and a novel coupled NADH/pyrophosphate (PP_i) detection

assay (Kittilä et al. 2016). For some of those assays the PCP domain of the synthetase needs to be in its active holo form, which was one of the reasons why the 6-H4 synthetase was expressed in the engineered *E. coli* strain BAP1. This strain co-expresses the *Bacillus subtilis* phosphopantetheinyl transferase *sfp* to assure the phosphopantetheinylation of the PCP domain (Pfeifer et al. 2001). Although the 6-H4 synthetase was expressed as insoluble protein, most likely due to its character as a membrane protein, the synthetase still might function for substrate specificity assays. This was indicated by a HPLC/MS analysis of the culture supernatant of BAP1, as the amount of AMP found in the supernatant was highly elevated when the expression of the 6-H4 synthetase was induced (data not shown). AMP is released in the second step of amino acid activation when the amino acid is covalently tethered to the PCP domain of the NRPS (Felnagle et al. 2008). The detection of the putative homopolymer of the 6-H4 synthetase is a next challenge that needs to be addressed in further research. For this goal, the expression of the synthetase has to be optimized. Zoonens & Miroux (2010) describe several parameters that need to be considered when expressing membrane proteins in *E. coli* and also give solution approaches. One interesting example to increase the rate of successful membrane protein expression is the co-expression of the b-subunit of *E. coli* adenosine triphosphate (ATP)-synthase. This triggers the proliferation of intracytoplasmic membranes. When this b-subunit is expressed as a fusion protein with the target protein, Zoonens & Miroux (2010) even suggest the use of the T7 expression system – the same system that was already used in this study. Another approach to optimize the expression of the 6-H4 synthetase would be to use different expression hosts. All three, ε-poly-L-lysine, poly(γ-L-diaminobutanoic acid) and poly(L-diaminopropionic acid) are synthetized by *Streptomyces* species, and poly(γ-glutamic acid) is mainly produced by *Bacillus* species. Using a gram-positive expression host could be more suitable for the 6-H4 synthetase.

Although further research needs to be performed to gain better insights into the nature of the synthetase and the product of the 6-H4 synthetase still needs to be identified, the presented results strongly suggest that the 6-H4 synthetase might be a novel homopoly(amino acid) synthesizing NRPS. Verification of this assumption would also make this synthetase and its putative homopolymer highly valuable for industrial applications. Both, ε-poly-L-lysine and poly(γ-glutamic acid) are already widely used in the industry. ε-poly-L-lysine is nontoxic towards humans and the environment, it is water soluble, biodegradable and edible. Due to its antimicrobial properties it is already used as a food preservative in Japan and is being examined for other applications as well (Chheda & Vernekar 2015). Poly(γ-glutamic acid) has the same beneficial characteristics as ε-poly-L-lysine. It and its derivatives are currently being tested for or already used in the food, medical, and wastewater industries, for example as food supplement, as thickener for fruit juices, as drug carrier or metal chelator, as biopolymer flocculant, as biocontrol

agent or as biodegradable plastic (Luo et al. 2016). Also poly(γ-L-diaminobutanoic acid) and poly(L-diaminoproprionic acid) show antimicrobial activities and are currently under investigation (Takehara et al. 2008; Xu et al. 2015).

Taking everything together, this novel NRPS-like synthetase from the *Sphagnum* moss metagenome needs to be further characterized and its putative homopolymer product needs to be experimentally verified. Should the synthetase prove to be homopolymer synthetizing, it would be the third characterized unusual NRPS known to produce homopoly(amino acids) and actually the first to be characterized from a metagenome.

6. Conclusion

Bioprospecting metagenomes has proven to be a promising tool for the detection of novel natural compounds. Both, function-based screenings and sequence-based screenings of metagenomes led to the identification of various enzymes and biologicals interesting for industrial applications (Coughlan et al. 2015; Lorenz & Eck 2005; Banik & Brady 2010). Many prominent classes of bioactive compounds of interest, like antibiotics, siderophores or antifungals, are synthetized by nonribosomal peptide synthetases (NRPSs).

Performing a comprehensive analysis of putative NRPSs identified by Müller et al. (2015) in a *Sphagnum* moss metagenomic fosmid library led to the identification of a NRPS system that might be responsible for the synthesis of a lipopeptide-type antibiotic. Since this class of antibiotics has proven to be very amenable to engineering (Felnagle et al. 2008), the NRPS system of this fosmid clone (2-F4) could be used for combinatorial biosynthesis. Sequence analysis of another *Sphagnum* moss metagenomic fosmid clone (6-H4) revealed an enormous number of enzyme-coding genes; some of the genes could be of interest for industrial applications. Further characterization of interesting enzymes originated from the metagenomic clone is advisable. The unusual structure of the NRPS-like gene of clone 6-H4 has strong similarities with homopoly(amino acid) synthetizing NRPSs. Homopoly(amino acids) like ε-poly-L-lysine and poly(γ-glutamic acid) have enormous potential for various industrial applications, for example as food preservative, as biocontrol agent or as drug carrier (Chheda & Vernekar 2015; Luo et al. 2016). Possible substrates for the discovered synthetase could be glutamic acid, glutamine, or 2,4-diaminobutyric acid, based on sequence information and HPLC/MS results. Should this novel NRPS-like synthetase prove in further studies to be homopolymer synthetizing, it would be the third characterized NRPS known to produce biopolymers consisting of only one type of amino acid. This synthetase would also be the first homopoly(amino acids) synthetizing NRPS to be characterized from a metagenome.

Summarizing the findings in this study it can be concluded that bioprospecting plant-associated metagenomes is indeed very promising for the identification of industrially interesting metabolites and enzymes. The analysis of only twelve NRPS-containing fosmid clones from a *Sphagnum* moss metagenomic library led the identification of putative NRPS machineries responsible for the synthesis of a lipopeptide-type antibiotic and an amino acid homopolymer and revealed an enormous number of enzymes of possible industrial interest.

References

Aakvik, T. et al., 2009. A plasmid RK2-based broad-host-range cloning vector useful for transfer of metagenomic libraries to a variety of bacterial species. *FEMS Microbiology Letters*, 296(2), pp.149–158.

Ahmed, E. & Holmström, S.J.M., 2014. Siderophores in environmental research: roles and applications. *Microbial Biotechnology*, 7(3), pp.196–208.

Ambrogelly, A., Palioura, S. & Söll, D., 2007. Natural expansion of the genetic code. *Nature Chemical Biology*, 3(1), pp.29–35.

Amoutzias, G.D., Chaliotis, A. & Mossialos, D., 2016. Discovery Strategies of Bioactive Compounds Synthesized by Nonribosomal Peptide Synthetases and Type-I Polyketide Synthases Derived from Marine Microbiomes. *Marine drugs*, 14(4), pii: E80.

Baba, T. et al., 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular systems biology*, 2:2006.0008.

Bakkevig, K. et al., 2005. Role of the *Pseudomonas fluorescens* alginate lyase (AlgL) in clearing the periplasm of alginates not exported to the extracellular environment. *Journal of bacteriology*, 187(24), pp.8375–84.

Banik, J.J. & Brady, S.F., 2010. Recent application of metagenomic approaches toward the discovery of antimicrobials and other bioactive small molecules. *Current opinion in microbiology*, 13(5), pp.603–9.

Barton, H.A. & Jurado, V., 2007. What's Up Down There? Microbial Diversity in Caves. *Microbe-American Society for Microbiology* 2(3), pp.132–138.

Bender, C.L., Alarcón-Chaidez, F. & Gross, D.C., 1999. *Pseudomonas syringae* phytotoxins: mode of action, regulation, and biosynthesis by peptide and polyketide synthetases. *Microbiology and molecular biology reviews* 63(2), pp.266–92.

Berg, G. et al., 2016. The plant microbiome explored: Implications for experimental botany. *Journal of Experimental Botany*, 67(4), pp.995–1002.

Berg, G. et al., 2014. Unraveling the plant microbiome: looking back and future perspectives. *Frontiers in Microbiology,* 5, p.148.

Blanusa, M. et al., 2010. Phosphorothioate-based ligase-independent gene cloning (PLICing): An enzyme-free and sequence-independent cloning method. *Analytical Biochemistry*, 406(2), pp.141–146.

Blatny, J.M. et al., 1997. Improved Broad-Host-Range RK2 Vectors Useful for High and Low Regulated Gene Expression Levels in Gram-Negative Bacteria. *Plasmid*, 38(1), pp.35–51.

Brady, S.F. & Clardy 2, J., Long-Chain N-Acyl Amino Acid Antibiotics Isolated from Heterologously Expressed Environmental DNA. *Journal of the American Chemical Society* 122(51), pp.12903-12904

Bragina, A. et al., 2014. The *Sphagnum* microbiome supports bog ecosystem functioning under extreme conditions. *Molecular Ecology*, 23(18), pp.4498–4510.

Breves, R. et al., 2007. Glycosyl hydrolases. U.S. Patent No. 7,300,782. Washington, DC: U.S. Patent and Trademark Office.

Candela, T. et al., 2009. *Fusobacterium nucleatum*, the first Gram-negative bacterium demonstrated to produce polyglutamate. *Canadian journal of microbiology*, 55(5), pp.627–32.

Challis, G.L. & Naismith, J.H., 2004. Structural aspects of non-ribosomal peptide biosynthesis. *Current Opinion in Structural Biology*, 14(6), pp.748–756.

Chheda, A.H. & Vernekar, M.R., 2015. A natural preservative ε-poly-L-lysine: fermentative production and applications in food industry. *International Food Research Journal*, 22(1), pp.23–30.

Chung, E.J. et al., 2008. Forest soil metagenome gene cluster involved in antifungal activity expression in *Escherichia coli*. *Applied and environmental microbiology*, 74(3), pp.723–30.

Clark, B.L., 2004. Characterization of a Catechol-Type Siderophore and the Detection of a Possible Outer Membrane Receptor Protein from *Rhizobium leguminosarum* strain IARI 312. *Dissertation*.

Collier, H.O. & Chesher, G.B., 1956. Antipyretic and analgesic properties of two hydroxyisophthalic acids. *British journal of pharmacology and chemotherapy*, 11(1), pp.20–6.

Conti, E. et al., 1997. Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *The EMBO journal*, 16(14), pp.4174–83.

Coughlan, L.M. et al., 2015. Biotechnological applications of functional metagenomics in the food and pharmaceutical industries. *Frontiers in microbiology,* 6, p.672.

Courtois, S. et al., 2003. Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. *Applied and environmental microbiology*, 69(1), pp.49–55.

Crosa, J.H. & Walsh, C.T., 2002. Genetics and assembly line enzymology of siderophore biosynthesis in bacteria. *Microbiology and molecular biology reviews* 66(2), pp.223–49.

Culligan, E.P. et al., 2014. Metagenomics and novel gene discovery: promise and potential for novel therapeutics. *Virulence*, 5(3), pp.399–412.

Daniel, R., 2005. The metagenomics of soil. *Nature Reviews Microbiology*, 3(6), pp.470–478.

Dreyfuss, M. et al., 1976. Cyclosporin A and C. *Applied Microbiology and Biotechnology, 3(2), pp.125-133.*

Durland, R.H. et al., 1990. Mutations in the *trfA* replication gene of the broad-host-range plasmid RK2 result in elevated plasmid copy numbers. *Journal of Bacteriology*, 172(7), pp.3859–3867.

Ehmann, D.E. et al., 2000. The EntF and EntE adenylation domains of *Escherichia coli* enterobactin synthetase: sequestration and selectivity in acyl-AMP transfers to thiolation domain cosubstrates. *Proceedings of the National Academy of Sciences of the United States of America*, 97(6), pp.2509–14.

Epstein, I. & Grossowicz, N., 1975. Purification and properties of glutamate dehydrogenase from a thermophilic bacillus. *Journal of bacteriology*, 122(3), pp.1257–64.

Felnagle, E.A. et al., 2008. Nonribosomal Peptide Synthetases Involved in the Production of Medically Relevant Natural Products. *Molecular Pharmaceutics*, 5(2), pp.191–211.

Ferrer, M. et al., 2009. Metagenomics for mining new genetic resources of microbial communities. *Journal of molecular microbiology and biotechnology*, 16(1–2), pp.109–23.

Firáková, S., Proksa, B. & Sturdíkova, M., 2007. Biosynthesis and biological activity of enniatins. *Die Pharmazie*, 62(8), pp.563–8.

Fujita, M.J. et al., 2011. Cloning and Heterologous Expression of the Vibrioferrin Biosynthetic Gene Cluster from a Marine Metagenomic Library. *Bioscience, Biotechnology, and Biochemistry*, 75(12), pp.2283–2287.

Gabor, E. et al., 2007. Updating the metagenomics toolbox. *Biotechnology Journal*, 2(2), pp.201–206.

Gardner, J.M. & Troy, F.A., 1979. Chemistry and biosynthesis of the poly(gamma-D-glutamyl) capsule in *Bacillus licheniformis.* Activation, racemization, and polymerization of glutamic acid by a membranous polyglutamyl synthetase complex. *The Journal of biological chemistry*, 254(14), pp.6262–9.

Gomez-Alvarez, V., Revetta, R.P. & Santo Domingo, J.W., 2012. Metagenomic analyses of drinking water receiving different disinfection treatments. *Applied and Environmental Microbiology*, 78(17), pp.6095–102.

Götz, S. et al., 2008. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic acids research*, 36(10), pp.3420–35.

Gunatilaka, A.A.L., 2006. Natural products from plant-associated microorganisms: distribution, structural diversity, bioactivity, and implications of their occurrence. *Journal of Natural Products*, 69(3), pp.509–26.

Hall, B.G., 2013. Building Phylogenetic Trees from Molecular Data with MEGA. *Molecular Biology and Evolution*, 30(5), pp.1229–1235.

Hamano, Y. et al., 2013. NRPSs and amide ligases producing homopoly(amino acid)s and homooligo(amino acid)s. *Natural Product Reports*, 30(8), p.1087.

Hamedi, J., Imanparast, S. & Mohammadipanah, F., 2015. Molecular, chemical and biological screening of soil actinomycete isolates in seeking bioactive peptide metabolites. *Iranian Journal of Microbiology*, 7(1), pp.23–30.

Han, J.W. et al., 2012. Site-directed modification of the adenylation domain of the fusaricidin nonribosomal peptide synthetase for enhanced production of fusaricidin analogs. *Biotechnology Letters*, 34(7), pp.1327–34.

Handelsman, J. et al., 1998. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry & Biology*, 5(10), pp.R245–R249.

Haugan, K. et al., 1995. The Host Range of RK2 Minimal Replicon Copy-Up Mutants Is Limited by Species-Specific Differences in the Maximum Tolerable Copy Number. *Plasmid*, 33(1), pp.27–39.

Hess, M. et al., 2011. Metagenomic Discovery of Biomass-Degrading Genes and Genomes from Cow Rumen. *Science*, 331(6016), pp.463-467.

Hirokawa, T., Boon-Chieng, S. & Mitaku, S., 1998. SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics*, 14(4), pp.378–379.

Hodges, T.W., Slattery, M. & Olson, J.B., 2012. Unique Actinomycetes from Marine Caves and Coral Reef Sediments Provide Novel PKS and NRPS Biosynthetic Gene Clusters. *Marine Biotechnology*, 14(3), pp.270–280.

Hofmann, K. & Stoffel, W., 1993. TMbase - A database of membrane spanning proteins segments. *Biol. Chem. Hoppe-Seyler*, 374(166).

Hopwood, D.A., 2009. Complex enzymes in microbial natural product biosynthesis. Part A: Overview Articles and Peptides. *Methods Enzymol.,* Volume 458., ISBN: 9780123745880.

Howden, B.P. et al., 2010. Reduced Vancomycin Susceptibility in *Staphylococcus aureus*, Including Vancomycin-Intermediate and Heterogeneous Vancomycin-Intermediate Strains: Resistance Mechanisms, Laboratory Detection, and Clinical Implications. *Clinical Microbiology Reviews*, 23(1), pp.99–139.

Hur, G.H., Vickery, C.R. & Burkart, M.D., 2012. Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. *Natural Product Reports*, 29(10), p.1074.

Husain, Q., 2010. β Galactosidases and their potential applications: a review. *Critical Reviews in Biotechnology*, 30(1), pp.41–62.

Ikai, H. & Yamamoto, S., 1997. Identification and analysis of a gene encoding L-2,4-diaminobutyrate:2-ketoglutarate 4-aminotransferase involved in the 1,3-diaminopropane production pathway in *Acinetobacter baumannii. Journal of bacteriology*, 179(16), pp.5118–25.

Iqbal, H.A., Craig, J.W. & Brady, S.F., 2014. Antibacterial enzymes from the functional screening of metagenomic libraries hosted in *Ralstonia metallidurans*. *FEMS microbiology letters*, 354(1), pp.19–26.

Iqbal, H.A., Feng, Z. & Brady, S.F., 2012. Biocatalysts and small molecule products from metagenomic studies. *Current Opinion in Chemical Biology*, 16(1–2), pp.109–116.

Janso, J.E. & Carter, G.T., 2010. Biosynthetic potential of phylogenetically unique endophytic actinomycetes from tropical plants. *Applied and Environmental Microbiology*, 76(13), pp.4377–86.

Johnston, C.W. et al., 2015. An automated Genomes-to-Natural Products platform (GNP) for the discovery of modular natural products. *Nature Communications*, 6, p.8421.

Kandler, O. et al., 1983. Occurrence of Poly-γ-D-Glutamic Acid and Poly-α-L-Glutamine in the Genera *Xanthobacter*, *Flexithrix*, *Sporosarcina* and *Planococcus*. *Systematic and Applied Microbiology*, 4(1), pp.34–41.

Kanokratana, P. et al., 2004. Diversity and abundance of Bacteria and Archaea in the Bor Khlueng Hot Spring in Thailand. *Journal of Basic Microbiology*, 44(6), pp.430–444.

Kaplan, J. & Ward, D.M., 2013. The essential nature of iron usage and regulation. *Current biology,* 23(15), pp.R642-6.

Kittilä, T., Schoppet, M. & Cryle, M.J., 2016. Online Pyrophosphate Assay for Analyzing Adenylation Domains of Nonribosomal Peptide Synthetases. *ChemBioChem*, 17(7), pp.576–584.

Kleinkauf, H. & Von Döhren, H., 1996. A nonribosomal system of peptide biosynthesis. *European Journal of Biochemistry*, 236(2), pp.335–51.

Konz, D. & Marahiel, M.A., 1999. How do peptide synthetases generate structural diversity? *Chemistry & Biology*, 6(2), pp.R39–R48.

Kries, H. & Hilvert, D., 2011. Tailor-Made Peptide Synthetases. *Chemistry & Biology*, 18(10), pp.1206–1207.

Krogh, A. et al., 2001. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. Edited by F. Cohen. *Journal of Molecular Biology*, 305(3), pp.567–580.

Kumar, A. & Bachhawat, A.K., 2012. Pyroglutamic acid: throwing light on a lightly studied metabolite. *Current Science*, 102(2), pp.288-297.

Lai, J.R., Koglin, A. & Walsh, C.T., 2006. Carrier Protein Structure and Recognition in Polyketide and Nonribosomal Peptide Biosynthesis. *Biochemistry*, 45(50), pp.14869–14879.

Laireiter, C.M. et al., 2014. Active anti-microbial effects of larch and pine wood on four bacterial strains. *BioResources*, 9(1), pp.273–281.

Lambalot, R.H. et al., 1996. A new enzyme superfamily - the phosphopantetheinyl transferases. *Chemistry & biology*, 3(11), pp.923–36.

Lehner, S.M. et al., 2013. Isotope-assisted screening for iron-containing metabolites reveals a high degree of diversity among known and unknown siderophores produced by *Trichoderma* spp. *Applied and Environmental Microbiology*, 79(1), pp.18–31.

Lewis, K., 2013. Platforms for antibiotic discovery. *Nature Reviews Drug Discovery*, 12(5), pp.371–387.

Liebeton, K. & Eck, J., 2004. Identification and Expression in *E. coli* of Novel Nitrile Hydratases from the Metagenome. *Engineering in Life Sciences*, 4(6), pp.557–562.

Lim, H.K. et al., 2005. Characterization of a forest soil metagenome clone that confers indirubin and indigo production on *Escherichia coli*. *Applied and Environmental Microbiology*, 71(12), pp.7768–77.

Ling, L.L. et al., 2015. A new antibiotic kills pathogens without detectable resistance. *Nature*, 517(7535), pp.455–459.

Liu, C. et al., 2011. Clinical practice guidelines by the infectious diseases society of america for the treatment of methicillin-resistant *Staphylococcus aureus* infections in adults and children: executive summary. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 52(3), pp.285–92.

Liu, Y., Zheng, T. & Bruner, S.D., 2011. Structural basis for phosphopantetheinyl carrier domain interactions in the terminal module of nonribosomal peptide synthetases. *Chemistry & biology*, 18(11), pp.1482–8.

Lorenz, P. & Eck, J., 2005. Metagenomics and industrial applications. *Nature Reviews Microbiology*, 3(6), pp.510–6.

Louden, B.C., Haarmann, D. & Lynne, A.M., 2011. Use of Blue Agar CAS Assay for Siderophore Detection. *Journal of Microbiology & Biology Education*, 12(1), pp.51–53.

Luo, Z. et al., 2016. Microbial synthesis of poly-γ-glutamic acid: current progress, challenges,

and future perspectives. *Biotechnology for biofuels*, 9(1), p.134.

MacNeil, I.A. et al., 2001. Expression and Isolation of Antimicrobial Small Molecules from Soil DNA Libraries. *J. Mol. Microbiol. Biotechnol*, 3(2), pp.301–308.

Maróti, G. et al., 2009. Discovery of [NiFe] hydrogenase genes in metagenomic DNA: cloning and heterologous expression in *Thiocapsa roseopersicina*. *Applied and Environmental Microbiology*, 75(18), pp.5821–30.

Mateo, N., Nader, W. & Tamayo, G., 2001. Bioprospecting. In *Encyclopedia of Biodiversity*. pp. 471–488.

McChesney, J.D., Venkataraman, S.K. & Henri, J.T., 2007. Plant natural products: Back to the future or into extinction? *Phytochemistry*, 68(14), pp.2015–2022.

McCormick, M.H. et al., 1954. Vancomycin, a new antibiotic. I. Chemical and biologic properties. *Antibiotics annual*, 3, pp.606–11.

Medema, M.H. et al., 2011. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Research*, 39(suppl_2), pp.W339–W346.

Mermod, N. et al., 1986. Vector for regulated expression of cloned genes in a wide range of gram-negative bacteria. *Journal of bacteriology*, 167(2), pp.447–54.

Meyer, J.-M. et al., 2002. Siderophore typing, a powerful tool for the identification of fluorescent and nonfluorescent pseudomonads. *Applied and Environmental Microbiology*, 68(6), pp.2745–53.

Miao, V. et al., 2006. Genetic Engineering in *Streptomyces roseosporus* to Produce Hybrid Lipopeptide Antibiotics. *Chemistry & Biology*, 13(3), pp.269–276.

Michelsen, C.F. et al., 2015. Nonribosomal peptides, key biocontrol components for *Pseudomonas fluorescens* In5, isolated from a Greenlandic suppressive soil. *mBio*, 6(2), p.e00079.

Mootz, H.D., Schwarzer, D. & Marahiel, M.A., 2002. Ways of Assembling Complex Natural Products on Modular Nonribosomal Peptide Synthetases. *ChemBioChem*, 3(6), p.490.

Mukherjee, P.K. et al., 2012. Functional analysis of non-ribosomal peptide synthetases (NRPSs) in *Trichoderma virens* reveals a polyketide synthase (PKS)/NRPS hybrid enzyme involved in the induced systemic resistance response in maize. *Microbiology*, 158(1), pp.155–165.

Müller-Esterl, W. et al., 2011. Biochemie: Eine Einführung für Mediziner und Naturwissenschaftler. Spektrum Akademischer Verlag, 2. Auflage, Heidelberg 2010, ISBN 9783827420039.

Müller, C.A. et al., 2015. Mining for Nonribosomal Peptide Synthetase and Polyketide Synthase Genes Revealed a High Level of Diversity in the *Sphagnum* Bog Metagenome. *Applied and Environmental Microbiology*, 81(15), pp.5064–72.

Müller, C.A., Obermeier, M.M. & Berg, G., 2016. Bioprospecting plant-associated microbiomes. *Journal of Biotechnology*, 235, pp.171–180.

Nagarajan, R., 1991. Antibacterial activities and modes of action of vancomycin and related glycopeptides. *Antimicrobial Agents and Chemotherapy*, 35(4), pp.605–9.

NCCLS, 2004. Method for antifungal disk diffusion susceptibility testing of yeasts; approved guideline. NCCLS document M44-A.

Newman, D.J. & Cragg, G.M., 2016. Natural Products as Sources of New Drugs from 1981 to 2014. *Journal of Natural Products*, 79(3), pp.629–661.

Nováková, J. & Farkašovský, M., 2013. Bioprospecting microbial metagenome for natural products. *Biologia (Poland)*, 68(6), pp.1079–1086.

Opelt, K., Berg, C. & Berg, G., 2007. The bryophyte genus *Sphagnum* is a reservoir for powerful and extraordinary antagonists and potentially facultative human pathogens. *FEMS Microbiology Ecology*, 61(1), pp:38-53.

Opelt, K. & Berg, G., 2004. Diversity and Antagonistic Potential of Bacteria Associated with Bryophytes from Nutrient-Poor Habitats of the Baltic Sea Coast. *Applied and*

Environmental Microbiology, 70(11), pp.6569–6579.

Park, C., Nichols, M. & Schrag, S.J., 2014. Two Cases of Invasive Vancomycin-Resistant Group B *Streptococcus* Infection. *New England Journal of Medicine*, 370(9), pp.885–886.

Payne, S.M., 1994. Detection, isolation, and characterization of siderophores. *Methods in Enzymology*, 235(November), pp.329–344.

Pfeifer, B.A. et al., 2001. Biosynthesis of Complex Polyketides in a Metabolically Engineered Stain of *E. coli*. *Science*, 291(2001), pp.1790–1792.

Philippot, L. et al., 2013. Going back to the roots: the microbial ecology of the rhizosphere. *Nature Reviews Microbiolgy*, 11(11), p.789.

Pimentel-Elardo, S.M. et al., 2012. Diversity of nonribosomal peptide synthetase genes in the microbial metagenomes of marine sponges. *Marine drugs*, 10(6), pp.1192–202.

Puk, O. et al., 2004. Biosynthesis of chloro-beta-hydroxytyrosine, a nonproteinogenic amino acid of the peptidic backbone of glycopeptide antibiotics. *Journal of bacteriology*, 186(18), pp.6093–100.

Rabausch, U. et al., 2013. Functional screening of metagenome and genome libraries for detection of novel flavonoid-modifying enzymes. *Applied and Environmental Microbiology*, 79(15), pp.4551–63.

Raymond, K.N., Dertz, E.A. & Kim, S.S., 2003. Enterobactin: an archetype for microbial iron transport. *Proceedings of the National Academy of Sciences of the United States of America*, 100(7), pp.3584–8.

Rondon, M.R. et al., 2000. Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Applied and Environmental Microbiology*, 66(6), pp.2541–7.

Rosano, G.L. & Ceccarelli, E.A., 2014. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Frontiers in Microbiology*, 5, p.172.

Rosconi, F. et al., 2013. Identification and structural characterization of serobactins, a suite of lipopeptide siderophores produced by the grass endophyte *Herbaspirillum seropedicae*. *Environmental Microbiology*, 15(3), pp.916–927.

Röttig, M. et al., 2011. NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Research*, 39(suppl_2), pp.W362–W367.

Roy, A.D. et al., 2010. Gene Expression Enabling Synthetic Diversification of Natural Products: Chemogenetic Generation of Pacidamycin Analogs. *Journal of the American Chemical Society*, 132(35), pp.12243–12245.

Schwyn, B. & Neilands, J.B., 1987. Universal chemical assay for the detection and determination of siderophores. *Analytical Biochemistry*, 160(1), pp.47–56.

Sharma, Y.R., 2012. Resources and Environment for Economic Development. *Academic Voices: A Multidisciplinary Journal*, 2(1), pp.86–89.

Simon, C. et al., 2009. Phylogenetic diversity and metabolic potential revealed in a glacier ice metagenome. *Applied and Environmental Microbiology*, 75(23), pp.7519–26.

Simon, C. & Daniel, R., 2011. Metagenomic analyses: Past and future trends. *Applied and Environmental Microbiology*, 77(4), pp.1153–1161.

Skaar, E.P., 2010. The battle for iron between bacterial pathogens and their vertebrate hosts. *PLoS pathogens*, 6(8), p.e1000949.

Sleator, R.D., Shortall, C. & Hill, C., 2008. Metagenomics. *Letters in Applied Microbiology*, 47(5), pp.361–366.

Stachelhaus, T. et al., 1998. Peptide Bond Formation in Nonribosomal Peptide Biosynthesis: Catalytic Role of the Condensation Domain. *Journal of Biological Chemistry*, 273(35), pp.22773–22781.

Stachelhaus, T., Mootz, H.D. & Marahiel, M.A., 1999. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chemistry & Biology*, 6(8), pp.493–505.

Strieker, M., Tanović, A. & Marahiel, M.A., 2010. Nonribosomal peptide synthetases: structures and dynamics. *Current Opinion in Structural Biology*, 20(2), pp.234–240.

Strobel, G. & Daisy, B., 2003. Bioprospecting for Microbial Endophytes and Their Natural Products Bioprospecting for Microbial Endophytes and Their Natural Products. , 67(4), pp.491–502.

Symeonidis, A. & Marangos, M., 2012. Iron and Microbial Growth. In *Insight and Control of Infectious Disease in Global Scenario*. InTech.

Takehara, M. et al., 2008. Poly(γ-L-diaminobutanoic acid), a novel poly(amino acid), coproduced with poly(ε-L-lysine) by two strains of *Streptomyces celluloflavus*. *FEMS Microbiology Letters*, 286(1), pp.110–117.

Teller, J.H., Powers, S.G. & Snell, E.E., 1976. Ketopantoate hydroxymethyltransferase. I. Purification and role in pantothenate biosynthesis. *The Journal of biological chemistry*, 251(12), pp.3780–5.

Torre, L.A. et al., 2015. Global cancer statistics, 2012. *CA: A Cancer Journal for Clinicians*, 65(2), pp.87–108.

Tripathi, D., Chandra, H. & Bhatnagar, R., 2013. Poly-L-glutamate/glutamine synthesis in the cell wall of *Mycobacterium bovis* is regulated in response to nitrogen availability. *BMC Microbiology*, 13, p.226.

Tuffin, M. et al., 2009. Metagenomic gene discovery: How far have we moved into novel sequence space? *Biotechnology Journal*, 4(12), pp.1671–1683.

Turnbaugh, P.J. et al., 2007. The human microbiome project. *Nature*, 449(7164), pp.804–10.

Uchiyama, T. & Miyazaki, K., 2009. Functional metagenomics for enzyme discovery: challenges to efficient screening. *Current Opinion in Biotechnology*, 20(6), pp.616–622.

Vandenende, C.S., Vlasschaert, M. & Seah, S.Y.K., 2004. Functional Characterization of an Aminotransferase Required for Pyoverdine Siderophore Biosynthesis in *Pseudomonas aeruginosa* PAO1. *Journal of Bacteriology*, 186(17), pp.5596–5602.

Visca, P., Imperi, F. & Lamont, I.L., 2007. Pyoverdine siderophores: from biogenesis to biosignificance. *Trends in Microbiology*, 15(1), pp.22–30.

Voget, S. et al., 2003. Prospecting for novel biocatalysts in a soil metagenome. *Applied and Environmental Microbiology*, 69(10), pp.6235–42.

Vorholt, J.A., 2012. Microbial life in the phyllosphere. *Nature Reviews Microbiology*, 10(12), pp.828–840.

Walsh, C.T., 2008. The chemical versatility of natural-product assembly lines. *Accounts of Chemical Research*, 41(1), pp.4–10.

Wang, G.Y. et al., 2000. Novel natural products from soil DNA libraries in a streptomycete host. *Organic Letters*, 2(16), pp.2401–4.

Wang, H. et al., 2014. Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. *Proceedings of the National Academy of Sciences of the United States of America*, 111(25), pp.9259–64.

Wang, H.-X. et al., 2008. Enriching plant microbiota for a metagenomic library construction. *Environmental Microbiology*, 10(10), pp.2684–91.

Whitman, W.B., Coleman, D.C. & Wiebe, W.J., 1998. Prokaryotes: the unseen majority. *Proceedings of the National Academy of Sciences of the United States of America*, 95(12), pp.6578–6583.

Williams, D.H. & Bardsley, B., 1999. The Vancomycin Group of Antibiotics and the Fight against Resistant Bacteria. *Angewandte Chemie International Edition*, 38(9), pp.1172–1193.

Winn, M. et al., 2016. Recent advances in engineering nonribosomal peptide assembly lines. *Nat. Prod. Rep.*, 33(2), pp.317–347.

Woolhouse, M. & Farrar, J., 2014. Policy: An intergovernmental panel on antimicrobial resistance. *Nature*, 509(7502), pp.555–7.

Wu, S. et al., 2011. WebMGA: a customizable web server for fast metagenomic sequence analysis. *BMC Genomics*, 12(1), p.444.

Wüthrich, K., 1986. NMR with Proteins and Nucleic Acids. *Europhysics News*, 17(1), pp.11–13.

Xu, J., 2006. Microbial ecology in the age of genomics and metagenomics: concepts, tools, and recent advances. *Molecular Ecology*, 15(7), pp.1713–1731.

Xu, Z. et al., 2015. Systematic unravelling of the biosynthesis of poly (L-diaminopropionic acid) in *Streptomyces albulus* PD-1. *Scientific Reports*, 5, p.17400.

Yamanaka, K. et al., 2008. ε-Poly-L-lysine dispersity is controlled by a highly unusual nonribosomal peptide synthetase. *Nature Chemical Biology*, 4(12), pp.766–772.

Yun, J. et al., 2004. Characterization of a novel amylolytic enzyme encoded by a gene from a soil-derived metagenomic library. *Applied and Environmental Microbiology*, 70(12), pp.7229–35.

Zerbe, K. et al., 2004. An Oxidative Phenol Coupling Reaction Catalyzed by OxyB, a Cytochrome P450 from the Vancomycin-Producing Microorganism. *Angewandte Chemie International Edition*, 43(48), pp.6709–6713.

Zhao, K. et al., 2011. The Diversity and Anti-Microbial Activity of Endophytic Actinomycetes Isolated from Medicinal Plants in Panxi Plateau, China. *Current Microbiology*, 62(1), pp.182–190.

Zoonens, M. & Miroux, B., 2010. Expression of Membrane Proteins at the *Escherichia coli* Membrane for Structural Studies. In *Methods in Molecular Biology (Clifton, NJ)*. pp. 49–66

.

List of Tables

Table 1: *Escherichia coli* strains and vector systems used in this project. 14
Table 2: NRPS fosmid clone IDs and the putative peptides encoded by the NRPS sequences. 15
Table 3: Different agar plates recipes, supplements and strains used for the CAS agar plate assay 23
Table 4: Identities of 2-F4 sequences with NRPSs playing a role in antimicrobial and toxin production. 39
Table 5: Retention values (R_F values) of EPI300 and clone 6-H4. 44
Table 6: Identified substances produced by clone 6-H4. 46
Table 7: Unassigned or novel substances produced by 6-H4. 47
Table 8: GO and COG annotations of the 6-H4 genes. 52

List of Figures

Figure 1: Overview of metagenomic library screening. 3
Figure 2: The three strategies of nonribosomal peptide synthesis. 5
Figure 3: Scheme of enzymes involved in enterobactin synthesis. 8
Figure 4: Structure of cyclosporin A. 9
Figure 5: Overview of three different approaches to investigate and characterize NRPS-containing fosmid clones. 13
Figure 6: CAS assay of *ΔentF* and EPI300. 31
Figure 7: Site-directed mutagenesis of the *trfA* gene on pRS48. 32
Figure 8: Verification of *trfA* A250V integration into the *ΔentF* chromosome by colony PCR. 33
Figure 9: Fosmid 2-C8 isolated from *trfA* A250V *ΔentF* after induction with *m*-toluate. 33
Figure 10: Growth of *E. coli* EPI300 and *ΔentF* measured as OD_{600} using iron-deficient MM9 with or without 2 µM $FeSO_4$ supplementation. 34
Figure 11: Siderophore detection assay (CAS assay) of *trfA* A250V *ΔentF* pCC2FOS_NRPS (A) and EPI300 pCC2FOS_NRPS (B) after 48 h of growth. 35
Figure 12: Siderohpore detection on CAS agar plates of clones trfA A250V *ΔentF* pCC2FOS_NRPS and EPI300 pCC2FOS_NRPS. 36
Figure 13: Antimicrobial effect of clone 2-F4. 38
Figure 14: Primer walking of clone 2-F4. 39
Figure 15: TLC analysis of the NRPS-containing fosmid clones. 41
Figure 16: Overlay of the HPLC chromatograms of clone 6-H4 and EPI300. 42
Figure 17: Preparative TLC of EPI300 and clone 6-H4. 43
Figure 18: HPLC chromatograms (positive mode) of clone 6-H4 extracted spots. 45
Figure 19: Blast2Go Pro analysis process. 49
Figure 20: General molecular function and biological process annotations of 6-H4 genes. 50
Figure 21: Annotation of 6-H4 genes using WebMGA/COG database. 51
Figure 22: Cluster analysis of the 6-H4 insert. 53
Figure 23: Generation of 6-H4_Synth expression vector. 54
Figure 24: SDS PAGE analysis of the heterologously overexpressed 6-H4 synthetase. 55
Figure 25: Transmembrane domain analysis of the 6-H4 synthetase. 56
Figure 26: Domain structure comparison of the 6-H4 synthetase with homopoly(amino acid) synthesizing NRPSs. 58
Figure 27: Phylogenetic analysis of the 6-H4 synthetase. 59
Figure 28: Model of the 6-H4 synthetase. 60
Figure 29: Pairwise comparison of the homopoly(amino acid) NRPSs and 6-H4 synthetase soluble domains. 60

List of Abbreviations

Amp	Ampicillin
APS	Ammonium persulfate
BLAST	Basic Local Alignment Search Tool
bp	Base pair
°C	Degree centigrade
$CaCl_2$	Calcium chloride
CAS	Chromazurol S
Cm	Chloramphenicol
COG	Clusters of Orthologous Groups
DNA	Deoxyribonucleic acid
e.g.	*Exempli gratia* (for example)
et al.	*Et alii* (and others)
Etc.	Etcetera
$FeCl_3$	Iron trichloride
$FeSO_4$	Iron sulfate
GO	Gene ontology
h	Hour
H_2O	Water
HCl	Hydrochloride
HPLC	High-performance liquid chromatography
IPTG	Isopropyl-β-D-thiogalactopyranosid
K_2HPO_4	Dipotassium phosphate
Kan	Kanamycin
kb	Kilo base
kDa	Kilo dalton
KH_2PO_4	Potassium dihydrogen phosphate
LB	Luria-Bertani
l	Liter
µg	Microgram
µl	Microliter

µM	Micromolar
M	Molar
min	Minute
$MgCl_2$	Magnesium chloride
mM	Millimolar
MM9	Minimal Medium 9
ml	Milliliter
MS	Mass spectrometry
NaCl	Sodium chloride
NaOH	Sodium hydroxide
Na_2HPO_4	Disodium hydrogen phosphate
NCBI	National Centre for Biotechnology Information
NH_4Cl	Ammonium chloride
nm	Nanometer
NMR	Nuclear magnetic resonance
NRP(S)	Nonribosomal peptide (synthetase)
ONC	Overnight culture
ORF	Open reading frame
PCR	Polymerase chain reaction
PIPES	Piperazine-N, N'-bis(2-ethanesulfonic acid)
%	Percentage
rpm	Round per minute
sec	Second
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
TAE	Tris-Acetate-EDTA
TB	Terrific Broth
Tc	Tetracycline
TEMED	N, N, N', N'-tetramethylethylenediamine
TLC	Thin-layer chromatography
Tris	Tris-hydroxymethyl aminomethane
UV	Ultraviolet
V	Volt

Attachment

Table A1: Primers for gap-closing sequencing. *Primers were designed using CLC Main Workbench. rev = reverse. fw = forward.*

	Primer name	Sequence
First gap-closing sequencing	D1 rev	5‘-TTCATAACCTTGACGCCCAC-3‘
	H11 rev	5‘-CTCGTCCTACATCCATCC-3‘
	E10 rev	5‘-CTTCTCTCTGTCTACGTGTC-3‘
	C12 rev	5‘-GCTGGTGATGTAGAAGTTGA-3‘
	H9 rev	5‘-CCTTTTTAACGATCTCTCCCA-3‘
	Contig 3 fw	5‘-GAATTGAGGGTGTAGTGGG-3‘
	B5 rev	5‘-CGATACCTACCCCTCACA-3‘
	F5 fw	5‘-AGATAGGTTCCAGCCGCA-3‘
	A8 rev	5‘-CAGACCGTAGCCTCCAT-3‘
	Contig 5 fw	5‘-TGAAGGTCGGAAACAGGG-3‘
	H1 rev	5‘-GTTTAGCGTGGGGATCTG-3‘
	Contig 7 fw	5‘-TTTGGCGAGCGTGTTCAT-3‘
	D8 rev fw	5‘-GAAGGTTTGCATGGGGAA-3‘
	G9 fw	5‘-TCCAGAAAGAAATCCGCC-3‘
	C4 rev	5‘-CGGGCGGCTATTTCTACT-3‘
	C4 fw	5‘-TCCTCCACGTCATAGCC-3‘
Second gap-closing sequencing	Contig 1-3 fw	5‘-GCTGACCTACACCGAACT-3‘
	Contig 1-3 rev	5‘-GCTTTCCCGTTGCGTTAT-3‘
	Contig 3-3 fw	5‘-CTTTTCATCCTGCGCAT-3‘
	Contig 3-3 rev	5‘-CCGGTATAGGTCTGGTC-3‘
	Contig 4-3 rev	5‘-GGTGGGCTTGAGGAAATAGG-3‘
	Contig 5-3 fw	5‘-GAAAAAATCGGTGGTGGG-3‘
	Contig 5-3 rev	5‘-CAGACCCCACTTCATAGC-3‘
	Contig 6-3 fw	5‘-GAAGAGGTTGCTGGCGA-3‘
	Contig 6-3 rev	5‘-TCCTTGATCCCGACCTTT-3‘
Third gap-closing sequencing	Contig_1	5‘-ATACCCTTAGCCCGACAG-3‘
	Contig_2	5‘-CCATAGTTCAGTGTCAGGT-3‘

Table A2: Buffer solutions and gel compositions for SDS-PAGE analysis.

Laemmli Loading Dye (4x)	**Electrophoresis buffer**	**Coomasie staining**
2.5 ml 1 M Tris-HCl (pH 6.8) 4 g Glycerol 0.4 g SDS 1 ml ß-Mercaptoethanol 0.002 g Bromphenol blue adjust to 10 ml with ddH_2O	144 g Glycin 30 g Tris-Base 10 g SDS adjust to 1 l ddH2O (pH)	0.25% (v/v) Coomasie Brilliant Blue R-250 45.5% (v/v) Methanol 9.2% (v/v) Acetic acid filtered through Whatmann 3MM-paper

Running Gel 7.5%		**Running Gel 10%**		**Stacking Gel 5%**	
ddH_2O	1780 µl	ddH_2O	1364 µl	ddH_2O	2770 µl
Polyacrylamide (30%)	1250 µl	Polyacrylamide (30%)	1667 µl	Polyacrylamide (30%)	665 µl
1 M Tris pH 8.8	1875 µl	1 M Tris pH 8.8	1875 µl	1 M Tris pH 6.8	500 µl
SDS 10%	50 µl	SDS 10%	50 µl	SDS 10%	40 µl
APS 10%	40 µl	APS 10%	40 µl	APS 10%	20 µl
TEMED	4 µl	TEMED	4 µl	TEMED	4 µl

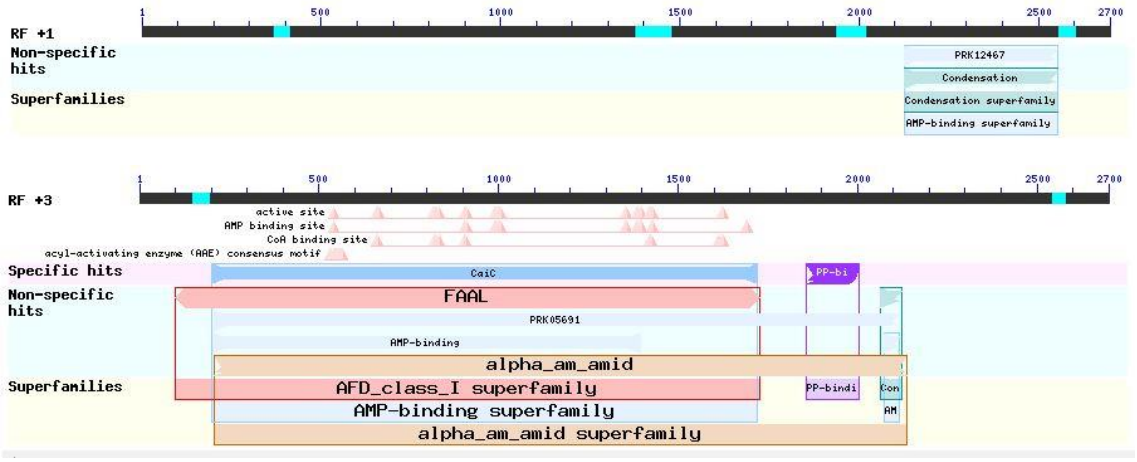


Figure A1: Conserved domains of the 2697 bp sequence of clone 2-F4. *Reading frame (RF) +3 reveals an A-domain with similarities to an acyl-CoA synthetase (CaiC) or a fatty acyl-AMP ligase (FAAL) followed by a phospho-pantetheine attachment site. RF +1 revealed parts of a condensation domain at the C-terminal end of the sequence.*

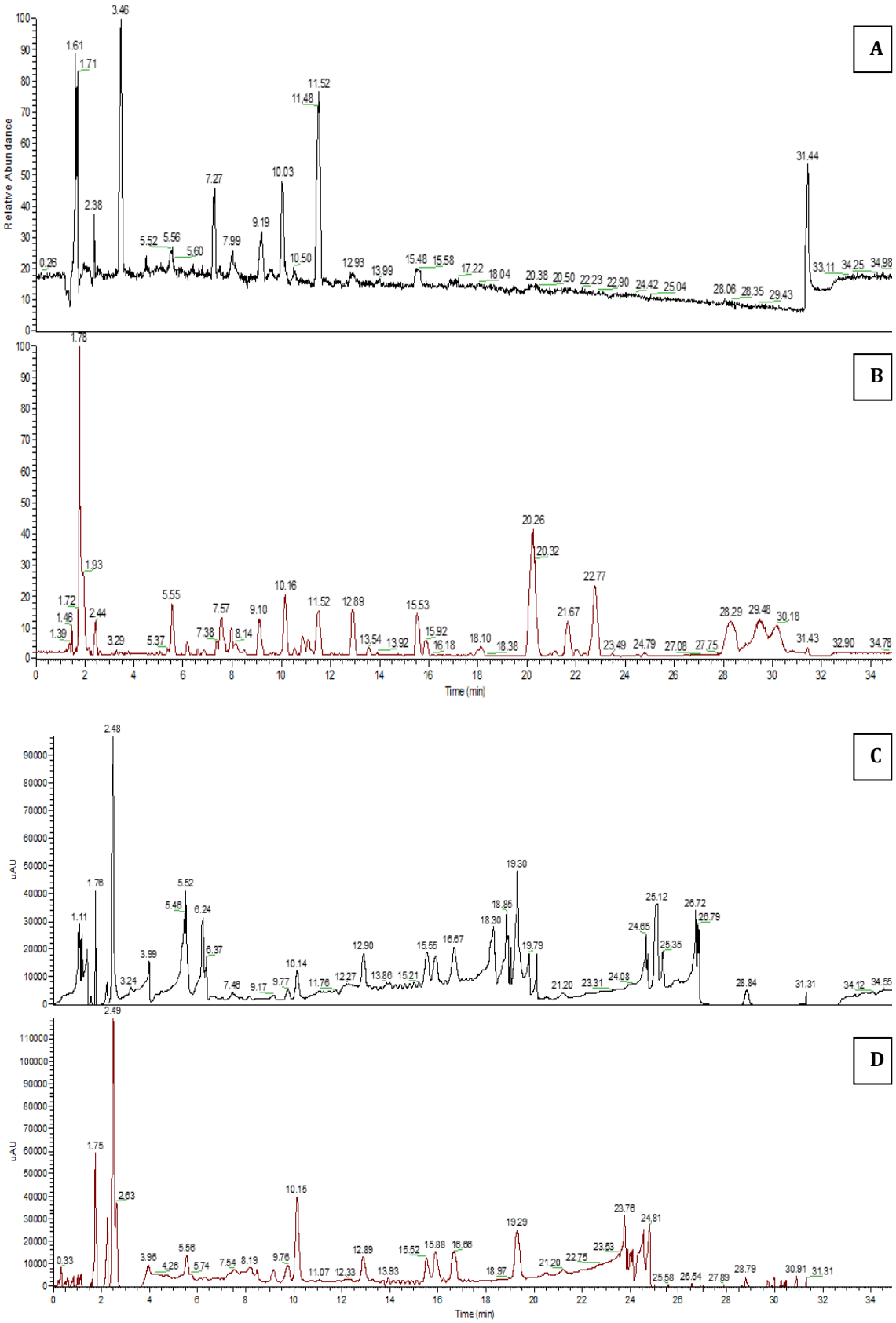


Figure A2: HPLC chromatograms of 6-H4 and EPI300 whole extracts. *EPI300 measured in positive mode (A) and negative mode (C). 6-H4 measured in positive mode (B) and negative mode (D).*

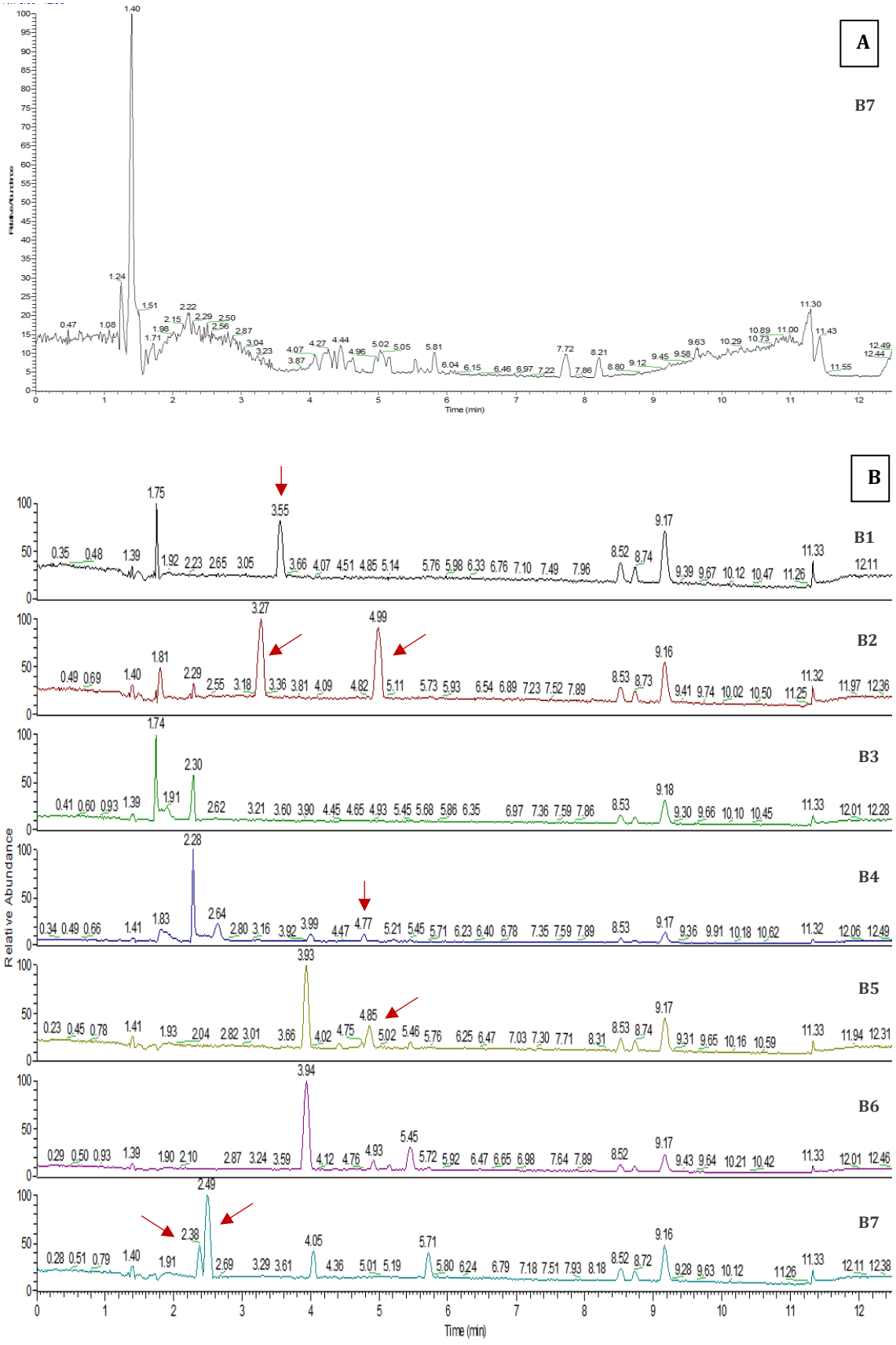


Figure A3: HPLC chromatograms of clone 6-H4 extracted spots (normal phase silica plates with ethyl acetate:acetic acid:water as solvent system). *(A) Spot B7 measured in positive mode. (B) Spots B1 to B7 measured in negative mode. The red arrows indicate peaks unique or strongly upregulated in comparison to EPI300. X-axis: time (min), y-axis: relative abundance (-).*

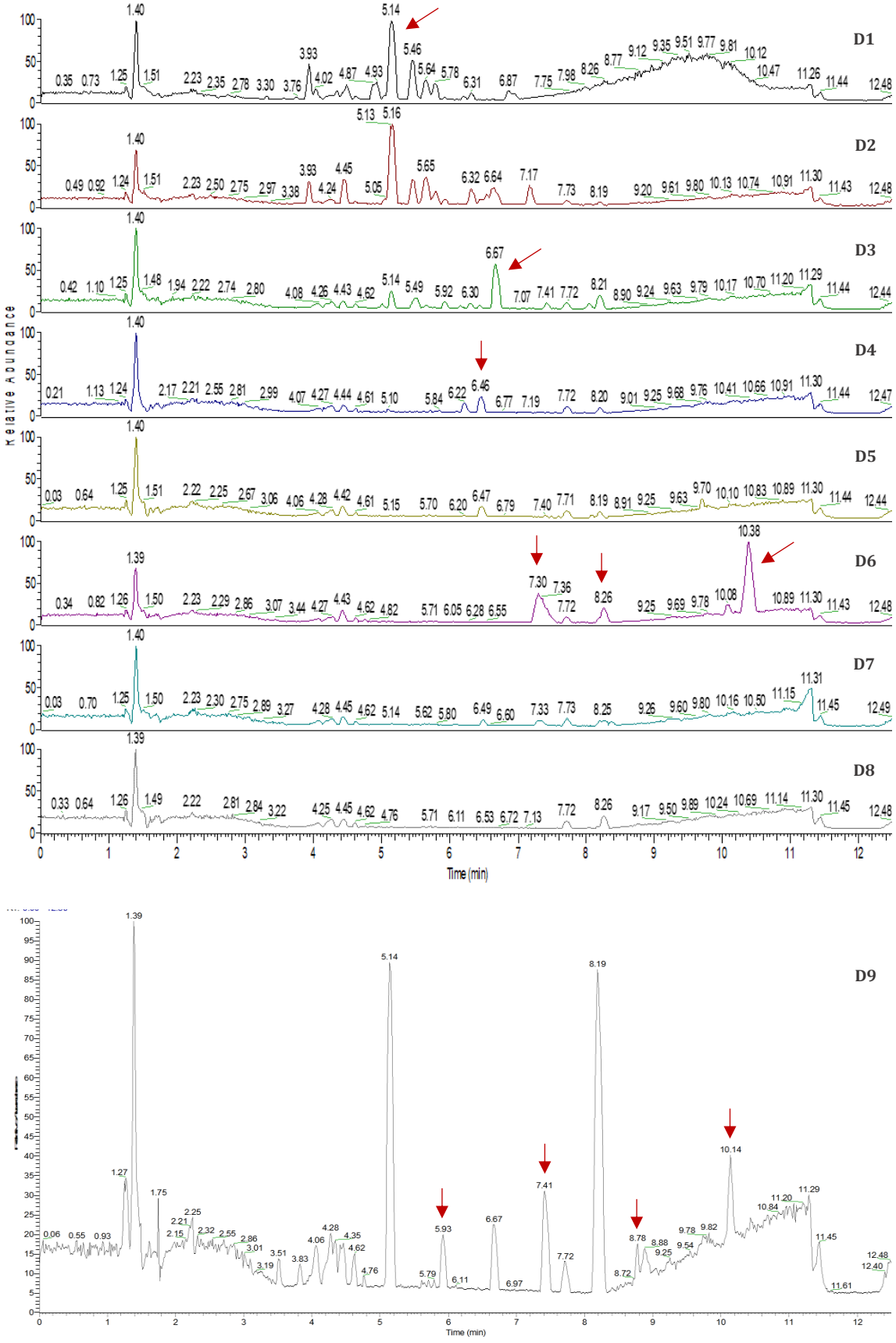


Figure A4: HPLC chromatogram of 6-H4 extracted spots (normal phase silica plates with ethyl acetate as solvent system, positive mode). *Spots from top to bottom: D1 to D9. The red arrows indicate peaks unique or strongly upregulated in comparison to EPI300. X-axis: time (min), y-axis: relative abundance (-).*

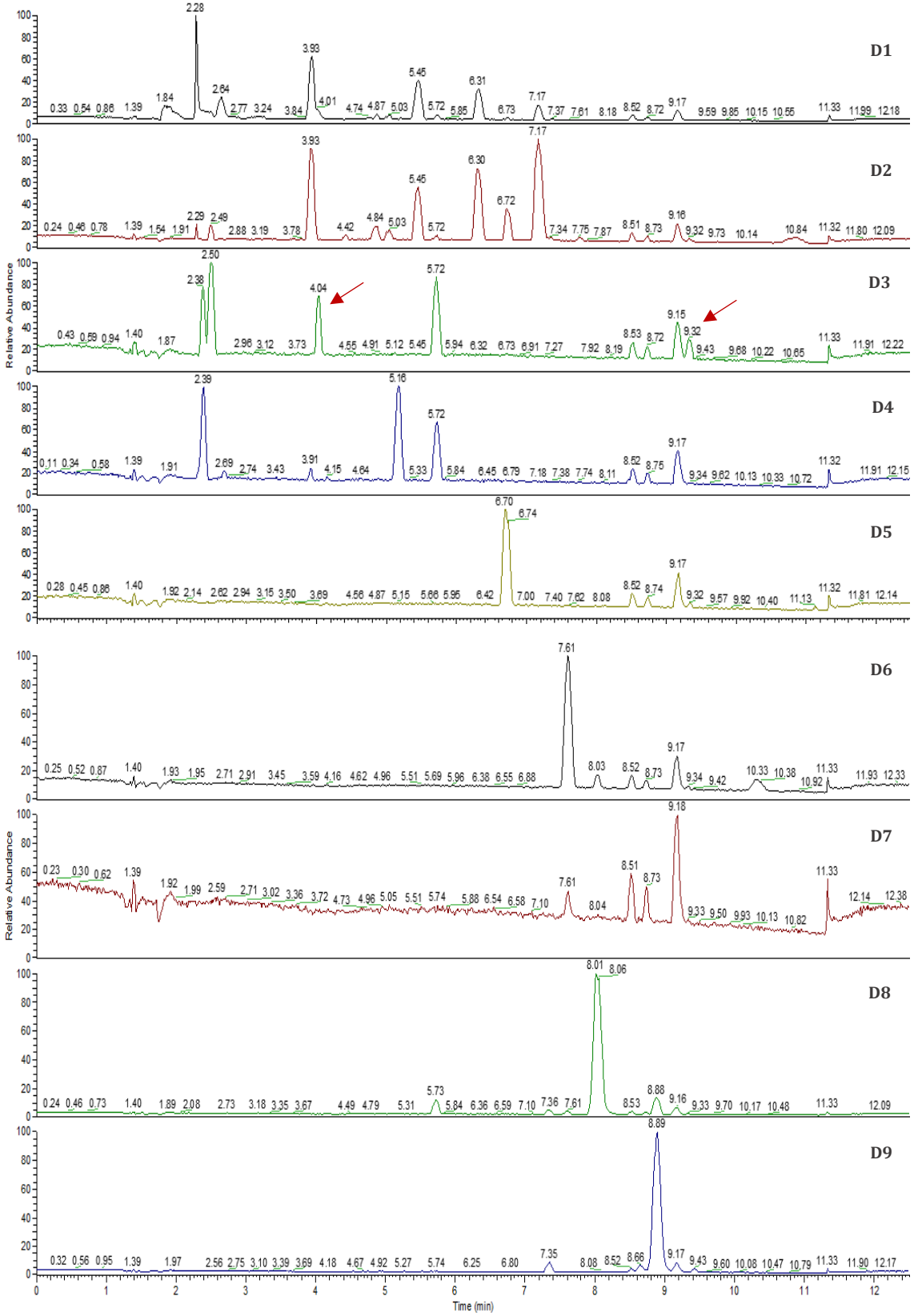


Figure A5: HPLC chromatograms of 6-H4 extracted spots (normal phase silica plates with ethyl acetate as solvent system, negative mode). *Spots D1 to D9. The red arrows indicate peaks unique or strongly upregulated in comparison to EPI300. X-axis: time (min), y-axis: relative abundance (-).*

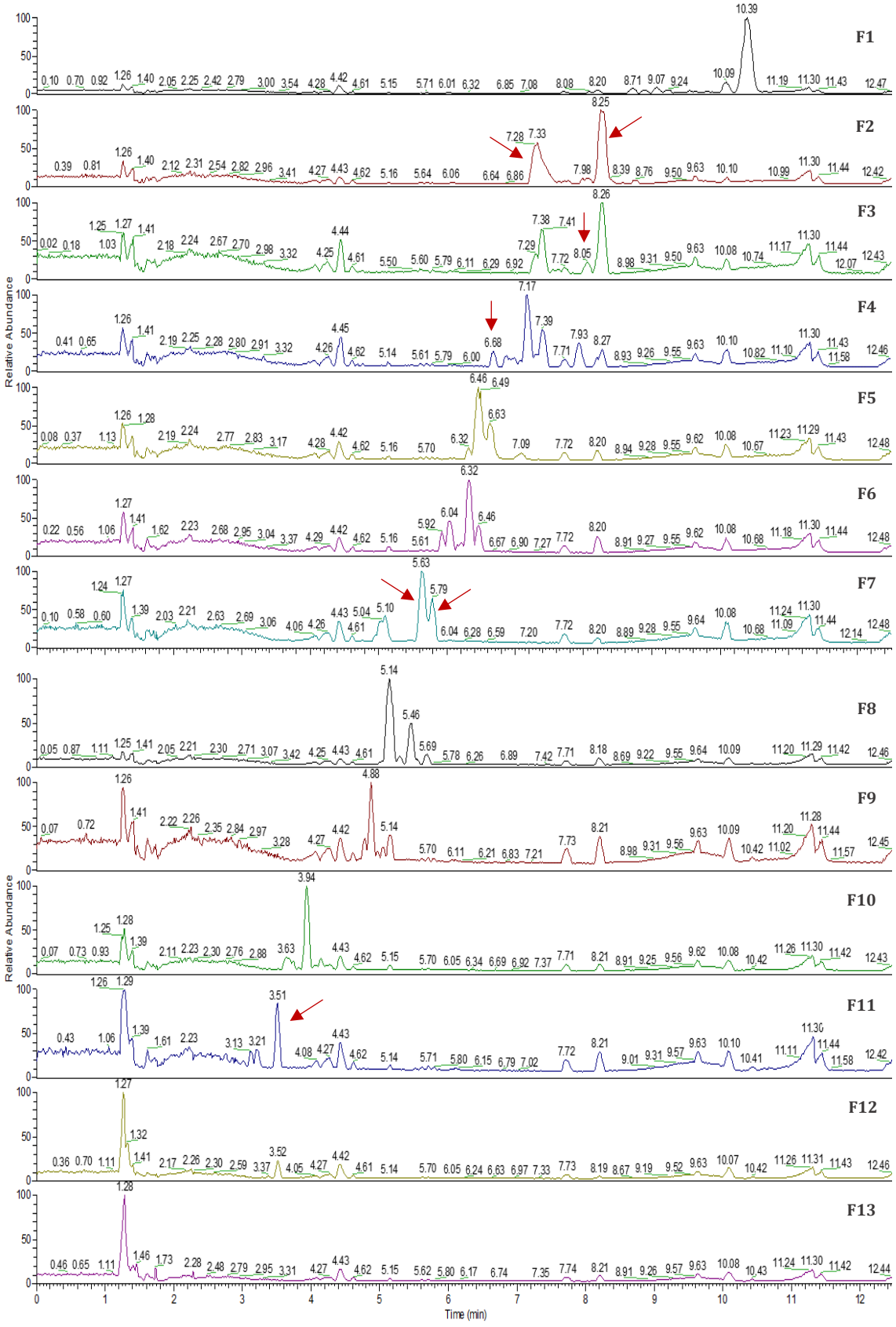


Figure A6: HPLC chromatograms of 6-H4 extracted spots (reverse phase silica plates with acetonitrile:water:formic acid as solvent system, positive mode). *Spots F1 to F13. The red arrows indicate peaks unique or strongly upregulated in comparison to EPI300. X-axis: time (min), y-axis: relative abundance (-).*

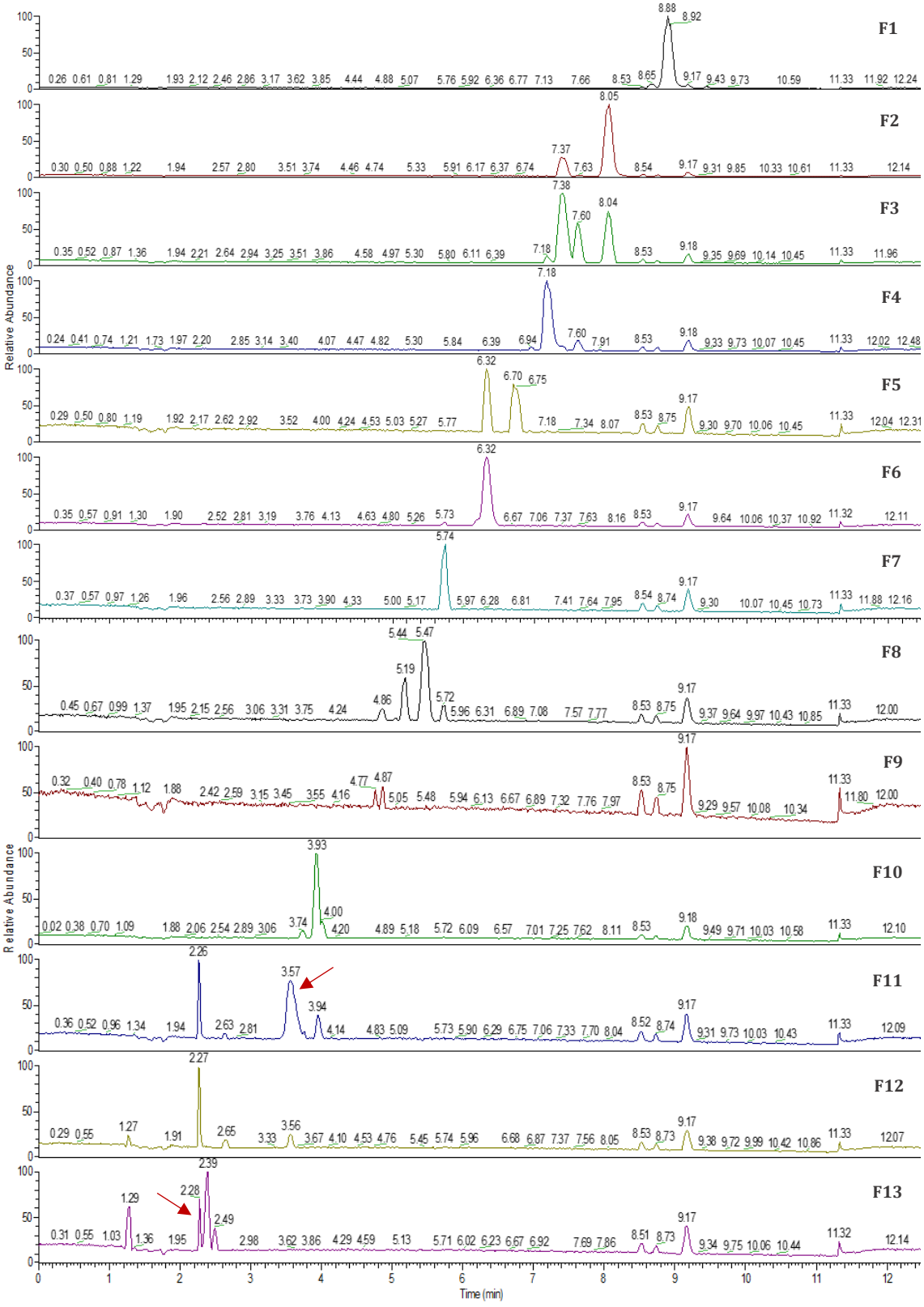


Figure A7: HPLC chromatograms of 6-H4 extracted spots (reverse phase silica plates with acetonitrile:water:formic acid as solvent system, negative mode). *Spots F1 to F13. The red arrows indicate peaks unique or strongly upregulated in comparison to EPI300. X-axis: time (min), y-axis: relative abundance (-).*

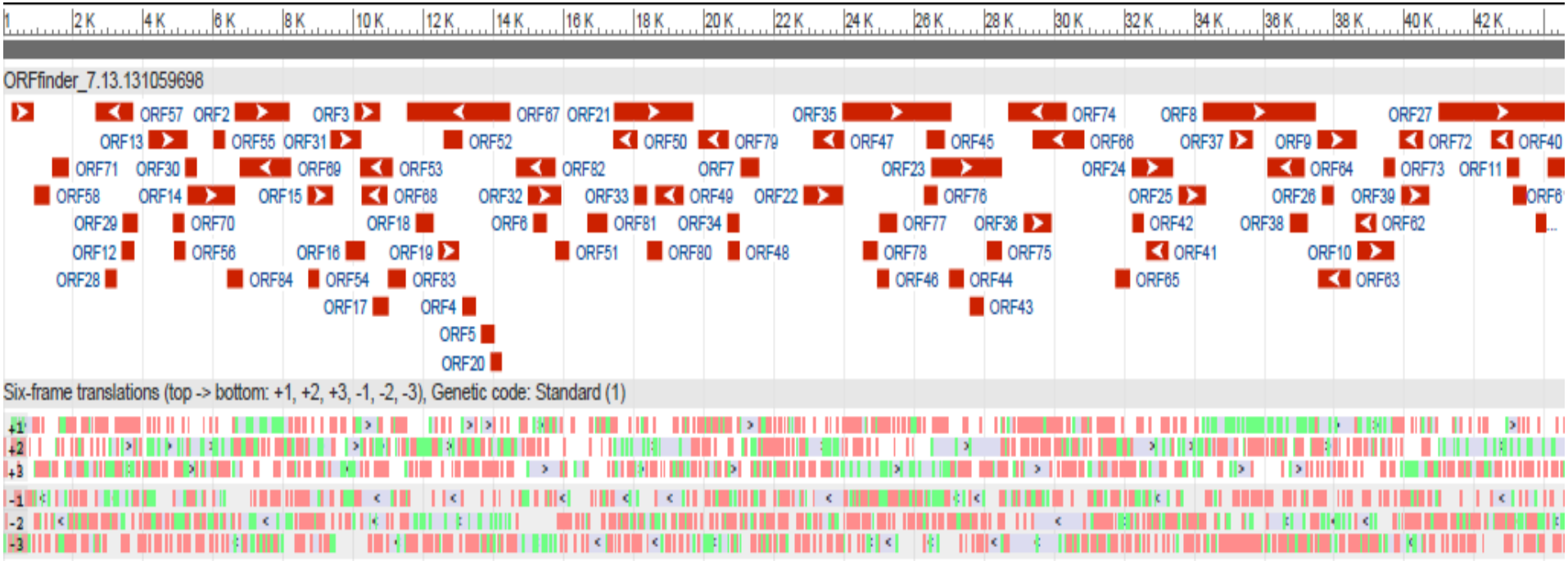


Figure A8: The 84 putative ORFs of clone 6-H4. *The insert sequence of NRPS clone 6-H4 was analyzed for putative ORFs using NCBI ORFfinder with following settings: minimal length of 300 nucleotides, standard genetic code and "ATG" only start codon. The figure shows the positions of the ORFs in the 44.7 kb insert fragment of 6-H4 and also displays the six-frame translations.*

Table A3: Blastp analysis of the 6-H4 ORFs. *For the Blast2Go Pro NCBI blast, the Blast expectation value threshold was set to 10 and the word size to 6. Blastp was performed against the non-redundant database, with the Blast description annotator turned on and the low complexity filter turned off. Length =amino acids. Hits = number of blast hits. Similarity mean of the blast hits in %. GO categories: F = molecular function, P = biological process, C = cellular component.*

Sequence Name	Description (Blast description annotator)	Length	#Hits	e-Value	Similarity Mean	GO Names	InterPro GO Names
ORF 1	peptidase M16	206	20	8.00E-39	65.75	F: catalytic activity	F: metal ion binding; F: catalytic activity
ORF 2	glycine dehydrogenase (aminomethyl-transferring)	519	20	0E0	85.0	P: glycine decarboxylation via glycine cleavage system; F: glycine dehydrogenase (decarboxylating) activity; P: oxidation-reduction process	F: catalytic activity; F: glycine dehydrogenase (decarboxylating) activity; P: oxidation-reduction process; P: glycine catabolic process
ORF 3	hypothetical protein UVI_02032900	241	1	9.77E0	50.0		no GO terms
ORF 4	hypothetical protein BN136_2946	126	4	1.25E-2	50.25	P: proteolysis; C: membrane; C: integral component of membrane; F: serine-type carboxypeptidase activity	no GO terms
ORF 5	FKBP12-rapamycin complex-associated	126	3	3.22E-1	55.33	F: nucleotide binding; F: ATP binding; P: phosphorylation; F: transferase activity; F: protein serine/threonine kinase activity; P: protein phosphorylation; F: kinase activity; F: macromolecular complex binding	no IPS match
ORF 6	---NA---	126					no IPS match
ORF 7	---NA---	176					no GO terms
ORF 8	GMC family oxidoreductase	1072	20	0E0	77.8	F: oxidoreductase activity, acting on CH-OH group of donors; F: flavin adenine dinucleotide binding; P: oxidation-reduction process	F: oxidoreductase activity, acting on CH-OH group of donors; F: oxidoreductase activity; F: flavin adenine dinucleotide binding; F: electron carrier activity; P: oxidation-reduction process; F: heme binding

ORF 9	DNA topoisomerase	366	20	8.15E-159	73.7	F: DNA topoisomerase activity	F: DNA binding; F: DNA topoisomerase type I activity; P: DNA topological change
ORF 10	aldo keto reductase	346	20	0E0	81.6	F: oxidoreductase activity	no GO terms
ORF 11	Hypothetical protein CBG15343	116	1	4.09E0	38.0		no GO terms
ORF 12	cell division	121	2	1.11E0	50.0		no IPS match
ORF 13	glycine cleavage system T	365	20	8.74E-158	74.25	P: metabolic process; F: transferase activity	F: aminomethyltransferase activity; P: glycine catabolic process
ORF 14	glycine dehydrogenase (aminomethyl-transferring)	451	20	0E0	80.2	P: small molecule metabolic process; F: oxidoreductase activity; P: nitrogen compound metabolic process	F: catalytic activity; P: nucleoside metabolic process; F: glycine dehydrogenase (decarboxylating) activity; P: oxidation-reduction process; P: glycine catabolic process
ORF 15	hypothetical protein ABS78_17995	240	20	2.87E-11	55.05		no GO terms
ORF 16	---NA---	180					no GO terms
ORF 17	---NA---	149					no GO terms
ORF 18	Uncharacterized	167	8	2.91E-4	53.25		no GO terms
ORF 19	hypothetical protein BGP14_22300	193	1	7.97E0	49.0		no GO terms
ORF 20	NUDIX domain-containing	104	3	3.22E-4	49.0	F: hydrolase activity	no IPS match
ORF 21	dTDP-4-dehydrorhamnose reductase	750	20	0E0	64.1	P: carbohydrate metabolic process; F: hydrolase activity, hydrolyzing O-glycosyl compounds	P: carbohydrate metabolic process; F: hydrolase activity, hydrolyzing O-glycosyl compounds
ORF 22	efflux RND transporter periplasmic adaptor subunit	372	20	7.13E-109	62.9	P: transmembrane transport; C: membrane	P: transmembrane transport; C: membrane; F: transporter activity
ORF 23	histidine kinase	671	20	1.02E-148	67.6	C: membrane	C: membrane; F: lipid binding; P: transport; F: transporter activity
ORF 24	LPPG:FO 2-phospho-L-lactate transferase	390	1	9.48E0	46.0		no GO terms

ORF 25	twin-arginine translocation pathway signal	256	20	6.41E-81	62.9		no GO terms
ORF 26	---NA---	110					no IPS match
ORF 27	peptide synthetase	1218	20	0E0	73.15	F: catalytic activity; C: membrane	F: catalytic activity; P: metabolic process
ORF 28	---NA---	108					no IPS match
ORF 29	Uncharacterized protein	139	10	3.06E-4	53.3		no GO terms
ORF 30	---NA---	109					no GO terms
ORF 31	hypothetical protein	292	20	6.75E-33	61.9		no GO terms
ORF 32	---NA---	315					no GO terms
ORF 33	---NA---	121					no GO terms
ORF 34	hypothetical protein ASD47_18895	109	2	6.39E-3	75.0		no GO terms
ORF 35	family heavy metal efflux RND transporter	1032	20	0E0	79.45	C: intrinsic component of membrane; F: transporter activity	P: cation transport; C: membrane; F: cation transmembrane transporter activity; C: integral component of membrane; P: transport; F: transporter activity
ORF 36	ATPase	261	1	8.89E0	46.0		no IPS match
ORF 37	hypothetical protein	214	2	2.33E-1	37.5	F: calcium ion binding; C: membrane; P: homophilic cell adhesion via plasma membrane adhesion molecules	no GO terms
ORF 38	---NA---	165					no GO terms
ORF 39	3-methyl-2-oxobutanoate hydroxymethyltransferase	263	20	6.03E-103	79.15	P: small molecule metabolic process; C: cytoplasm; F: metal ion binding; P: nitrogen compound metabolic process; P: biosynthetic process; F: transferase activity	P: pantothenate biosynthetic process; F: 3-methyl-2-oxobutanoate hydroxymethyltransferase activity; F: catalytic activity
ORF 40	---NA---	206					no GO terms
ORF 41	23S rRNA (adenine(2030)-N(6))-methyltransferase	210	20	4.86E-51	59.75	F: transferase activity	F: rRNA methyltransferase activity; P: rRNA base methylation

ORF 42	alpha beta hydrolase	104	6	9.87E-1	46.67	F: catalytic activity; F: transferase activity; F: hydrolase activity	no GO terms
ORF 43	---NA---	127					no IPS match
ORF 44	DNA repair and recombination RAD54B isoform X3	135	3	9,39E-1	59.0		no GO terms
ORF 45	hypothetical protein BN961_01950	170	4	3.23E-1	46.5	P: signal transduction by protein phosphorylation; F: protein histidine kinase activity; F: signal transducer activity; F: nucleotide binding; F: phosphorelay sensor kinase activity; P: chemotaxis; F: ATP binding; C: cytoplasm; P: phosphorelay signal transduction system; F: transferase activity, transferring phosphorus-containing groups; P: phosphorylation; F: transferase activity; F: kinase activity; P: peptidyl-histidine phosphorylation	no IPS match
ORF 46	hypothetical protein GY15_08565	112	3	1.53E-2	52.0	C: membrane; C: integral component of membrane	no GO terms
ORF 47	---NA---	290					no IPS match
ORF 48	transposase IS3 IS911 family	105	20	2.93E-51	83.8	F: nucleic acid binding; F: transposase activity; P: DNA metabolic process	F: RNA polymerase binding; F: DNA binding; P: regulation of DNA-templated transcription, elongation
ORF 49	lactate dehydrogenase	265	1	8.92E0	38.0		no IPS match
ORF 50	non-ribosomal peptide synthetase	230	2	2.44E0	63.0		no IPS match
ORF 51	TolC family	132	20	2.05E-31	72.2	C: membrane	no GO terms
ORF 52	---NA---	177					no GO terms
ORF 53	---NA---	309					no GO terms
ORF 54	hypothetical protein	101	4	3.01E0	51.75	C: membrane; C: integral component of membrane	

ORF 55	Uncharacterized	108	3	2.78E-3	59.67		no GO terms
ORF 56	hypothetical protein	106	2	2.61E0	45.5		no GO terms
ORF 57	short-chain dehydrogenase	355	20	1.24E-119	68.25	F: oxidoreductase activity; C: membrane; C: integral component of membrane; P: oxidation-reduction process	F: oxidoreductase activity
ORF 58	hypothetical protein	142	20	2.32E-43	68.9		no IPS match
ORF 59	Zn-ribbon-like motif-containing	162	19	2.13E-13	50.16		no GO terms
ORF 60	PA14 domain	108	2	3.38E0	48.0	F: calcium ion binding; C: membrane; P: homophilic cell adhesion via plasma membrane adhesion molecules	no GO terms
ORF 61	---NA---	132					no GO terms
ORF 62	2,3-bisphosphoglycerate-independent phosphoglycerate mutase	192	12	2.79E-4	48.92	C: cytoplasm; F: metal ion binding; F: catalytic activity; P: metabolic process; P: glycolytic process; F: phosphoglycerate mutase activity; F: manganese ion binding; P: glucose catabolic process; F: 2,3-bisphosphoglycerate-independent phosphoglycerate mutase activity; F: isomerase activity	no IPS match
ORF 63	hypothetical protein SP67_30670	302	4	1.44E-13	49.75	C: membrane; C: integral component of membrane	no GO terms
ORF 64	hypothetical protein AYR66_12630	349	4	1.83E-6	47.0		no GO terms
ORF 65	DUF4112 domain-containing	137	20	7.42E-28	60.15	C: membrane; C: integral component of membrane	no GO terms
ORF 66	TonB-dependent receptor	490	20	7.10E-93	76.5	C: membrane; F: receptor activity; P: transport	
ORF 67	multidrug transporter	978	20	0E0	75.9	C: membrane	
ORF 68	short-chain dehydrogenase	242	20	3.21E-82	71.3	F: oxidoreductase activity; P: oxidation-reduction process	no GO terms
ORF 69	metal-dependent RNase	487	18	2.16E-34	49.28		no IPS match
ORF 70	---NA---	106					no GO terms
ORF 71	---NA---	156					no GO terms

ORF 72	cell division	224	4	2.92E-2	43.25	P: cell division; F: ligase activity	no GO terms
ORF 73	---NA---	105					no IPS match
ORF 74	TonB-dependent receptor	554	20	9.05E-169	68.05	C: membrane; C: cell outer membrane; F: receptor activity; P: transport	C: membrane; F: receptor activity; P: transport
ORF 75	---NA---	140					no GO terms
ORF 76	DUF898 domain-containing [Devosia enhydra]	125	2	1.01E0	52.5		no GO terms
ORF 77	hypothetical protein Barb4_02461	169	6	1.27E-2	45.67		no GO terms
ORF 78	hypothetical protein Barb4_02461	136	4	1.31E-2	49.0		no GO terms
ORF 79	Integrase catalytic region	286	20	0E0	90.45	F: nucleic acid binding; P: DNA metabolic process	
ORF 80	---NA---	139					no GO terms
ORF 81	TolC family	188	20	3.85E-15	49.35	C: membrane; C: integral component of membrane; F: lipid binding; P: transport; F: transporter activity; C: plasma membrane	no GO terms
ORF 82	efflux RND transporter periplasmic adaptor subunit	371	20	1.16E-107	67.95	P: transmembrane transport; C: membrane	P: transmembrane transport; C: membrane; F: transporter activity
ORF 83	transcription elongation factor	166	20	2.00E-66	81.0	F: nucleic acid binding; F: protein binding; P: transcription, DNA-templated; P: regulation of metabolic process; P: translation	
ORF 84	---NA---	146					no IPS match

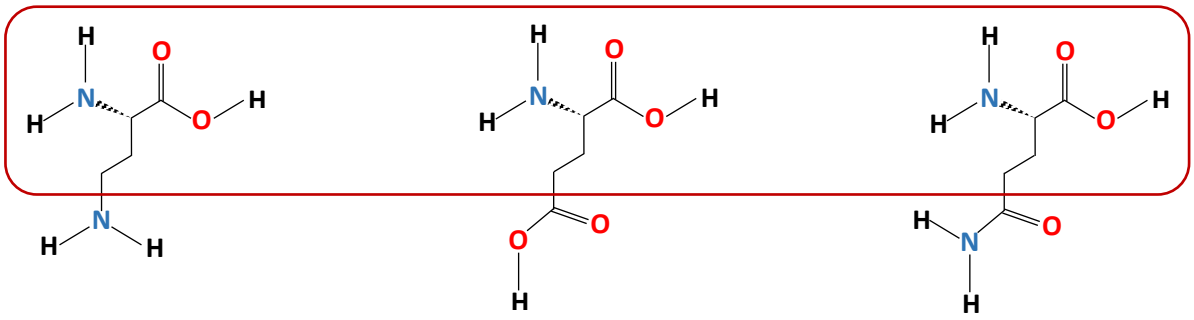
Attachment Table 4: WebMGA RPSBLAST results against the COG database.

#Query	Hit	E-value	Identity	Score	Hit-length	Description	Class	Class description
ORF 1	COG0612	1.00E-6	20	46.7	438	Predicted Zn-dependent peptidases	R	General function prediction only
ORF 2	COG1003	0E0	47	637	496	Glycine cleavage system protein P (pyridoxal-binding), C-terminal domain	E	Amino acid transport and metabolism
ORF 8	COG2010	2.00E-11	30	65.4	150	Cytochrome c, mono- and diheme variants	C	Energy production and conversion
ORF 8	COG2303	2.00E-53	22	204	542	Choline dehydrogenase and related flavoproteins	E	Amino acid transport and metabolism
ORF 9	COG3569	1.00E-93	47	336	354	Topoisomerase IB	L	Replication, recombination and repair
ORF 10	COG0667	4.00E-98	46	351	316	Predicted oxidoreductases (related to aryl-alcohol dehydrogenases)	C	Energy production and conversion
ORF 13	COG0404	1.00E-89	41	323	379	Glycine cleavage system T protein (aminomethyltransferase)	E	Amino acid transport and metabolism
ORF 14	COG0403	1.00E-142	40	499	450	Glycine cleavage system protein P (pyridoxal-binding), N-terminal domain	E	Amino acid transport and metabolism
ORF 21	COG1091	2.00E-47	33	184	281	dTDP-4-dehydrorhamnose reductase	M	Cell wall/membrane/envelope biogenesis
ORF 21	COG2723	2.00E-15	22	78.1	460	Beta-glucosidase/6-phospho-beta-glucosidase/beta-galactosidase	G	Carbohydrate transport and metabolism
ORF 22	COG0845	1.00E-21	25	97.9	372	Membrane-fusion protein	M	Cell wall/membrane/envelope biogenesis
ORF 23	COG1538	1.00E-69	34	258	457	Outer membrane protein	MU	Multiple classes
ORF 27	COG1020	3.00E-86	38	314	642	Non-ribosomal peptide synthetase modules and related proteins	Q	Secondary metabolites biosynthesis, transport and catabolism
ORF 27	COG0663	3.00E-5	28	45.3	176	Carbonic anhydrases/acetyltransferases, isoleucine patch superfamily	R	General function prediction only
ORF 35	COG3696	0E0	40	1006	1027	Putative silver efflux pump	P	Inorganic ion transport and metabolism
ORF 39	COG0413	1.00E-97	59	349	268	Ketopantoate	H	Coenzyme transport and metabolism

						hydroxymethyltransferase		
ORF 41	COG2961	2.00E-44	35	172	279	Protein involved in catabolism of external DNA	R	General function prediction only
ORF 48	COG2963	5.00E-5	31	39.5	116	Transposase and inactivated derivatives	L	Replication, recombination and repair
ORF 51	COG1538	2.00E-15	30	74.4	457	Outer membrane protein	MU	Multiple classes
ORF 57	COG0300	4.00E-38	30	152	265	Short-chain dehydrogenases of various substrate specificities	R	General function prediction only
ORF 66	COG1629	2.00E-8	18	54.5	768	Outer membrane receptor proteins, mostly Fe transport	P	Inorganic ion transport and metabolism
ORF 67	COG0841	0E0	32	788	1009	Cation/multidrug efflux pump	V	Defense mechanisms
ORF 68	COG1028	1.00E-15	28	77.2	251	Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)	IQR	Multiple classes
ORF 74	COG1629	9.00E-13	17	69.1	768	Outer membrane receptor proteins, mostly Fe transport	P	Inorganic ion transport and metabolism
ORF 79	COG2801	2.00E-23	34	102	232	Transposase and inactivated derivatives	L	Replication, recombination and repair
ORF 81	COG1538	9.00E-9	26	53.6	457	Outer membrane protein	MU	Multiple classes
ORF 82	COG0845	1.00E-23	27	104	372	Membrane-fusion protein	M	Cell wall/membrane/envelope biogenesis
ORF 83	COG0782	4.00E-23	33	101	151	Transcription elongation factor	K	Transcription

Table A5: Molecular function annotation. *GO annotation was performed with Blast2Go Pro. The GO level was set to 1 and intermediate GO terms were filtered.*

Blast2Go Pro Molecular Function Annotation	
GO-Terms	**ORFs**
catalytic activity	1, 2, 8, 9, 10, 13, 14, 41, 21, 27, 39, 48, 57
transporter activity	22, 23, 35, 51, 67, 82
oxidoreductase activity	2, 8, 10, 14, 57
nucleic acid binding	9, 48, 73, 83
transferase activity	13, 39, 41
receptor activity	66, 74
ion binding	1, 8
coenzyme binding	8
tetrapyrrole binding	8
metal ion binding	1
transposase activity	48
electron carrier activity	8
protein binding	83
hydrolase activity	21
isomerase activity	9
nucleotide binding	8



2,4-diaminobutyric acid L-glutamic acid L-glutamine

Figure A9: Structures of 2,4-diaminobutyric acid, L-glutamic acid and L-glutamine. *The consistent regions are highlighted.*